Detection of Salient Object Using Pixel Blurriness

Yi-Chong Zeng* and Chi-Hung Tsai†
*Advanced Research Institute, Institute for Information Industry, Taipei, Taiwan
E-mail: yichongzeng@iii.org.tw Tel: +886-2-66072958
†Advanced Research Institute, Institute for Information Industry, Taipei, Taiwan
E-mail: brick@iii.org.tw Tel: +886-2-66072922

ABSTRACT— In this paper we propose a method to detect salient object in still image and non-slow motion background video. The key technique is measuring pixel blurriness. Generally speaking, salient object was taken in focus, pixels within salient object should be sharper than those within background. In the first step image intensity is extracted, and then four different-size average filters are applied to intensity. Subsequently, variation of intensity differences (VID) is computed among the original intensity and four blurred versions. The VID is employed to represent degree of pixel blurriness. Finally, a thresholding method is applied to pixels' blurriness in order to distinguish salient object from background, and salient object is composed of low-blurriness pixels. The experiment results demonstrate that the proposed method is efficient in detection of salient object in still image and non-slow motion background video. Moreover, our method has better detection performance than the two compared methods.

I. INTRODUCTION

In spite of photography skill, when people take a picture, target object should be in focus. Focus is highly associated to image sharpness/blurriness. In the past, researchers proposed many schemes to define criterion for assessment of sharpness/blurriness. In [1], Crete et al. used luminance differences as criterion of image blurriness. In [2], Wee and Paramesran implemented singular value decomposition (SVD) to an image and traced several large eigenvalues to represent image sharpness. Computing means and variances of image luminance, Tsomko et al. [3] used those measures to classify image blurriness into three categories: globally sharp, average quality, and global blurry. Avcıbaş et al. analyzed edge information to generate blur metric [4].

Detection of salient object is helpful for many applications, such as, object segmentation, object recognition, image/video content analysis, and annotation. Forming a saliency map was mentioned in many literatures. In [5], Itti et al. used multi-scale features to form saliency map, those features were extracted based on color, intensity and orientation. They claimed that their method is robust to noise. Liu et al. [6] introduced local, regional, and global features to define salient object, which are multi-scale contrast, center-surround histogram, and color spatial-distribution. They computed saliency map using those features. Zhang et al. [7] proposed a method to form saliency map using natural statistics, which was performed with difference of Gaussian (DoG) filters and linear independent component analysis (ICA) filters. In [8], Achanta et al. computed difference between blurred pixel value and mean pixel value to form saliency maps of color and luminance. The idea of image signature was introduced in [9]. Hou et al. exploited sign of discrete cosine transform (DCT) coefficient referred to as image signature, and then generated saliency map for image.

In this paper we present a method to detect salient object using pixel blurriness. The idea comes from variation of degraded signal. As signal is smoothed by different-size lowpass filters, variation of signal differences to intense-variant signal is bigger than moderate-variant signal is. For two-dimensional image, we consider that in-focus image and out-of-focus image are equivalent to intense-variant signal and moderate-invariant signal, respectively. Similarly, variation of intensity differences to in-focus image is larger than out-of-focus image is. Under this circumstance, we compute variation of intensity differences of a block instead of an image, then, the measure is employed to assess degree of pixel blurriness. Generally speaking, salient object is taken in focus, and it is composed of low-blurriness pixels. Subsequently, a thresholding method is applied to pixels' blurriness to filter out blurry pixels and maintain pixels of salient object. The rest of this paper is organized as follows: in Section II we describe the relationship between blurriness and measure. The proposed method is introduced in Section III. The experimental results will be shown in Section IV, and the concluding remarks will be drawn in Section V.

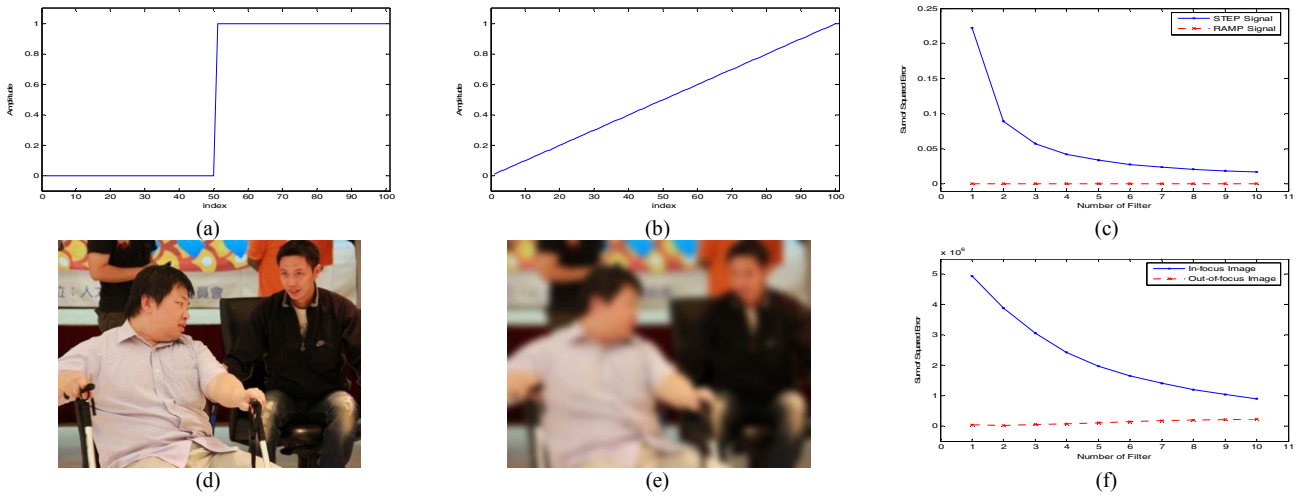
II. THE RELATIONSHIP BETWEEN BLURRINESS AND MEASURE

It is easily understandable that intense-variant signal has larger high-frequency energy than moderate-variant signal does. If signal is smoothed by lowpass filter (such as average filter, Gaussian filter), smoothed signal will maintain low-frequency energy. Therefore, after implementing lowpass filtering, intense-variant signal loses more high-frequency energy than moderate-variant signal does. In other words, large difference occurs between intense-variant signal and its smoothed one, and insignificant difference occurs between moderate-variant signal and its smoothed one.

Figs. 1(a) and 1(b) show a step signal and a ramp signal, respectively. The step signal has steep change at the 51st element, thus, it is treated as an intense-variant signal. The ramp signal is increased gradually, and considered as a moderate-variant signal. Both two signals are smoothed by averages filters with different-size setting. Sum of squared differences (SSD) is formulated by,

$$\varepsilon_k = \sum_{i=1}^{L} (s_{k-1}(i) - s_k(i))^2, \qquad (1)$$

where $s_k(i)$ denotes the i-th element of the k-level smoothed signal, and L is the length of the signal. Fig.1(c) shows two



(a) (b) (c)

(d) (e) (f)

Fig.1. (a) A 100–point step signal, (b) a 100–point ramp signal, and (c) the SSD curves of (a) and (b) against 10 different-size average filters. (d) An in-focus image, (e) an out-of-focus image, and (f) the SSID curves of (d) and (e) against 10 different-size average filters.

SSD curves for the step signal and the ramp signal. It is obvious that the step signal has larger variation of SSD values than the ramp signal does. Hence, signal smoothness is highly associated with variation of SSD values.

Similarly, as image is taken out of focus, image content becomes blurry and edges are smooth. On the contrary, in-focus image has sharp edges. Therefore, we consider out-of-focus image as two-dimensional moderate-variant signal and in-focus image as two-dimensional intense-variant signal. Once image is blurred by average filters with different-size setting, in-focus image loses more high-frequency energy than out-of-focus image does. For example, Figs. 1(d) and 1(e) show an in-focus image and an out-of-focus image, respectively. We compute sum of squared intensity differences (SSID), and the SSID curves of Figs.1(d) and 1(e) are illustrated in Fig.1(f). Obviously, the figure depicts that image blurriness is also highly associated with variation of intensity differences.

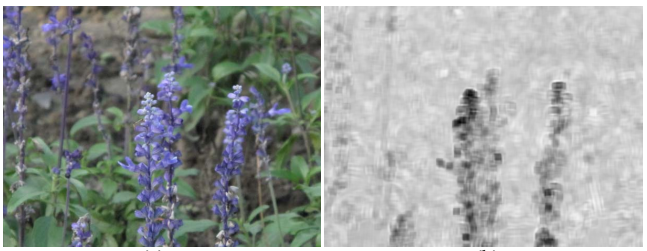
III. The Proposed Method

As we mentioned in Section II, image blurriness is highly associated with variation of intensity difference (VID). Extending the idea of blurriness measuring, VID is computed based on a block instead of an image, and the measure is employed to assess degree of pixel blurriness. In the first step a sliding window of sized $L_w \times L_h$ is employed to extract overlapping blocks over an image. We compute VID of block and it is pixel blurriness. Similarly, every block is blurred by four different-size average filters, i.e., 7×7, 11×11, 15×15, and 19×19 sizes. Assuming that $B_{0,j}$ denotes the j-th block, it is processed by a two-dimensional average filter Φ_k, and then its k-level blurred result $B_{k,j}$ is defined below,

$$B_{k,j} = B_{0,j} \otimes \Phi_k, \text{and } k = \{1,2,3,4\}, \quad (2)$$

where the size of Φ_k is $(4k+3)\times(4k+3)$, and $\|\Phi_k\|_1=1$. Sum of squared intensity difference between two blurred blocks is given by,

$$\psi_{k,j} = \left\| B_{k,j} - B_{k-1,j} \right\|_2^2, \text{and } k = \{1,2,3,4\}, \quad (3)$$



(a) (b)

Fig.2. The test image and its saliency map. (a) A color image of sized 1104×828, and (b) the saliency map. The light-gray pixel and the dark-gray pixel correspond to big blurriness and small blurriness, respectively.

where $\psi_{k,j}$ is the SSID between the k-level and (k-1)-th level blurred blocks, and the symbol $\|X\|_2$ represents the $L2$-norm of the vector X. Subsequently, a straight line is estimated and it approximates to four SSID values (including $\psi_{1,j}$, $\psi_{2,j}$, $\psi_{3,j}$, and$\psi_{4,j}$), then, the slope of line is named variation of intensity difference as well as pixel blurriness. Slope of line, which is abbreviated as s_j, is computed by the following equation,

$$\begin{aligned} s_j &= \frac{-\sum_{k=1}^{4} k \sum_{k=1}^{4} \psi_{k,j} + 4 \sum_{k=1}^{4} k \psi_{k,j}}{4 \sum_{k=1}^{4} k^2 - \left(\sum_{k=1}^{4} k\right)^2} \\ &= -0.3\psi_{1,j} - 0.1\psi_{2,j} + 0.1\psi_{3,j} + 0.3\psi_{4,j}. \end{aligned} \quad (4)$$

Equation (4) conducts a simplified form to compute pixel blurriness. Figs. 2(a) and 2(b) show a test image and its saliency map. The saliency map consists of pixels' blurriness. In Fig.2(b), the light-gray pixel corresponds to the blurry pixel with big blurriness, and the dark-gray pixel corresponds to the sharp pixel with small blurriness. Obviously, Mealy sage consists of the sharp pixels.

However, we are concerned about influence of illumination changing to accuracy of blurriness measure. Illumination changing can be considered as intensity offsetting in a small area (such as a block), which is defined as $\hat{B}_{0,j} = B_{0,j} + c$ and c is a constant. The offset block $\hat{B}_{0,j}$ is processed by different-size average filters, and the SSID between two blurred results of $\hat{B}_{0,j}$ is computed. Equations (2) and (3) can be rewritten as the follows:

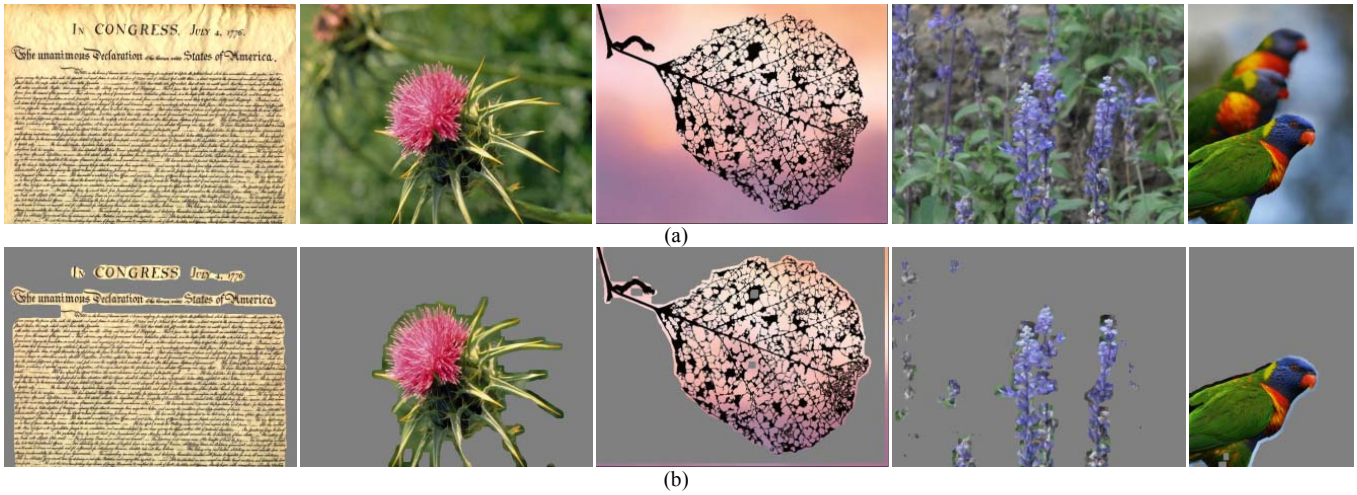


Fig.3. Salient object detection for still images. (a) Five test images (last image is downloaded from [10]), and (b) the detection results of (a)

$$\begin{cases} \hat{B}_{k,j} = \hat{B}_{0,j} \otimes \Phi_k = B_{0,j} \otimes \Phi_k + c\,, \\ \psi_{k,j} = \left\| \hat{B}_{k,j} - \hat{B}_{k-1,j} \right\|_2^2 = \left\| B_{k,j} - B_{k-1,j} \right\|_2^2. \end{cases} \quad (5)$$

According to (5), it is obvious that the SSID value of $B_{0,j}$ is the same as that of $\hat{B}_{0,j}$. Therefore, it demonstrates that the proposed pixel blurriness is invariant to illumination changing.

Subsequently, we design a way to detect salient object. If pixel blurriness is bigger than a threshold, pixel is filtered out. On the contrary, pixel is kept. The detection criterion of salient object is defined as,

$$O_j = \begin{cases} 1 & s_j \le \tau \\ 0 & s_j > \tau \end{cases} \quad (6)$$

If O_j =1, it means that the j-th pixel is identified as a part of salient object; otherwise, the j-th pixel is identified as background. The threshold τ is computed according to $\tau = \frac{1}{HW}\sum_{j=1}^{H\times W} s_j$, where H and W represent the height and the width of an image, respectively. Consequently, the detected result is refined by median filtering.

As the proposed method is implemented to video, salient objects in all frames are detected individually. Subsequently, we compose fragile detection results to reconstruct a complete salient object, which is defined as follows:

$$\begin{cases} j_k^* = \operatorname{argmax}_{j\in\Omega} \| I_{\mathrm{R}}(j) - I_k(j) \|_2 \\ I_{\mathrm{R}}(j_k^*) \leftarrow I_{\mathrm{R}}(j_k^*) \cup I_k(j_k^*) \\ \alpha(j_k^*) \leftarrow \alpha(j_k^*) + 1 \end{cases} \quad (7)$$

where I_{R} and I_k are, respectively, the reconstruction frame image and the k-th frame. j_k^* denotes pixel coordinate in the k-th frame, and it has been identified as a part of salient object by (6). $\alpha(j)$ represents the number of pixel at the j-th pixel coordinate. Consequently, the complete salient object is defined below,

$$\hat{O}_j = \begin{cases} 1 & \alpha(j) > t \\ 0 & \alpha(j) \le t \end{cases} \quad (8)$$

As $\hat{O}_j = 1$, the j-th pixel is identified as a part of refined salient object. In the experiments the threshold is set t=2.

IV. The Experiment Results

A. Detection of Salient Object in Still Image

First, we tested 200 still images to detect salient objects, and large-dimension image was scaled close to 640×480. The objectives of image scaling have two: (1) the proposed method detects salient object without taking a lot of time for large-dimension image; (2) salient object also can be detected in rescaled image. In the first experiment, the parameters were set to L_w=23 and L_h=23. Fig.3(a) shows the five test images, and the corresponding detection results are shown in Fig.3(b). The average computing time of 200 images was 2.55 second, and the programs were run on MATLAB v6.5 with 1.5GHz CPU with 504MB RAM.

B. Detection of Salient Object in Video

In the second experiment four videos were tested. Those videos consist of non-slow motion background frames. Fig.4(a) shows the five frames extracted from one of the test video, and the initial detection results are shown in Fig.4(b). It is obvious that some pixels were misidentified as parts of salient objects. After implementing salient object reconstruction by (7) and (8), we obtained the complete salient objects as shown in Fig.4(c). Obviously, most of misidentified pixels have been filtered out.

C. Comparison

We compared the proposed method with two existing approaches, Zhang et al.'s method [7] and Hou et al.'s method [9]. Two issues were considered in the comparison: computing time and detection performance. Three methods were implemented to 100 images of sized 640×480. Zhang et al.'s method needs more time than the other methods, which spends 110.02 second to form saliency map, averagely. Hou et al.'s method and ours need 0.29 second and 3.93 second, respectively. Fig.5 shows the average recall-precision curves of three methods for the 100 images, the proposed method has best performance among three methods. We just need the 3.57% computed time of implementing Zhang et al.'s method.

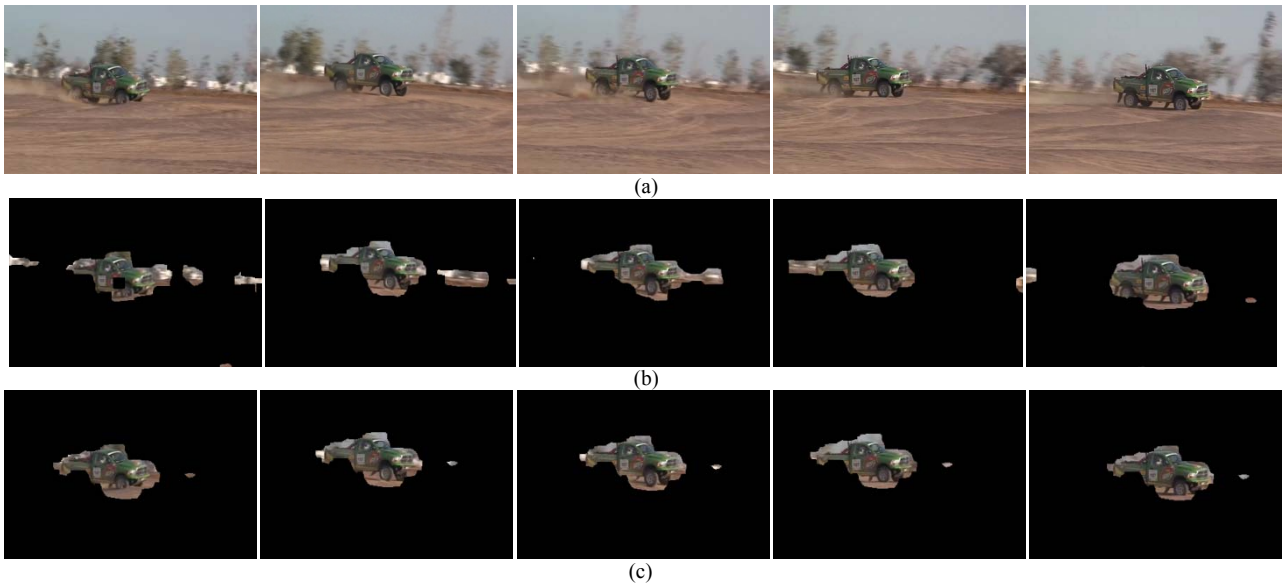


Fig.4. Salient object detection for non-slow motion background video. (a) Five frames extracted from the video [11], (b) the initial detection results, and (c) the complete salient objects

In addition to calculate accuracy rates of salient object and background detection, the maximum accuracy rates of Zhang et al.'s, Hou et al.'s, and our methods are, respectively, 91.6%, 85.5%, and 94.4%.

V. Conclusion

A salient object detection method is proposed in this paper. We measure variation of intensity difference as criterion for representation of pixel blurriness. Small-blurriness pixel is identified as a part of salient object. The proposed method is robust to illumination changing, and the experiment results demonstrate that it is capable of distinguish salient object from background. The average accuracy rate for detection of salient object is 94.4%, and it is better than the compared methods. Consequently, the proposed method is efficient to detect salient object in still image and non-slow motion background video.

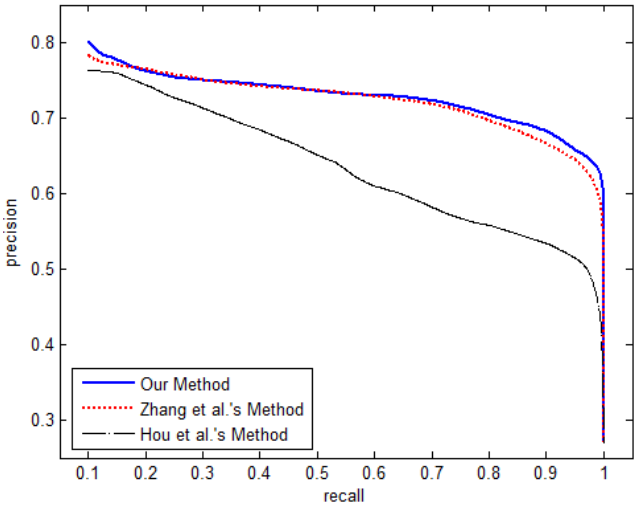


Fig.5 The recall-precision curves of three methods

References

[1] F. Crete, T. Dolmiere, P. Ladret, M. Nicolas, "The blur effect: Perception and estimation with a new no-reference perceptual blur metric," in *Proc. SPIE—Electronic Imaging Symp. Conf. Human Vision and Electronic Imaging*, San Jose, CA, 2007, vol. XII, p. EI6492–16, 2007.

[2] C.-Y. Wee, R. Paramesran, "Measure of image sharpness using eigenvalues," *Information Sciences*, vol.177, no.12, pp.2533-2552, June 2007.

[3] E. Tsomko, H. J. Kim, J. Paik, and I.-K. Yeo, "Efficient method of detecting blurry images," *Journal of Ubiquitous Convergence Technology*, vol.2, no.1, pp.27-39, May 2008.

[4] İ. Avcıbaş, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *Journal of Electronic Imaging*, vol.11, no.2, pp.206-223, April 2002.

[5] L. Itti, C. Koch, E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on PAMI*, vol.20, no.11, pp.1254-1259, 1998.

[6] T. Liu, J. Sun, N.-N. Zheng, X. Tang, H.-Y. Shum, "Learning to detect a salient object," *CVPR*, pp.1-8, 17-22 June 2007.

[7] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, G. W. Cottrell, "SUN: A Bayesian framework for saliency using natural statistics," *Journal of Vision*, vol.8, no.7, article 32, pp.1-20, Dec. 2008. http://cseweb.ucsd.edu/~l6zhang/

[8] R. Achanta, S. Hemami. F. Estrada, S. Süsstrunk, "Frequency-tuned salient region detection," *CVPR,* pp.1597-1604, 20-25 June 2009.

[9] X. Hou, J. Harel, and C. Koch, "Image Signature: Highlighting sparse salient regions", *IEEE Trans. on PAMI*, vol.34, no.1, pp.194-201, Jan. 2012.
http://www.klab.caltech.edu/~harel/share/gbvs.php

[10] Test image, http://www.panoramio.com/photo/2060248

[11] Test video, http://www.youtube.com/watch?v=jPQ4cBxdxi4