Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2015, Article ID 875735, 8 pages
http://dx.doi.org/10.1155/2015/875735

*Research Article*

# Saliency Mapping Enhanced by Structure Tensor

**Zhiyong He,[1] Xin Chen,[2] and Lining Sun[1]**

[1]*School of Mechanical and Electric Engineering, Soochow University, Suzhou 215021, China*
[2]*NovuMind Inc., Santa Clara, CA 95131, USA*

Correspondence should be addressed to Zhiyong He; he-zhiyong@139.com

Received 4 June 2015; Revised 11 September 2015; Accepted 27 September 2015

Academic Editor: Paolo Del Giudice



We propose a novel efficient algorithm for computing visual saliency, which is based on the computation architecture of Itti model. As one of well-known bottom-up visual saliency models, Itti method evaluates three low-level features, color, intensity, and orientation, and then generates multiscale activation maps. Finally, a saliency map is aggregated with multiscale fusion. In our method, the orientation feature is replaced by edge and corner features extracted by a linear structure tensor. Following it, these features are used to generate contour activation map, and then all activation maps are directly combined into a saliency map. Compared to Itti method, our method is more computationally efficient because structure tensor is more computationally efficient than Gabor filter that is used to compute the orientation feature and our aggregation is a direct method instead of the multiscale operator. Experiments on Bruce's dataset show that our method is a strong contender for the state of the art.

## 1. Introduction

*Visual saliency* (also called *visual salience*) refers to the quality or state by which information stands out relative to its neighbors and often attracts human attention [1]. In the subsequent stages, the salient images are preferentially taken as inputs instead of the whole image. As a consequence, visual saliency has been widely applied to various computer vision tasks such as segmentation [2, 3], image retargeting [4–6], object detection [7, 8], image collection [9], and object recognition [10].

Koch and Ullman introduced a basic biologically inspired architecture of visual saliency, referred to as Koch and Ullman model [11]. Then Itti et al. presented a computational architecture to implement and verify the Koch and Ullman model [12]. As summarized in [13], most of the implementation techniques of the visual saliency models generally have three stages: (1) *extraction:* extracting low-level features at locations over the image plane, (2) *activation:* forming activation maps from the features, and (3) *normalization/combination:* normalizing the activation maps and combining them into a single saliency map.

For Itti method, the objective of the first stage is to extract three low-level features, intensity, color, and orientation, and followed by using Difference of Gaussian (DOG) to form total of forty activation maps. Finally, a linear operator is employed to normalize these maps, in which the most salient location is selected by the winner-take-all neural network to generate a saliency map. However, results of Itti method are sometimes blurry and prefer small and local features, which are less useful for some further computer vision applications such as object segmentation and detection.

Despite many advances of the visual saliency made in recent years, the various evaluation results in [14] indicate that there are still some questions about the mechanism of visual saliency. In addition to a motivation of investigation on some questions such as low-level features and the combination of activation maps, in this paper, we also focus on the performance of algorithm and whether the saliency results can greatly benefit computer vision applications.

Results from some recent research works have shown that features of edge and corner also play important roles in visual saliency [13, 15, 16]. In our study, we also note that orientation features are less likely to win in the combination of activation maps. Moreover, Gabor filters used for orientation extraction are computationally expensive.

We therefore propose an efficient method to compute saliency map, referred to as *structure tensor* (ST) saliency map. The computational architecture of our method is shown

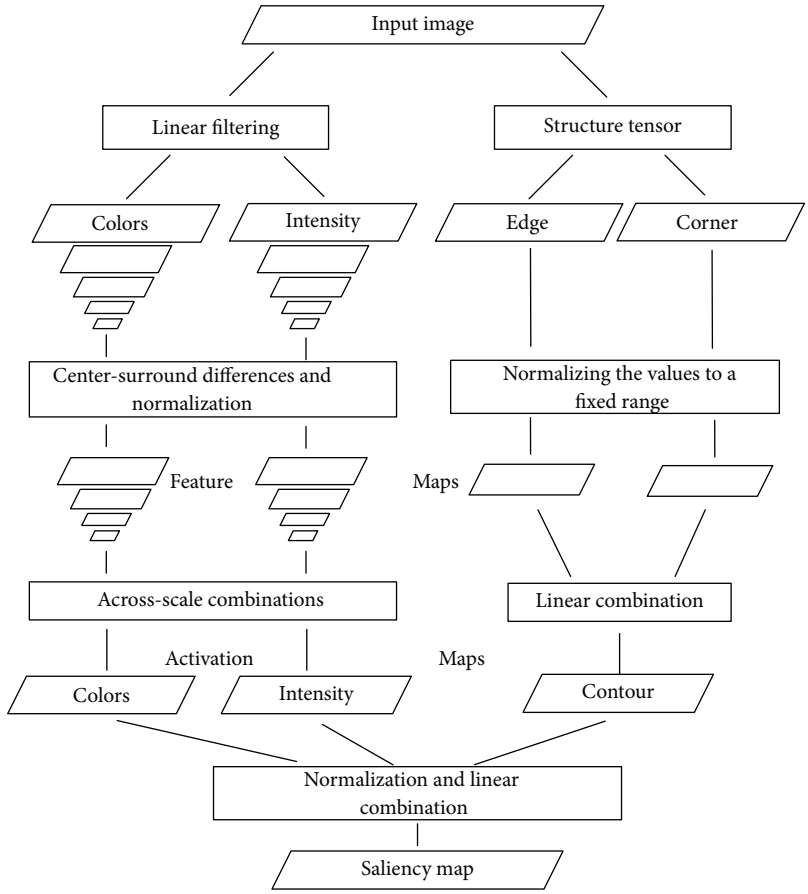


Figure 1: General architecture of our method. In our method, we call the activation map generated by edge and corner features contour activation map. The final saliency map combined these activation maps into ST saliency map.

in Figure 1 and is the same as Itti method in feature extraction and activation maps generation for intensity and color features. The features of edge and corner are extracted by structure tensor and directly combined into an activation map, called contour map. After obtaining three activation maps, we use linear combination to aggregate activation maps to a saliency map instead of multiscale combination and winner-take-all rule.

This paper makes two major contributions as follows:

(1) We propose a novel efficient algorithm to calculate the saliency map. Compared to other methods performed on a challenging dataset, besides the best performance achieved, the results of our method obtain sharper boundaries which are useful in some further applications such as object segmentation and detection.

(2) Our work has shown that edge and corner are two important low-level features in saliency generation.

The paper is organized as follows. Section 2 briefly reviews the state-of-the-art methods with particular emphasis on saliency algorithms related to Itti method, and Section 3 introduces some backgrounds of structure tensor and formally describes our algorithm of saliency map computation. In Section 4, we present our experimental results and quantitative evaluations on a challenging dataset and discuss them. This paper closes with a conclusion of our work in Section 5.

## 2. Related Work

Visual saliency methods are generally categorized into biologically inspired methods and computationally oriented methods. There is an extensive literature on the areas, but here we mention just a few relevant papers. Some surveys are found in [17–19], and some recent progress is reported in [20].

Koch and Ullman [11] proposed a basic architecture of biologically inspired methods and defined a saliency map as a topographic map that represents conspicuousness of scene locations. Their work also introduced a winner-take-all neural network that selects the most salient location and employs an inhibition of return mechanism to allow the focus of attention to shift to the next most salient location. Then Itti et al. presented a computational model to implement and verify Koch and Ullman model. Since then, the works related to the saliency map have quickly become one of the hot research fields.

Itti method employs a Difference of Gaussian (DOG) operator to evaluate color, intensity, and orientation features

to generate total of forty activation maps and across-scale-combines these maps into a saliency map. Besides the expensive computation, one big problem of Itti method is that the results are sometimes blurry and prefers small purely local features. On the other hand, many algorithms of computer vision need input features related to contours because they require the distinct boundary information. Recently, some methods have been proposed to obtain sharp edges, for example, local dissimilarities at the pixel level [21], multiscale DOG [22], and histogram analysis [15]. However, the common problem of these methods is that they are more sensitive to the noises.

As mentioned in the previous section, improving on Itti method, we propose an efficient algorithm for calculating the saliency map, and the computational architecture of our method is shown in Figure 1. The computational architecture of our method is similar to Itti method, and our method evaluates intensity, color, edge, and corner features instead of intensity, color, and orientation features. The structure tensor is used to extract the features of the edge and corner. In the final step, we use the linear combination to generate a saliency map instead of the winner-take-all rule of Itti method.

## 3. The Proposed Saliency Model

In this section, we briefly introduce the background of structure tensor and formally describe our algorithm.

*3.1. Introduction to Structure Tensor.* In mathematics, structure tensor is a matrix representation of partial derivative information. In the field of image processing and computer vision, it typically represents the gradient or edge information and has a more powerful description of local patterns as opposed to the directional derivative through its coherence measure [23, 24].

There are two categories of structure tensor: linear structure tensor and nonlinear structure tensor. Compared to the nonlinear structure tensor, the linear structure tensor is fast and easy to implement with Fast Fourier Transform (FFT). We therefore select the linear structure tensor to extract the features of edges and corners.

Given an image $I(x, y)$, if pixel $(x, y)$ translates to $(x + \Delta x, y + \Delta y)$, the energy $E$ is defined as

$$E = \sum_{(u,v)\in W(x,y)} w(u,v)\left(I(u,v) - I(u+\Delta x, v+\Delta y)\right)^2, \quad (1)$$

where $W(x, y)$ is a window at center point $(x, y)$ and $w(u, v)$ is a weight function at pixel $(x, y)$. In the rest of this paper, $\sum_{(u,v)\in W(x,y)} w(u, v)$ is simply written as $\sum_w$.

It is approximated by a first-order Taylor series:

$$I(u+\Delta x, v+\Delta y) \approx I(u,v) + \frac{\partial I}{\partial x}(u,v)\Delta x + \frac{\partial I}{\partial y}(u,v)\Delta y = I(u,v) + \left[\frac{\partial I}{\partial x}(u,v), \frac{\partial I}{\partial y}(u,v)\right]\begin{bmatrix}\Delta x \\ \Delta y\end{bmatrix}. \quad (2)$$

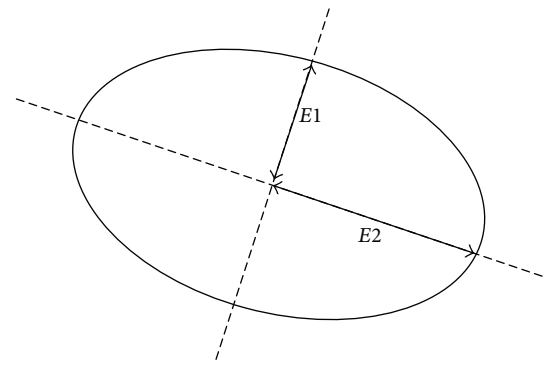


Figure 2: The relation between ellipse and structure tensor.

Hence, (1) can be rewritten as

$$E = \sum_w \left(I(u,v) - I(u+\Delta x, v+\Delta y)\right)^2 \approx [\Delta x, \Delta y]\,\mathbf{T}\begin{bmatrix}\Delta x \\ \Delta y\end{bmatrix}, \quad (3)$$

where matrix $\mathbf{T}$ is

$$\mathbf{T} = \sum_w \begin{bmatrix}\left(\frac{\partial I}{\partial x}\right)^2 & \frac{\partial I}{\partial x}\cdot\frac{\partial I}{\partial y} \\ \frac{\partial I}{\partial x}\cdot\frac{\partial I}{\partial y} & \left(\frac{\partial I}{\partial y}\right)^2\end{bmatrix} = \begin{bmatrix}\sum_w\left(\frac{\partial I}{\partial x}\right)^2 & \sum_w\frac{\partial I}{\partial x}\cdot\frac{\partial I}{\partial y} \\ \sum_w\frac{\partial I}{\partial x}\cdot\frac{\partial I}{\partial y} & \sum_w\left(\frac{\partial I}{\partial y}\right)^2\end{bmatrix}. \quad (4)$$

Matrix $\mathbf{T}$ is a structure tensor, which is also considered as a covariance matrix.

We also consider (3) as an approximation of a binomial function, and from a view of geometry, a binomial function is an ellipse where short axis and long axis are represented as eigenvalues $E1$ and $E2$, respectively. The direction of ellipse is determined by the eigenvectors. As shown in Figure 2, the equation of ellipse is written as

$$[\Delta x, \Delta y]\,\mathbf{T}\begin{bmatrix}\Delta x \\ \Delta y\end{bmatrix} = 1. \quad (5)$$

(1) Input:
(2) Input image **I**: three-channel and size $(m, n)$
(3) Output:
(4) Edge feature map **A**: one channel and size $(m1, n1)$
(5) Corner feature map **B**: one channel and size $(m1, n1)$
(6) Contour activation map **C**: one channel and size $(m1, n1)$
(7) Begin
(8) Resize the input image **I** to $(m1, n1)$, called Im-Re
(9) **for** $j \leftarrow 1, n1$ **do**
(10) **for** $i \leftarrow 1, m1$ **do**
(11) For Im-Re, calculate structure tensor $J_\sigma$ using (6)
(12) Calculate eigenvalues $\lambda_1$ and $\lambda_2$ using (7) and (8), respectively
(13) $A(i, j) = \lambda_1 - \lambda_2$
(14) $B(i, j) = \lambda_1 + \lambda_2$
(15) **end for**
(16) **end for**
(17) Normalize **A** and **B** into a fixed range $[0 \cdots 1]$
(18) Combine normalized **A** and normalized **B** into **CT**
(19) End

Algorithm 1: Algorithm of contour activation map.

Based on (4), some types of structure tensor have been constructed. In our work, we use a linear structure tensor to analyze the input image, and it is defined as

$$J_\sigma = \sum_{i=1}^{3} \begin{bmatrix} K_\sigma * \left(\frac{\partial I_i}{\partial x}\right)^2 & K_\sigma * \left(\frac{\partial I_i}{\partial x} \cdot \frac{\partial I_i}{\partial y}\right) \\ K_\sigma * \left(\frac{\partial I_i}{\partial x} \cdot \frac{\partial I_i}{\partial y}\right) & K_\sigma * \left(\frac{\partial I_i}{\partial y}\right)^2 \end{bmatrix}, \quad (6)$$

where $K_\sigma$ is a Gaussian kernel with variance $\sigma$ and $*$ is a convolution operator. The parameter $i$ is the image channel number.

For any kind of structure tensor, we use $\left[\begin{smallmatrix} G & F \\ F & H \end{smallmatrix}\right]$ to simply represent matrix **T** of (4). Then the two eigenvalues are calculated as

$$\lambda_1 = \frac{G + H + \sqrt{(G - H)^2 + 4F^2}}{2}, \quad (7)$$

$$\lambda_2 = \frac{G + H - \sqrt{(G - H)^2 + 4F^2}}{2}. \quad (8)$$

*3.2. Contour Activation Map.* As shown in Figure 1, we calculate the activation maps of color and intensity with Itti method, and the contour activation map is detailed in Algorithm 1.

For computation of **A** and **B**, we do not need to compute $\lambda_1$ and $\lambda_2$ with (7) and (8) and add and subtract these values to calculate **A** and **B**. We directly compute them as

$$\begin{aligned} \lambda_1 - \lambda_2 &= \sqrt{(G - H)^2 + 4F^2}, \\ \lambda_1 + \lambda_2 &= G + H. \end{aligned} \quad (9)$$

In the final step, we combine feature maps into a contour activation map **CT** as follows:

$$\mathbf{CT} = \frac{1}{2}\left(N(A) + N(B)\right), \quad (10)$$

where $N(A)$ is the normalized edge feature map and $N(B)$ is the normalized corner feature map.

*3.3. ST Saliency Map Generation.* We assume that all features equally contribute to the ST saliency map generation. After obtaining the contour activation map, the intensity activation map, and the color activation map, we combine them into a saliency map as follows:

$$S = \frac{1}{3}\left(\bigoplus(\mathbf{CL}) + \bigoplus(I) + \bigoplus(\mathbf{CT})\right), \quad (11)$$

where $\bigoplus$ is a normalization operation which is defined in [12], **CL** is the color activation map, **I** is the intensity activation map, and **CT** is the contour activation map.

Some saliency maps of our method are shown in Figure 3 and these maps have distinct boundaries.

## 4. Experimental Results

In this section, we present subjective evaluation and quantitative analysis of our method and some state-of-the-art methods and analysis of performance of these methods.

*4.1. Saliency Maps.* We compared saliency maps of our method with saliency maps of some state-of-the-art methods including Itti method [12], Attention Based on Information Maximization (AIM) method [25, 26], Dynamic Visual Attention (DVA) method [27], Graphic-Based Visual Saliency (GBVS) method [13], and Image Signature (IS) method [28]. The MATLAB implementation of these methods is based on

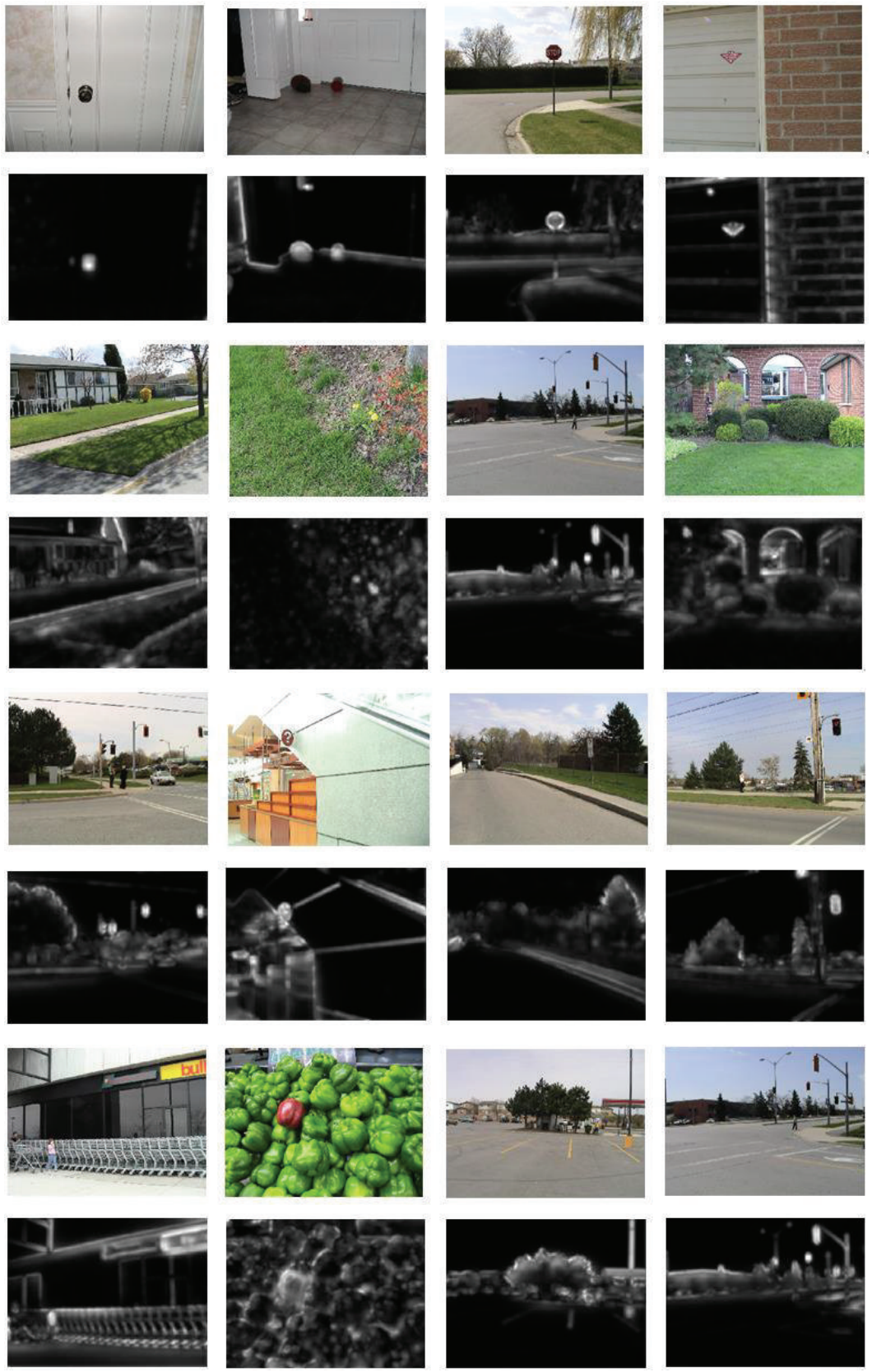

Figure 3: Saliency maps of our method. The odd row is the input images from Bruce dataset, and the even row is the saliency generated by our method. Obviously, the saliency maps of our method have sharp edges.


Figure 4: Saliency maps on the Bruce dataset. (a) Input image, (b) our method, (c) Itti method, (d) AIM method, (e) DVA method, (f) GBVS method, and (g) IS method using LAB color space. Since our method includes the edge and corner information, saliency maps of our method have sharp edges that are useful for the further steps in some computer vision tasks.

the codes on the authors' websites. Saliency maps are shown in Figure 4.

*4.2. Analysis of Performance.* We evaluated our method on Bruce dataset containing 120 natural images with eye fixation ground truth data. In Bruce dataset, the size of all images is $681 \times 511$. Some of methods are sensitive to different sizes of the input image. As a consequence, in order to fairly evaluate results of different methods, we resize the input images to the same size ($170 \times 128$) for each method.

Results from perceptual research works [29, 30] have found that human fixations have strong center bias which may affect the performance of a saliency algorithm. To remove this center bias, following the procedure of Tatler et al.'s work [29], Hou et al. [28] introduced ROC Area Under the Curve (AUC) score to quantitatively evaluate the performance of different algorithms. Good results should maximize the ROC AUC score. To compare the ROC AUC scores, we follow the computation method provided by [28], but the size ($170 \times 128$) is different with two input image sizes used in [28]. Comparison of the ROC AUC scores is shown in Figure 5.

We conducted our tests on a laptop with Intel Dual-Core i5-4210U 1.7 GHz CPU and 4 G RAM memory. All codes were written in MATLAB.

The execution times of the methods are summarized in Figure 6, in which the time is an average time of 120 images.

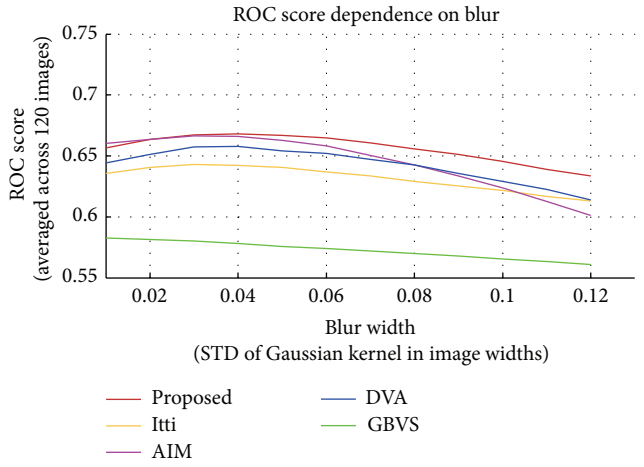


Figure 5: Comparison of the ROC AUC scores of all methods. Our method achieves the best ROC AUC score.

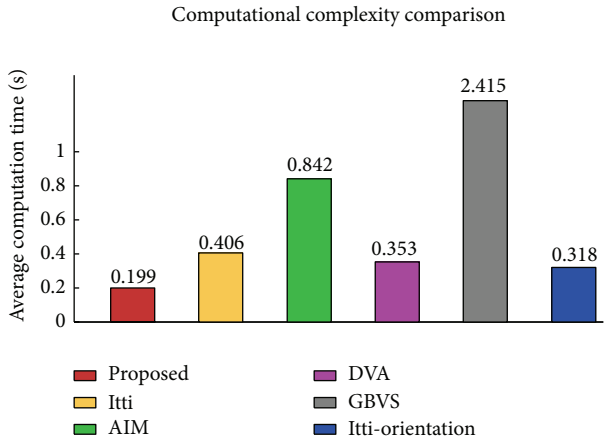


Figure 6: Results of the performance of these different methods. Time measurements are given in seconds. The results are the average times of 120 images of Bruce dataset.

The figure shows that our method is about twice as fast as Itti method and outperforms other state-of-the-art methods. The reason lies in two parts. First, structure tensor is an efficient algorithm of feature extraction. Second, we directly aggregate three activation maps into a saliency map. It is obvious that the performance will increase greatly if our method is implemented by C/C++, and it should satisfy most of the real time applications.

## 5. Conclusion

In this paper we have proposed an efficient algorithm for computing the saliency map, which has a distinct boundary that contributes to further computer vision applications such as segmentation and detection. The computational architecture of our method is close to Itti method, but we have made two improvements in low-level features extraction and combination of activation maps. Since features of edge and corner are important cues in visual saliency, we use a linear structure tensor to extract these features. The reason that our algorithm is efficient lies in the following: (1) linear structure tensor is an efficient feature extraction algorithm and (2) our linear combination method is fast. On the basis of experimental results on Bruce dataset, our method has shown that some computer vision tasks, in particular real time applications, can benefit from our method as a preprocessing step.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This work was supported in part by the Chinese National Natural Science Foundation (NSFC 61473201, 51405320), the Natural Science Foundation of Jiangsu Province (BK20150339), and the Science and Technology Program of Suzhou (SYG201424).

## References

[1] M. Carrasco, "Visual attention: the past 25 years," *Vision Research*, vol. 51, no. 13, pp. 1484–1525, 2011.

[2] J. Han, K. N. Ngan, M. Li, and H.-J. Zhang, "Unsupervised extraction of visual attention objects in color images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 1, pp. 141–145, 2006.

[3] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, "Segmenting salient objects from images and videos," in *Computer Vision—ECCV 2010*, vol. 6315 of *Lecture Notes in Computer Science*, pp. 366–379, Springer, Berlin, Germany, 2010.

[4] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Transactions on Graphics*, vol. 26, no. 3, article 10, 2007.

[5] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, 2012.

[6] D. Vaquero, M. Turk, K. Pulli, M. Tico, and N. Gelfand, "A survey of image retargeting techniques," in *Applications of Digital Image Processing XXXIII*, vol. 7798 of *Proceedings of SPIE*, pp. 779–814, SPIE Optical Engineering + Applications, San Diego, Calif, USA, August 2010.

[7] A. Oliva, A. Torralba, M. S. Castelhano, and J. M. Henderson, "Top-down control of visual attention in object detection," in *Proceedings of the International Conference on Image Processing (ICIP '03)*, pp. I-253–I-256, September 2003.

[8] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 853–860, Providence, RI, USA, June 2012.

[9] M.-M. Cheng, N. J. Mitra, X. Huang, and S.-M. Hu, "SalientShape: group saliency in image collections," *The Visual Computer*, vol. 30, no. 4, pp. 443–453, 2014.

[10] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?" in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. II-37–II-44, IEEE, July 2004.

[11] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.

[12] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[13] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS '06)*, pp. 545–552, Vancouver, Canada, December 2006.

[14] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 409–416, Providence, RI, USA, June 2011.

[15] T. Liu, Z. Yuan, J. Sun et al., "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011.

[16] R. Valenti, N. Sebe, and T. Gevers, "Image saliency by isocentric curvedness and color," in *Proceedings of the 12th IEEE International Conference on Computer Vision*, pp. 2185–2192, IEEE, Kyoto, Japan, September-October 2009.

[17] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, 2013.

[18] A. Borji, H. R. Tavakoli, D. N. Sihite, and L. Itti, "Analysis of scores, datasets, and models in visual saliency prediction," in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV '13)*, pp. 921–928, Sydney, Australia, December 2013.

[19] S. Frintrop, E. Rome, and H. I. Christensen, "Computational visual attention systems and their cognitive foundations: a survey," *ACM Transactions on Applied Perception*, vol. 7, no. 1, article 6, 2010.

[20] Z. Bylinskii, T. Judd, A. Borji et al., "MIT Saliency Benchmark," 2015, http://saliency.mit.edu/index.html.

[21] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proceedings of the 11th ACM International Conference on Multimedia (MM '03)*, pp. 374–381, ACM, November 2003.

[22] L. Itti and P. F. Baldi, "Bayesian surprise attracts human attention," in *Advances in Neural Information Processing Systems*, pp. 547–554, MIT Press, 2005.

[23] T. Brox, J. Weickert, B. Burgeth, and P. Mrázek, "Nonlinear structure tensors," *Image and Vision Computing*, vol. 24, no. 1, pp. 41–55, 2006.

[24] U. Köthe, "Edge and junction detection with an improved structure tensor," in *Pattern Recognition*, pp. 25–32, Springer, Berlin, Germany, 2003.

[25] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS '05)*, pp. 155–162, Vancouver, Canada, December 2005.

[26] N. D. B. Bruce and J. K. Tsotsos, "Saliency, attention and visual search: an information theoretic approach," *Journal of Vision*, vol. 9, no. 3, article 5, 2009.

[27] X. Hou and L. Zhang, "Dynamic visual attention: searching for coding length increments," in *Advances in Neural Information Processing Systems*, pp. 681–688, MIT Press, 2009.

[28] X. Hou, J. Harel, and C. Koch, "Image signature: highlighting sparse salient regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 194–201, 2012.

[29] B. W. Tatler, R. J. Baddeley, and I. D. Gilchrist, "Visual correlates of fixation selection: effects of scale and time," *Vision Research*, vol. 45, no. 5, pp. 643–659, 2005.

[30] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, "SUN: a bayesian framework for saliency using natural statistics," *Journal of Vision*, vol. 8, no. 7, article 32, 2008.