Multi-scale Feature Fusion for 3D Saliency Detection

Gang Pan1, An-zhi Wang2*, Bao-lei Xu3 and weihua OU2

1School of Intelligent Manufacturing, Sichuan University of Arts and Science, Dazhou, Sichuan, 635000, China
2Guizhou Normal University, Guiyang, Guizhou, 550025, China
3Teaching Affairs Office, Sichuan University of Arts and Science, Dazhou, Sichuan, 635000, China

20110037@sasu.edu.cn *Corresponding author’s e-mail: andyscu@163.com

Abstract. 3D saliency detection aims to take advantage of the disparity map, depth map and color information to automatically detect informative objects from natural scenes. Although studies have concentrated on this issue in recent years, there are challenges such as how to leverage disparity map or depth map effectively to compute depth-induced saliency, and how to fuse optimally multiple visual features and cues. A novel 3D saliency detection approach is proposed, which fuses local contrast, region contrast, texture feature, depth cue, and location cue into a unified saliency computation framework. Results show that the proposed approach achieves significant and consistent improvements on other advanced methods in the RGBD1000 datasets.

1. Introduction

Saliency detection is a method that simulates the human visual system to detect the location and area information of interest in an image. Accurate extraction of these location and region information not only improves efficiency and accuracy of image processing, but also plays a critical role in dealing with key problems in machine vision such as image segmentation[1], object recognition[2-4], image compression[5] and image retrieval[6]. Since saliency detection plays an important role in machine vision study, it has attracted researchers’ attention. They proposed approaches of saliency detection, for example, computing image saliency based on its brightness, color and direction [7], detecting image saliency based on color and texture feature[8], detecting salient objects in images based on contrast feature of image’s Euclidean distance in space and in color space from global perspective[9], local perspective[10] or combination of global and local perspective[11]. Against the technical background, these approaches achieved better detection results; however, with improvement of image pixel, heavy computation of image processing might appear. In order to reduce heavy computation and decrease complexity of image processing, Ren et al. [12] proposed a method taking certain pixel points neighboring and similar in terms of color, brightness, texture and the like as super-pixels, which would be computed. Wang et al. [13] came up with a multi-scale approach which could be used to make up for a deficiency of different scales segmentation so as to decrease inaccurate segmentation of scales which would influence saliency detection structure. The above-mentioned studies got better results of saliency detection; however, due to limitations of technical conditions, these approaches were used based on two-dimensional information, while human vision system evolves in the three-dimensional world, so studying saliency detection approaches in three-dimensional information...
plays a significant role in studying vision saliency detection. With improvement of three-dimensional sensor technology, shape, color and distance information of object could be acquired, and sensitivity of scenes is improved, which facilitate the research of saliency detection on image in three-dimensional scenes. Therefore, this paper proposes a 3D saliency detection algorithm based on multi-scale feature fusion, whose advantages are that it is able to combine local contrast, region contrast and texture feature in two-dimensional scenes with depth cue in three-dimensional scenes in order to detect image’s saliency. It can better deal with the interference factors in the complex background. The contrast effect of detection results of the algorithm proposed in this paper and those of present popular algorithm is shown in Figure 1.

![Saliency Detection Comparison](image-url)

Figure 1. Example of Saliency Detection comparing our method with recent state-of-the-art methods

2. Related Work

This paper focuses on related work of multi-scale algorithm, local contrast, region contrast, texture feature, location cue and depth cue and briefly introduces them. Local saliency model estimates image’s saliency by computing difference between image region and its local neighboring region. Itti et al.[7] was the first to compute saliency in this way and get saliency map by computing difference of center-surround based on colors, intensity and orientations features. Based on the theory, Harel et al. [10] proposed a saliency model which formed activation maps on certain feature channels, normalizing them in a way which highlights conspicuity and admits combination with other maps. Chen et al.[14] put forward a measurement of local contrast, which gets local contrast map of input image by measuring difference between current location and its neighboring region. Nouri et al. [15] utilized local patches adapting to different sizes to describe grid saliency. These models used local contrast to form saliency mapping, which usually highlight object’s margin. In order to decrease influence of local contrast’s shortcomings on results of saliency detection, Tong et al. [16] proposed an algorithm of salient object detection based on linear coding by making up for and fusing local and global information, which could make saliency detection result more consistent and robust. Based on super-pixel segmentation, Cheng et al.[9] defined region saliency, using contrast of current region and other regions and space relation of regions. Wang et al.[17] utilized color and texture feature, getting multi-scale simplified image based on super-pixel and via region assimilation to reduce image’s complexity and improve accuracy of detection of salient object. Being different from the
above-mentioned saliency detection studies, this paper proposes a three-dimensional saliency detection approach fusing local contrast, region contrast, texture feature and depth cue into a unified saliency computation framework and detects image’s saliency in three-dimensional scenes. Its main contributions are as follows.

(1). A saliency detection algorithm featured by multi-scale and multi-feature is proposed, including multi-scale segmentation, saliency computation of different salient models, formation of saliency map and multi-scale fusion of different saliency maps.

(2). Multi-feature and depth prior information is fused to improve saliency detection.

3. Proposed Approach

Algorithms in this paper mainly involve multi-scale segmentation, Saliency Computation Method and multi-scale fusion of saliency map.

3.1. Multi-scale Image Segmentation.

Difficulty of discovering and acquiring different images’ features varies from one scale to other scale. In order to acquire image’s feature, this paper, based on\cite{18-20}, utilizes image segmentation algorithm of SLIC\cite{21} to segment the input image I into three super-pixel scales in coarse-to-fine method, setting the number of super-pixel as $N_k \in \{100,200,300\}$, and $k \in \{0,1,2\}$ as image scale and computing the saliency value ($S_k$) in each scale respectively.

3.2. Saliency Computation Method

3.2.1. Computation of Local Contrast.

In image saliency detection, saliency of a super-pixel refers to difference between super-pixel and color, location and features of its neighboring super-pixels, namely, the greater the difference is, the larger the contrast is, and the more attention could be attracted. The mathematical expression of contrast is Euclidean distance, so the longer the square distance is, the bigger the change of color and texture feature is. In order to compute contrast of saliency of a super-pixel in a SLIC\cite{21-22} segmented super-pixel image and its neighboring region, a computation function of a local contrast is designed by using space distance and color distance of two neighboring super-pixels. The formula is as follows.

$$S_{local}(SR) = \frac{1}{mk} \sum_{SP_i \in NBSP_i} \exp \left( -\frac{D_s(SP_i, SP_j)^2}{\sigma_1^2} \right) D_c(SP_i, SP_j)$$  \hspace{1cm} (1)

In the formula, SP$_i$ refers to the super-pixel of i, $i \in [1,N_k]$; $m$ means the number of neighboring regions; $K$ is normalization factor, which guarantees the sum of weighted coefficient of space distance is one; NBSP$_i$ refers to aggregate of neighbors of the SP$_i$.

$$D_s(SP_i, SP_j) = \sqrt{(x_{SP_i} - x_{SP_j})^2 + (y_{SP_i} - y_{SP_j})^2}$$ is Euclidean distance of two neighboring super-pixel regions; $\sigma_1$ is influence coefficient of distance of super-pixel regions on local saliency;

$$D_c(SP_i, SP_j) = \sqrt{(\tilde{x}_{SP_i} - \tilde{x}_{SP_j})^2 + (\tilde{y}_{SP_i} - \tilde{y}_{SP_j})^2 + (\tilde{a}_{SP_i} - \tilde{a}_{SP_j})^2 + (\tilde{b}_{SP_i} - \tilde{b}_{SP_j})^2}$$ is Euclidean distance of average gray of super-pixel region.

3.2.2. Computation of Region Contrast.

Details of salient region detected via local contrast are much clearer, but they are easier to be trapped in local optimum. In order to make up for the shortcoming, Cheng et al.\cite{9} integrated global space relation with region contrast computation from global perspective to effectively detect salient region and avoid local optimum; meanwhile, it considered influence of space distance of two super-pixel regions on saliency value. However, under the circumstance of two region contrasts being the same, the saliency value of central region is usually greater than that of neighboring regions\cite{23-24}. Therefore, we weight central location in region saliency result in order to get the final region saliency map.
\[ S_{Global}(SP_k) = \sum_{SP_k \neq SP_r} \exp \left( -\frac{D_S(SP_k, SP_r)}{\sigma^2} \right) w(SP_r) D_C(SP_k, SP_r) \exp \left( -\frac{D_S(SP_k - I_{center})^2}{\sigma^3} \right) \]

In the formula, \( D_S(SP_k, SP_r) \) is space distance of \( SP_k \) and \( SP_r \); \( w(SP_r) \) is weight of \( SP_r \); \( \sigma^2 \) is strength coefficient of space weight. The bigger the \( \sigma^2 \) is, the less the influence is, and the contrast of farther regions would make greater contributions to saliency value of current region, otherwise, it will get opposite result. \( I_{center} \) is coordinate of image center; \( \sigma^3 \) is influence coefficient of region location on region saliency, namely, the bigger the \( \sigma^3 \) is, the less the influence of region location on saliency; otherwise, the greater the influence.

### 3.2.3. Extraction of Texture Feature

Having features of rotation invariance and stronger anti-noise capability, texture is a global statistical characteristics which is mainly used to describe surface properties of image or image region’s corresponding scenery. There are usually numerous complex textures in image background such as grassland and brick wall, which might interfere with saliency detection so that result of saliency detection would be influenced. Therefore, it plays an important role in improving effect of saliency detection to extract texture feature of image and integrate it with other features via taking complementary advantages. Texture feature of an image is expressed by gradient[25], so for an input image \( I(x, y) \), its texture feature could be expressed as

\[ \text{Feature}_G = (I \times P_x)^2 + (I \times P_y)^2 \]

In the formula, \( P_x \) and \( P_y \) are horizontal and vertical Prewitt operators respectively. The convolution of discrete gabor wavelet transform of Image I [26] is expresses as

\[ G_{MN}(x, y) = \sum \sum I(x - h, y - w) \psi_{MN}(h, w) \]

In this formula, \( h \) and \( w \) are filter mask size variables, and \( \psi_{MN} \) is complex conjugate of \( \psi_{MN} \), among which \( \psi_{MN}(x, y) \) is expressed as

\[ \psi_{MN}(x, y) = \frac{1}{2\pi \sigma_x \sigma_y} \exp \left( -\frac{1}{2} \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \exp (j2\pi W x) \]

In the formula, \( W \) is complex modulation frequency of Gaussian function. Gabor filter is used to filter texture feature image. ST, the texture saliency map, could be expressed as

\[ S_T = \text{Feature}_G \times \psi_{MN}(x, y) \]

### 3.2.4. Computation of Depth Prior

In visual perception, the nearest object could attract the most attention of observer[27-29]. Therefore, depth feature plays a significant role in saliency detection. In order to explore relation between depth and saliency, Lang et al.[30] used Gaussian mixture model to learn depth prior, which describes nonlinear relation between saliency in different DOF and depth. In order to avoid influence exerted by DOFs on saliency detection, we firstly normalize absolute depth value of image to range from 0 to 1 [0,1], setting \( D \) and \( S \) as depth image and saliency map of image I, so depth distribution and probability of salient object could be expressed as \( P(D|S) \) and \( P(D) \). Once \( D \) is given, we can compute posterior probability, namely, \( P(S|D) = \frac{P(D|S)P(S)}{P(D)} \). According to conclusion of [31], posterior distribution basically tallies with formula of \( S_{DP}(SP_l) = \sqrt{1 - DP(SP_l)} \). Therefore, the formula could be used to estimate depth prior map. In the formula, \( DP(SP_l) \) is average depth value of \( SP_l \)(the super-pixel region). The salient map of depth prior could be expressed as

\[ S_{DP} = \sum_{l=1}^{N_k} \sqrt{1 - DP(SP_l)} \]

### 3.3. Multi-scale Feature Fusion

Corresponding to each scale, after extracting texture feature of each super-pixel image, a texture saliency map is formed, then texture saliency map is fused with depth saliency map on corresponding scale[32], and finally saliency map on this scale \( S_r = S_{Local} \otimes S_{Global} \otimes S_T \otimes S_{DP} \) is formed. The three saliency maps on corresponding scales are fused into the final saliency map S as
\[ S = S_0 + S_1 + S_2 \]  

In order to reduce impact on salient object in process of fusing saliency map, the PageRanking algorithm was adopted to optimize images so as to gain better performance of detecting salient object.

4. Experiments

4.1. Experiment Environment and Dataset

Hardware platform of experiment environment in this paper is i5-6300, 2.8GHz CPU, 4GB memory and software development tool is Matlab2014a. In order to evaluate validity and reliability of the proposed algorithm, this paper, based on datasets of RGBD1000[31], contrasts the algorithm with current nine popular detection algorithms in terms of evaluation indexes including precision, recall, F-measure, PR Curve and MAE. The nine popular detection algorithms are GR[33], FT[34], MSS[35], GC[36], GU[37], HC[9], SEG[38], LC[14] and RC[9]. The contrast method was derived from public codes and detection results given by authors of the above-mentioned nine studies.

4.2. Visual Comparison

Results of the algorithm and the eight algorithms based on RGBD1000 are shown in Figure 2.

As can be seen from Figure 2, FT, GC, GU, HC, LC, MSS and SEG can identify location of salient object, and meanwhile identify certain background area so that they cannot suppress impact exerted by background area on saliency detection. RC can identify a general location of salient object and certain
meaningless background area, and sometimes it identifies surrounding areas of a salient object as a salient area. The algorithm proposed not only identifies salient object wholly and accurately, but also suppresses interference of background areas, which performs better than other algorithms.

4.3 Quantitative Comparison
The better performance of the algorithm proposed were shown by analyzing the PR Curve, Mean Absolute Error and F-measure value of other nine algorithms on RGBD1000. The PR Curve was firstly obtained by calculation on RGBD1000. The contrast effect of the proposed algorithm and other nine algorithms is shown in Figure 3. The larger the cover area of PR Curve is, the better the performance of the algorithm is. As can be seen from Figure 3, the PR Curve’s cover area of the proposed algorithm is larger than those of other algorithms, which shows proposed algorithm gains better accuracy than other algorithms in terms of saliency detection.

In order to contrast the proposed algorithm with other algorithms in terms of calculated saliency value and real saliency value, MAE of these algorithms obtained on RGBD1000 and the results were shown in Figure 4. As can be seen from Figure 4, the MAE of the proposed algorithm is the smallest, which is the result of weighting and fusing of local contrast saliency map, region contrast saliency map, texture saliency map and depth cue map and then using PageRanking to optimize images so that fallout rate is decreased and saliency value is striking. Therefore, general saliency value is closer to real value.

4.4 Failures
The proposed algorithm achieves state-of-the-art performance. However, as the algorithm is based on features of 2D and depth cue of 3D of the image, so if the image is with extreme low contrast, it will influence final performance. Figure 5 shows some failures. If there were complex background textures in picture and saliency of the object in depth cue was not striking, poor detection results might be gained. Main reason lies in that the unified saliency computation framework did not study low level features, which cannot distinguish effectively salient object from complex textures.
5. Conclusions

This paper proposes a 3D saliency detection algorithm based on multi scale feature fusion, comprehensively considering image’s multi scale factors. On the basis of segmenting input images into three super-pixel scales, it integrates local contrast, region contrast, texture feature and depth cue, obtaining fusion saliency maps on different scales. Then, it fuses the saliency maps on three scales into the final saliency map. Finally, it conducts experiments on RGBD1000. Experimental results show the algorithm proposed makes improvement in accuracy and recall rate, proving it is necessary to consider image’s depth cue information in salient object detection. In conclusion, the proposed algorithm gains better performance in terms of visual effect, PR Curve and F-measure value; meanwhile, it fits in with several image scenes, having better universality and applicability. For saliency detection, there are several critical problems which need to be handled in the future, including extraction and fusion of multi-feature, estimation of multi-prior algorithm parameter and optimization of processed models.

Acknowledgment

We thank reviewers for valuable suggestions to improve the paper. This work is supported by The National Natural Science Funds Fund(Grants No.61962010), Scientific Research Fund of Sichuan Provinicial Education Department(Grants No.18ZB0507), The 2019 Startup Research Project Sponsored by Guizhou Normal University for Youth doctor under Grant GZNUD[2018]32, Sichuan University of Arts and Science (Grants No.2016KZ007Y, No.2017ZZ004Y, 2017ZZ003G).

References

[1] Vese L A, Chan T F.A multiphase level set framework for image segmentation using the Mumford and Shah model. International Journal of Computer Vision, 2004,50(3):271-293.
[2] LIU Zhen, JIANG Hui, WANG Libin. SAR ATR method based on sparse representation. Computer Engineering and Applications, 2014, 50 (10): 212-215.
[3] Wang A, Wang M. RGB-D Salient Object Detection via Minimum Barrier Distance Transform and Saliency Fusion[J]. IEEE Signal Processing Letters, 2017, 24(5):663-667.
[4] Wang A, Wang M, Li X, et al. A Two-Stage Bayesian Integration Framework for Salient Object Detection on Light Field[J]. Neural Processing Letters, 2017, 46(1):1-12.
[5] Christopoulos C, Skodras A, Ebrahimi T. The JPEG2000 still image coding system: an overview. IEEE Transactions on Consumer Electronics, 2000, 16(4):1103-1127.
[6] Ma Y F, Zhang H J. Contrast-based image attention analysis by using fuzzy growing//Eleventh ACM International Conference on Multimedia, 2003: 374-381.
[7] 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, November 1998.
[8] LU Li-zhen, LIU Ren-yi, LIU Nan.Remote Sensing Image Retrieval Using Color and Texture Fused Features.Journal of Image and Graphics, 2004,9(3):74-78
[9] Cheng M M, Mitra N J, Huang X, et al. Global Contrast Based Salient Region Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3):569-582.
[10] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” Advances in Neural Information Processing Systems, December 2006, pp. 545–552.
[11] Liu ST,Jiang KH, Liu ZX, et al.Saliency detection of infrared images based on local feature and global feature. Journal of Optoelectronics-Laser,2018,29(3):333-339.
[12] Ren X, Malik J. Learning a classification model for segmentation. Iccv, 2003, 1:10-17 vol.1.
[13] Wang Gang, Wang Xiaodong, Chen Chao, et al. A Multi-scale Superpixels Saliency Detection Algorithm. Computer Engineering,2016, 42(7): 257-260, 266.
[14] Chen C L P, Li H, Wei Y, et al. A Local Contrast Method for Small Infrared Target Detection. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(1):574-581.
[15] A. Nouri, C. Charrier, and O. Lezoray, “Multi-scale mesh saliency with local adaptive patches for viewpoint selection,” Signal Processing: Image Communication, vol. 38, pp. 151–166, October 2015.
[16] N. Tong, H. Lu, Y. Zhang, and X. Ruan, “Salient object detection via global and local cues,” Pattern Recognition, vol. 48, no. 10, pp. 3258–3267, October 2015.
[17] Wang H, Dai L, Cai Y, et al. Salient object detection based on multi-scale contrast[J]. Neural Networks, 2018, 101.

[18] Wang J J, Liu Z Y. Multi-scale saliency detection based on composition prior. Journal of Image and Graphics, 2015, 20(12): 1664-1673.

[19] Cheng Pei-rui. Visual Saliency Detection Algorithm Based on Multi-scale Region Contrast[D]. Chinese Academy of Sciences (Changchun Institute of Optics, Fine Mechanics and Physics), 2015.

[20] Song H, Zhi L, Du H, et al. Depth-Aware Salient Object Detection and Segmentation via Multiscale Discriminative Saliency Fusion and Bootstrap Learning[J]. IEEE Transactions on Image Processing, 2017, 26(9): 4204-4216.

[21] Achanta R, Shaji A, Smith K, et al. Slic superpixels[R]. Lausanne, Vaud, Switzerland: Swiss federal Institute of Technology, 2010.

[22] Fan D P, Gong C, Cao Y, et al. Enhanced-alignment Measure for Binary Foreground Map Evaluation[J]. 2018.

[23] Jiang H Z, Wang J D, Yuan Z J, et al. Automatic salient object segmentation based on context and shape prior [C]// Proceedings of the 22nd British Machine Vision Conference. Dundee, Britain: BMVA Press, 2011: 110.1-110.12.

[24] Zhu C, Zhang W, Li T H, et al. Exploiting the Value of the Center-dark Channel Prior for Salient Object Detection[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2019, 10(3): 32.1-32.20.

[25] Zong Baobao, Li Chaofeng, SangQingbing. 3D Image Saliency Detection Based on Log-Gabor Filtering and Saliency Map Fusion Optimization[J]. Laser & Optoelectronics Progress. 2019, 56(08): 109-115.

[26] Singh S M, Hemachandran K. Content-based image retrieval using color moment and gabor texture feature[C]// International Conference on Machine Learning & Cybernetics. IEEE, 2010.

[27] Ding Y, Liu Z, Huang M, et al. Depth-aware saliency detection using convolutional neural networks[J]. Journal of Visual Communication & Image Representation, 2019, 61: 1-9.

[28] Wang W, Shen J. Deep Visual Attention Prediction[J]. IEEE Transactions on Image Processing, 2018, PP(99): 1-1.

[29] Zhu C, Cai X, Huang K, et al. PDNet: Prior-model Guided Depth-enhanced Network for Salient Object Detection[J]. 2018.

[30] Lang C, Nguyen T V, Katti H, et al. Depth Matters: Influence of Depth Cues on Visual Saliency[C]// ECCV (2). Springer-Verlag, 2012.

[31] Ren J, Gong X, Yu L, et al. Exploiting global priors for RGB-D saliency detection[C]// Computer Vision & Pattern Recognition Workshops. IEEE, 2015.

[32] Zhou X, Zhi L, Sun G, et al. Improving Saliency Detection Via Multiple Kernel Boosting and Adaptive Fusion[J]. IEEE Signal Processing Letters, 2016, 23(4): 517-521.

[33] Yang C, Zhang L, Lu H. Graph-Regularized Saliency Detection With Convex-Hull-Based Center Prior[J]. IEEE Signal Processing Letters, 2013, 20(7): 637-640.

[34] A R, H S, E F, et al. Frequency-tuned salient region detection[C]// In proc. CVPR. Florida, USA: IEEE, 2009: 1597–1604.

[35] Bai Yongjian, Xiong Shuhua, Wu Xiaoliang, et al. Infrared and visible images fusion based on FDST and MSS[J]. Science Technology and Engineering, 2017, 17(6): 215–219.

[36] Zhang L. H., Ai J. W., Jiang B. W., et al. Saliency Detection via Absorbing Markov Chain With Learnt Transition Probability[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2018, PP(99):1-1.

[37] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, “Efficient salient region detection with soft image abstraction,” in ICCV, 2013, pp. 1529–1536.

[38] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä,” a, “Segmenting salient objects from images and videos,” in ECCV, 2010.