# Saliency Detection Integrated with Depth Information

R.D. Tang, C.R. Yu, S.H. Qian, W.Y. Yu

School of Electronic and Information Engineering
South China University of Technology
Guangzhou, P.R. China 510641

*Abstract-*Most bottom-up saliency detection methods only perform outstandingly on simple image but when operating on complex one, results are not satisfactory. In a complex scene image, more image features demand being taking into consideration, the most important one of which is the spatial relation of different objects. Aiming at this problem, a new image saliency detection method is proposed based on frequency-tuned (FT) salient region detection. In our method, depth map of the image are constructed and then depth information of every object are added to the calculation of saliency to improve FT. Experimental results illustrate that our proposed method outperforms the original one and thus the salient regions are more evident in the ultimate saliency map.

*Keywords-image saliency detection; complex scene image; spatial relation; depth information*

## I. INTRODUCTION

Human beings have strong visual perception ability after the evolution for a long time. Analysing the surrounding environment and locating the objects that we are interested in efficiently and quickly is quite easy. Salient region detection has a wide range of application such as object recognition [1], image segmentation [2,3] and image compression[4], etc. Moreover, it has already been applied in visual prosthesis to help human with visual impairment [5]. Many existing algorithms of saliency detection including IT [6], AC [7], RC [8], and FT [9] and so on make use of basic image features like color, intensity, orientation to determine contrast of regions with respect to their surroundings.

In a complex image, objects which are closer to us can be more attractive, inspiring us to utilize depth information to calculate their saliency. Depth information enables us to get geometry, location, direction information from the picture. The cues such as texture, gradients, color, etc can be used for depth perception[10].There are several methods that have been put forward to get depth map, for example, using laser radar imaging[13], coordinate measuring machine[14], or movie fringe technique[11,12]. For single monocular images, Markov Random Fields (MRF) can be utilized to construct their depth maps [15]. Since FT [9] is one of the top-performance methods for saliency detection in recent study, we propose a new method based on FT and add depth information to the calculation of saliency.

## II. FREQUENCY SALIENT REGION DETECTION

FT (frequency-tuned salient region detection) [9] is an algorithm to analyze the salient region of an image in frequency domain. The high frequency components of an image express details of the image, while the low frequency components, which we utilize more in salient region detection, reveal the general information. Considering that much noise is contained in the high frequency components, we need to determine the appropriate highest cut-off frequency $\omega_h$. To remove the high frequency noise, we use a combining DoG band pass filter [9]. In practice, FT transforms RGB color space to Lab color space and then utilizes the method of global contrast to calculate its saliency map after a Gaussian Blur filter [16]. The salient value of a pixel p can be formulated as:

$$S(p) = \| I_{\omega_h}(p) - I_\mu \| \qquad (1)$$

Where $I_{\omega_h}(p)$ the image is feature of pixel p after Gaussian blur and $I_\mu$ is the mean image feature vector. The Euclidean distance of these vectors determines its saliency value.

## III. METHOD FT INTEGRATED WITH DEPTH INFORMATION

In our method, we add the depth feature to FT and calculate the saliency map. In many cases, the nearer objects are more attractive to us while the farther ones appeal less, which inspires us that taking the depth information into consideration may help us get a more accurate saliency map.

Before calculation, we need to acquire the actual depth information of each pixel of the image and construct a depth map about the original image, where we use gray value to represent the actual depth of the pixel in visual field. An example of a corresponding depth map constructed from the original map is shown in Fig.1.



FIGURE I.    (A) ORIGINAL MAP; (B) DEPTH MAP.

In order to weaken the noise of background, for the property of exponential function, we utilize exponential damping combined with depth information to revise the saliency value obtained by FT. If we assume the depth value of pixel $p$ is $d$ (ranging from 0 to 255, larger value meaning larger distance), we can construct an decaying exponential function $\exp[-(\frac{d}{\sigma_d})^{\sigma_s}]$, implying that the revised saliency value of pixel $p$ is:

$$S(p) = \| I_{\omega_h}(p) - I_\mu \| \cdot \exp[-(\frac{d}{\sigma_d})^{\sigma_s}] \qquad (2)$$

Where $\sigma_s$ and $\sigma_d$ determine the attenuation intensity.

Consequently, the background in the distance would be attenuated to a large extent, while the saliency value of front objects almost remains the same or gets larger.

## IV. PRE-PROCESSING

We can notice that when we use the above method to combine the result of FT with the depth information, it may pose a problem that some indistinctive objects take a large saliency value for its front position. To reduce problems above, we should take some pre-processing on the image before combing the depth information. We choose to increase the value of the salient objects and decrease that of the indistinctive ones, by which we can highlight the salient objects. In experiment, we still utilize exponential function to do image pre-processing. Assume that the max salient value obtains by FT is $S_{max}$ and the average salient value is $S_{avg}$. The salient value of pixel $p$ after pre-processing can be formulated as:

$$S'(p) = S(p) \cdot \exp[(S(p) - S_{avg})/(S_{max} - S_{avg})] \qquad (3)$$

But remember not all the pixels should be pre-processed. Several pixels may have an extremely larger salient value than that of the most ones which are also salient. Once we pre-process the several outstanding pixels, the difference between the most of salient pixels and the maximum will be enlarged because of the non-linearity property of exponential function. When we normalize the salient value of the pixels, most of the salient pixels may actually obtain a lower salient value compared with not being pre-processing. Therefore, a threshold is of necessity. Only those with the salient value lower than the threshold should be pre-processed and those higher than the threshold should be remained, and in this way, more salient pixels will come close to the maximum.

Using Eqn (1), after pre-processing, the ultimate salient value $S'(p)$ of pixel p can be calculated as:

$$S'(p) = S(p) \cdot \exp[(S(p) - S_{avg})/(S_{max} - S_{avg}) - (\frac{d}{\sigma_d})^{\sigma_s}] \qquad (4)$$

where $S(p)$ can be computed from Eqn (1).

## V. EXPERIMENTAL RESULTS

In our proposed method, we make $\sigma_s$ and $\sigma_d$ equal 5 and 235 respectively to obtain proper attenuation intensity in Eqn (2). We compare our proposed method with FT and other saliency detection methods, and results are shown in Fig.2.



FIGURE II. FROM TOP TO BOTTOM: ORIGINAL IMAGES, SALIENCY MAPS GENERATED FROM SR [17], IT, FT AND OUR PROPOSED METHOD RESPECTIVELY.

We can see from Fig.2 that saliency maps generated by our proposed method fairly highlight the salient regions. Our proposed method outperforms the original ones and the salient regions are more distinctive. Operated on a complex image, many methods fail to remove the less salient regions like the background. Thus oftentimes the background component is

mistakenly assumed to be salient one. In contrast, adding depth information to the original method FT helps to consider the spatial relation of objects in the image. Hence our method performs well in attenuating the background component and effectively eliminates the effect of background. Therefore, it is easier to recognize the salient region of the image since interference by background is removed.

## VI. CONCLUSIONS

We propose a novel saliency detection method based on FT integrated with depth information in this paper. In our method, spatial relationship of objects in the image is taken into consideration and thus front objects may have a larger salient value. Experimental result indicates that our method outperforms original method FT and when dealing with complex images, it presents better saliency detection results.

## ACKNOWLEDGMENTS

## REFERENCES

[1] U. Rutishauser, D. Walther, C. Koch, and P. Perona. Is bottom-up attention useful for object recognition? IEEE Conference on Computer Vision and Pattern Recognition, 2, 2004.

[2] J. Han, K. Ngan, M. Li, and H. Zhang. Unsupervised extraction of visual attention objects in color images. IEEE,Transactions on Circuits and Systems for Video Technology, 16(1):141–145, 2006.

[3] B. C. Ko and J.-Y. Nam. Object-of-interest image segmentation based on human attention and semantic region clustering. Journal of Optical Society of America A, 23(10):2462–2470, 2006.

[4] C. Christopoulos, A. Skodras, A. Koike, and T. Ebrahimi. The JPEG2000 still image coding system: An overview. IEEE Transactions on Consumer Electronics, 46(4):1103–1127, 2000.

[5] Weizhen FU. A Bottom-up Visual Saliency-Based image processing strategy for objects recognition under Simulated Prosthesis Vision. March 2013

[6] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. IEEE PAMI, 1998. 1, 4, 6

[7] R. Achanta, F. Estrada, P. Wils, and S. Susstrunk. Salient region detection and segmentation. In ICVS, 2008. 1, 6

[8] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang, and S. M. Hu. Global contrast based salient region detection. In CVPR, 2011. 1, 4, 6

[9] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned Salient Region Detection. In 2009 IEEE Conference on Computer Vision and pattern Recognition, number Ic, pages 1597-1604. IEEE, June 2009

[10] Make3D: Depth Perception from a Single Still Image. Ashutosh Saxena, Min Sun and Andrew Y. Ng.

[11] D. Nizan, dtal. Measurement and Use of Registed Reflectance and Range Data in Scene Analysis. Processding of IEEE.1997,(Vol.65):206~220

[12] R.A.Lewis, A Scanning Laser Range Finder for A Robotic Vehicle. Proceeding of 5th Hoint Conf. on AI.1997:762~768

[13] M.Hersman,etal. Coherent Laser Radar Application to 3D Vision and Metrology.Proc of Vision 87 Conf., London,1987:465~579

[14] L.Wu, etal.Recognition and Measurement of 3D Objects for Robot Vision. Proc. Of 1995 the International Conf. On Intelligence,Knowledge and Integration for Manufacturing. Nanjing, Sep.1995:299~314

[15] Ashutosh Saxena, Sung H. Chung, Andrew Y. Ng., Learning depth from single monocular images, In NIPS 18, 2005.

[16] Erik Reinhard. High dynamic range imaging: Acquisition, Display, and Image-Based Lighting. Morgan Kaufmann, 2006, pp. 233–234.

[17] X. Hou and L. Zhang, Saliency detection: A spectral residual approach, in IEEE CVPR, 2007, pp.1-8.