Color Spectrum Normalization: Saliency Detection Based on Energy Re-allocation

Zhuoliang Kang[1] and Junping Zhang[2,⋆]

[1] Department of Communication Science and Engineering
Fudan University, Shanghai, 200433, China
[2] Shanghai Key Lab of Intelligent Information Processing
School of Computer Science, Fudan University, Shanghai, 200433, China
zhuoliangkang@gmail.com, jpzhang@fudan.edu.cn

Abstract. Spectrum normalization is a process shared by two saliency detection methods, Spectral Residual (SR) and Phase Fourier Transform (PFT). In this paper, we point out that the essence of spectrum normalization is the re-allocation of energy. By re-allocating normalized energy in particular frequency region to the whole background, the salient objects are effectively highlighted and the energy of the background is weakened. Considering energy distribution in both spectral domain and color channels, we propose a simple and effective visual saliency model based on Energy Re-allocation mechanism (ER). We combine color energy normalization, spectrum normalization and channel energy normalization to attain an energy re-allocation map. Then, we convert the map to the corresponding saliency map using a low-pass filter. Compared with other state-of-the-art models, experiments on both natural images and psychological images indicate that ER can better detect the salient objects with a competitive computational speed.

1 Introduction

Visual saliency detection plays an important role in human vision system since it helps people allocate sensory and computational resources to the most valuable information among the vast amount of incoming visual data [1]. Furthermore, saliency has a broad range of applications in computer vision and engineering fields, such as object recognition, tracking and image resizing [2].

During last decade, a large number of computational models on bottom-up visual saliency have been developed. Some models focus on the center-surround contrast which is closely related to the biological property of human visual system. In these models, saliency is detected either from the center-surround contrast of several features including color, intensity and orientation [3], or from regions of maximal self-information which is a measure of local contrast [4], or from the most discriminant center-surround elements [5]. The other models regard saliency as a kind of image complexity [6,7,8,9]. Lowe [7] referred to the intensity variation in an image as the complexity. Sebe [8] derived the complexity from the absolute value of the coefficients of a wavelet decomposition of the image. Kadir [9] defined the complexity as the entropy of the distribution of local intensities. Recently, Achanta et al. [10] employed a method to attain saliency maps with well-defined boundaries (AC).

[⋆] Corresponding author.

G. Qiu et al. (Eds.): PCM 2010, Part I, LNCS 6297, pp. 248–259, 2010.

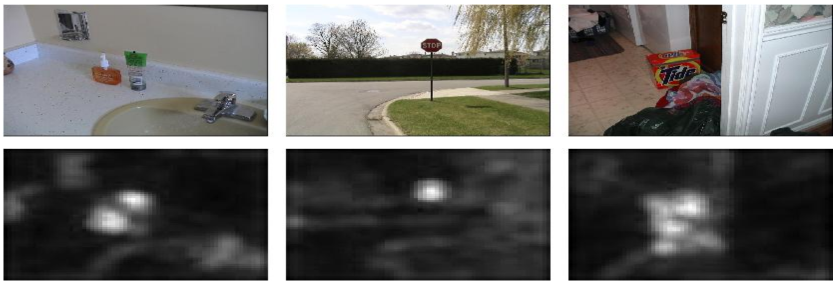

Fig. 1. Top row: Original images collected by Bruce et al. [4]. Bottom row: Their corresponding saliency maps obtained by `ER`.

Compared with the methods in spatial domain, a remarkable advantage of the spectrum based methods is that they have faster computational speed which is essential for practical applications. Two saliency detection methods based on spectral process have been proposed. Hou and Zhang [11] suggested that the spectral residual (`SR`) is a crucial factor to reveal the salient location in an image. Then, Guo et al. [12] proved that a competitive performance can be achieved using the phase information alone (`PFT`). In fact, these two models employ a similar strategy, i.e., setting the amplitudes of all frequency components as close as possible. However, they fail to explain why this spectrum process is reasonable for saliency detection.

Note that when watching an image, people tend to focus on objects with unique properties and ignore background with similar properties (e.g., sky, sea, grass). We observe in Section 2 that the patches in the background resemble each other in the aspect of energy distribution. Color is another important property for people to distinguish the salient objects and the background. Furthermore, Field [13,14] pointed out that the redundant properties of background in natural images can be exploited to produce more effective representations of the original scenes, and one of the most efficient code systems for such representation is based on Fourier Transform.

We thus study saliency detection based on the Energy Re-allocation model (`ER`) in the spectral domain. In `ER`, the energy of regular parts of an image is weakened by setting the energy at all frequencies to be equal. Furthermore, color energy normalization, spectrum normalization and channel energy normalization are combined to weaken the energy of background and make the salient object prominent. The remainder of this paper is organized as follows. In Section 2, we interpret the spectrum normalization mechanism which is shared in `ER`, `SR` and `PFT`. In Section 3, we propose the `ER` model. In Section 4, the performance of `ER` is evaluated on the natural and psychological images. We conclude the paper in Section 5.

2 Spectrum Normalization

Among the spectrum-based methods, `SR` utilizes spectral residual [11] and `PFT` employs the phase Fourier transformation [12] for saliency detection. Although these two methods share the process of setting the amplitudes of all frequency components as close as possible, they fail to clarify that their intrinsic mechanism is, as we point out, spectrum normalization.

Spectrum distribution, which reflects the distribution of energy in the frequency domain, is an important property of both psychological symbols and natural images. Generally speaking, similar spectrum distributions can be observed from similar textures [15], similar psychological symbols, and the patches of the same background in an image. Given an image $\boldsymbol{I}(x,y)$, energy distribution in the frequency domain is represented as follows:

$$A(u,v)e^{-j\phi(u,v)} = \mathcal{F}[\boldsymbol{I}(x,y)] \tag{1}$$

$$E(u,v) = |A(u,v)|^2 \tag{2}$$

where $\mathcal{F}$ denotes a Fourier Transform, $A(u,v)$ and $E(u,v)$ are the amplitude and energy of the two-dimensional spectral component at frequency (u,v) respectively, and $\phi(u,v)$ is the corresponding phase information.

In order to analyze the degree of similarity of spectrum distributions between two image patches $(\boldsymbol{I}_a(x,y), \boldsymbol{I}_b(x,y))$ of size $m \times n$, we define a new criterion, named *Normalized Shared Energy Proportion* (NSEP), as follows:

$$NSEP_{ab} = \frac{1}{mn}\sum_{u,v}\frac{2J_{ab}(u,v)}{E_a(u,v)+E_b(u,v)} \tag{3}$$

where

$$J_{ab}(u,v) = \min(E_a(u,v), E_b(u,v)) \tag{4}$$

denotes an energy distribution located at frequency (u,v), which is commonly owned by the two patches. The $NSEP_{ab}$ represents the normalized proportion of energy distribution shared by the two patches. A larger NSEP means larger proportion of energy distribution is shared by the two patches.

To empirically prove the rationale of NSEP, we perform experiments on both psychological symbols and natural image patches. For simplicity, we extract the patches of background, objects and psychological symbols manually. Firstly, we choose several psychological symbols in patches of size 100×100, including circle, non-closed circle, lines with different orientations shown in Fig. 2 as a test set. We also select nine patches

Table 1. The *Normalized Shared Energy Proportion* of different psychological symbols and natural image patches

Symbol Pairs	NSEP	Patch Pairs (Sky)	NSEP
(a,b)	0.72	(a,b)	0.50
(a,c)	0.32	(a,c)	0.10
(c,d)	0.44	(b,c)	0.11
Patch Pairs (Water)	NSEP	Patch Pairs (Dog)	NSEP
(a,b)	0.47	(a,b)	0.48
(a,c)	0.10	(a,c)	0.14
(b,c)	0.13	(b,c)	0.16

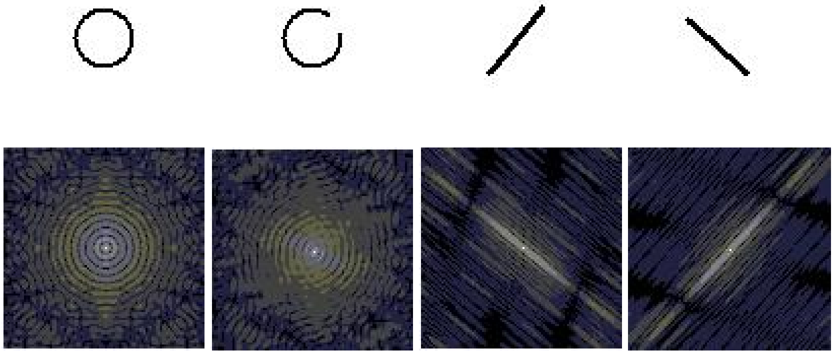

Fig. 2. Psychological symbols (Top row) and their log spectrums (Bottom row). From left to right: (a) circle, (b) unclosed circle, (c) and (d) lines with different orientations.

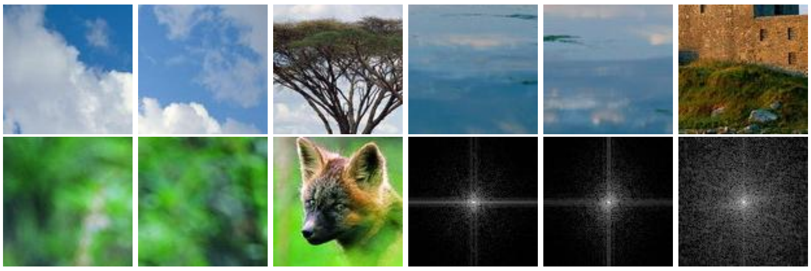

Fig. 3. Background and object patches cropped from three groups of natural images. For each group, from left to right: (a) (b) patches of background (c) patch of object. The last three images in the last row is the corresponding log spectrums of patches in the first three images in the same row.

of size 100×100 from three natural images [11]. Two patches in each natural image are cut from the same background, and the rest patches are three objects. The difference between background and object is that in most natural images, the background is regular and mainly contains spectral components in specific frequency region with corresponding color, whereas the object usually corresponds to a broader frequency region. The results reported in Tab. 1 indicate that the similar symbols and patches with the same background are of higher NSEP. For example, as shown in the last row of Fig. 3, the energy distribution of patches a, b resembles each other in the low frequency region, and the energy of the "dog" image patch is distributed in whole frequency domain.

As we know, the phase spectrum embodies the information about the *location* and the corresponding *proportion* of the energy to be assigned [16]. We perform the spectrum normalization by preserving the phase information of source image, and setting the spectrum amplitudes at all frequencies to a constant $\mathcal{K}$. Then we achieve an energy reallocation map by reversing the normalized spectrum to the spatial domain. Given an

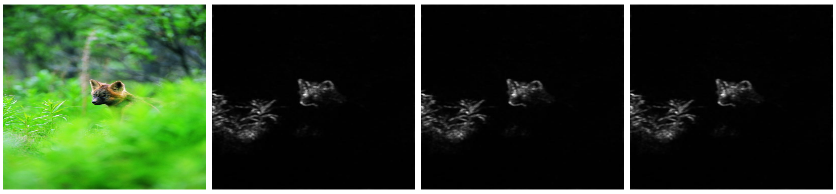

Fig. 4. Saliency maps with different $\mathcal{K}$. From left to right: a input image used in [11], corresponding saliency maps with $\mathcal{K}$=1, 100, 10000.

image $\boldsymbol{I}(x, y)$, more specifically, the procedure of spectrum normalization is formulated as follows:

$$\begin{aligned} \boldsymbol{A}(u, v)e^{-j\phi(u,v)} &= \mathcal{F}[\boldsymbol{I}(x, y)] \\ \boldsymbol{M}(x, y) &= \mathcal{F}^{-1}[\mathcal{K} \cdot e^{-j\phi(u,v)}] \end{aligned} \tag{5}$$

where $\mathcal{F}^{-1}$ denotes an inverse Fourier transformation. $e^{-j\phi(u,v)}$ describes the phase spectrum of the image. The value of the energy re-allocation map at location (x, y), $\boldsymbol{M}(x, y)$, is calculated based on Eq. (5). The value of $\mathcal{K}$ does not affect the result because the comparable energy is utilized in the saliency detection process, which is demonstrated in Fig. 4. Therefore, we set $\mathcal{K}$ as 1 in the rest of this paper.

The essence of the spectrum normalization is to normalize the energy at the whole frequency domain. By keeping the phase information unchanged and forcing the energy located at all frequencies to a constant, the frequency components of higher energy in a source image will be relatively weakened. In the source images, these frequency components correspond to the similar parts, e.g., similar psychological symbols and background. Consequently, the salient objects will be “pop-out” as the re-allocated energy of the objects is higher than those of the similar parts.

We take the lines with different orientations (symbol c and symbol d in Fig. 2) as elements of a psychological test image. Let distractors (symbol d) be of equal energy distribution, and N_d be the number of distractors d shown in Fig. 5. For conceptual simplicity, we assume there is no spectral leakage at the background. The total energy E_{total} is calculated as the sum of energy of each symbol c and d:

$$E_{total} = \sum_{u,v}(E_c(u, v) + N_d E_d(u, v)) \tag{6}$$

The exclusive energy distributions of the each symbol c or d at frequency (u, v), $W_c(u, v)$ and $W_d(u, v)$, are formulated as:

$$\begin{aligned} W_c(u, v) &= E_c(u, v) - J_{cd}(u, v) \\ W_d(u, v) &= E_d(u, v) - J_{cd}(u, v) \end{aligned} \tag{7}$$

where the commonly owned part of the energy distribution at frequency (u, v), $J_{cd}(u, v)$, is computed based on Eq. (4).

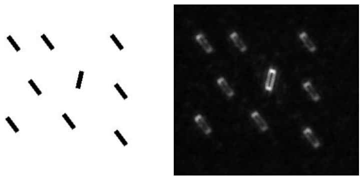

Fig. 5. A psychological test image and its energy re-allocation map

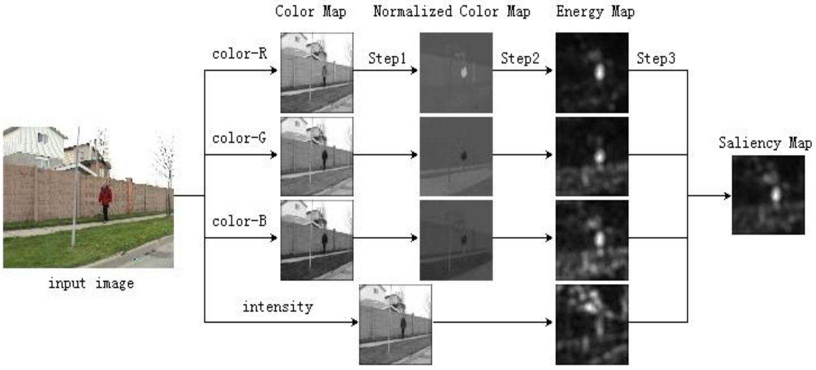

Fig. 6. The framework of ER model includes four channels, which are R, G, B and intensity channel. Step 1 is the color energy normalization demonstrated in Eq. (10). Step 2 is spectrum normalization demonstrated in Eq. (11), and Step 3 is the channel energy normalization and combination in Eq. (12) and Eq. (13).

By spectrum normalization, the normalized energy will be re-assigned to all their respective symbols. The exclusive part of energies, $W_c(u, v)$ and $W_d(u, v)$, is respectively kept by a single object b and shared by all distractors d whose number is N_d. The re-allocated energies at each object c and distractor d are equal to:

$$E_c^* = \sum_{u,v}(W_c(u,v) + \frac{J_{cd}(u,v)}{1+N_d}) \tag{8}$$

$$E_d^* = \sum_{u,v}(\frac{W_d(u,v)}{N_d} + \frac{J_{cd}(u,v)}{1+N_d}) \tag{9}$$

It is obvious that since the normalized exclusive energy of distractors d, $W_d(u, v)$, is shared by all the symbols d, the energy at each symbol d is weakened and thus symbol c becomes salient as illustrated in Fig. 5.

As for most natural images, the energy distribution of the regular background is concentrated in specific frequency region with specific color. With spectrum normalization, the normalized energy of background in particular frequency region will be evenly allocated into the whole image. Meanwhile, more normalized energy will be concentrated on the object since the object contains spectral components in a broader frequency region.

3 Energy Re-allocation Model

In our model, we utilize spectrum normalization, which is shared by `SR` [11] and `PFT` [12], to achieve saliency detection. Since color information can be utilized as an useful clue for saliency detection, color energy normalization and channel energy normalization are also used in our model to normalize energy in each color channel. For better understanding, the `ER` model is illustrated in Fig. 6.

Firstly, in order to distinguish the energy in color channels and intensity channel, we normalize the energies in RGB color channels to get normalized color maps - $\boldsymbol{\mathcal{I}}_{R,G,B}(x,y)$, which can be formulated as follows:

$$\boldsymbol{\mathcal{I}}_i(x,y) = \frac{\boldsymbol{O}_i(x,y)}{\boldsymbol{O}_I(x,y)}, \quad i = R, G, B \tag{10}$$

where $\boldsymbol{O}_*(x,y)(* = R, G, B)$ denotes the original color maps in the R,G,B channels, respectively. $\boldsymbol{O}_I(x,y)$ is the corresponding intensity map. Let spectrum normalization shown in Eq. (5) be $\mathcal{N}$, the corresponding energy re-allocation maps in their respective channels are represented as follows:

$$\begin{aligned} \boldsymbol{M}_I(x,y) &= \mathcal{N}(\boldsymbol{O}_I(x,y)) \\ \boldsymbol{M}_i(x,y) &= \mathcal{N}(\boldsymbol{\mathcal{I}}_i(x,y)), \quad i = R, G, B \end{aligned} \tag{11}$$

Note that the salient object with a strong response in one channel should not be masked by noise or by less-saliency object with normal responses in several channels. We thus normalize the energy of each energy re-allocation map to get the comparable energy instead of the prior energy amount. Different from the previous normalization mechanism [3] which sets the energy value to a fixed range, we normalize the energy by setting the total energy in each channel to be the same:

$$\mathcal{M}_i(x,y) = \frac{\boldsymbol{M}_i(x,y)}{\sum_{x,y} \boldsymbol{M}_i(x,y)}, \quad i = R, G, B, I \tag{12}$$

Finally, we combine the normalized energy re-allocation maps shown in Eq. 12 into a single energy map.

$$\boldsymbol{S}(x,y) = \sum_{i=R,G,B,I} \mathcal{M}_i(x,y) \tag{13}$$

4 Experiments

To evaluate the performance of `ER`, we carry out experiments on both natural images and psychological images. We also compare `ER` with five state-of-the-art models, i.e., `SR` [11], `PFT` [12], `AC` [10], `AIM` [4], `STB` [3]. Among these methods, `ER`, `SR` and `PFT` are performed in the spectral domain.

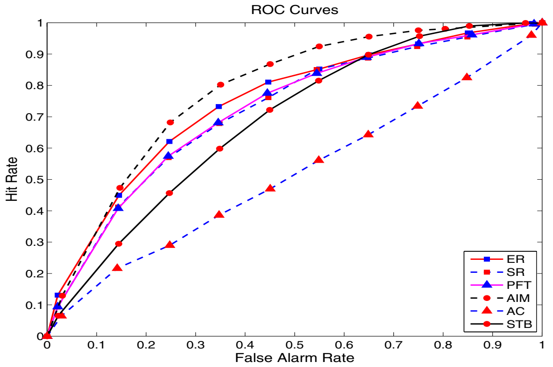

Fig. 7. ROC Curves for ER and five state-of-the-art methods

4.1 Natural Images and Psychological Patterns

We use the images and eye fixation data collected by Bruce et al. [4], in which 120 natural images are included, as a test set to evaluate the performance of these saliency detection methods. The low-pass filter is an average one of size 3 × 3. In ER, SR and PFT, the saliency map's resolutions are set to be 64 × 64 in all experiments. For SR and PFT, we extract the saliency maps in RGB channels separately, then combine them to obtain the final saliency map. For AC, AIM and STB, the default parameter settings are employed in all experiments. Some qualitative comparison results with eye fixation density maps and saliency maps are shown in Fig. 8. We also perform quantitative experiments to test the performance of these saliency methods. The results of their ROC curves and corresponding AUC values are shown in Fig. 7 and Table. 2. Among these methods, AIM obtains the best ROC results and AC has the worst performance on the eye fixation test. A possible reason is that AC's main goal is to segment objects well and thus sacrifices its saliency performance to eye fixation problem. Furthermore, compared with other state-of-the-art methods in spectral domain, ER shows more similar results with the eye fixation data, and detects the salient region better.

An important way to verify the effectiveness of a saliency detection method is to justify whether its resultant maps are consistent with psychological patterns shown in Fig. 9. Since AIM has a specific saliency mechanism for psychological patterns, we only report the results of ER and other four state-of-the-art methods. The parameters are set the same as in the experiments for natural images. The results in Fig. 9 indicate that ER is effective to detect these "pop-out" psychological patterns. Another fundamental

Table 2. AUC (Areas under the curve of ROC) performances of ER and five state-of-the-art methods

Methods	AUC	Methods	AUC
ER	0.75	AC	0.59
SR	0.71	AIM	0.77
PFT	0.71	STB	0.68

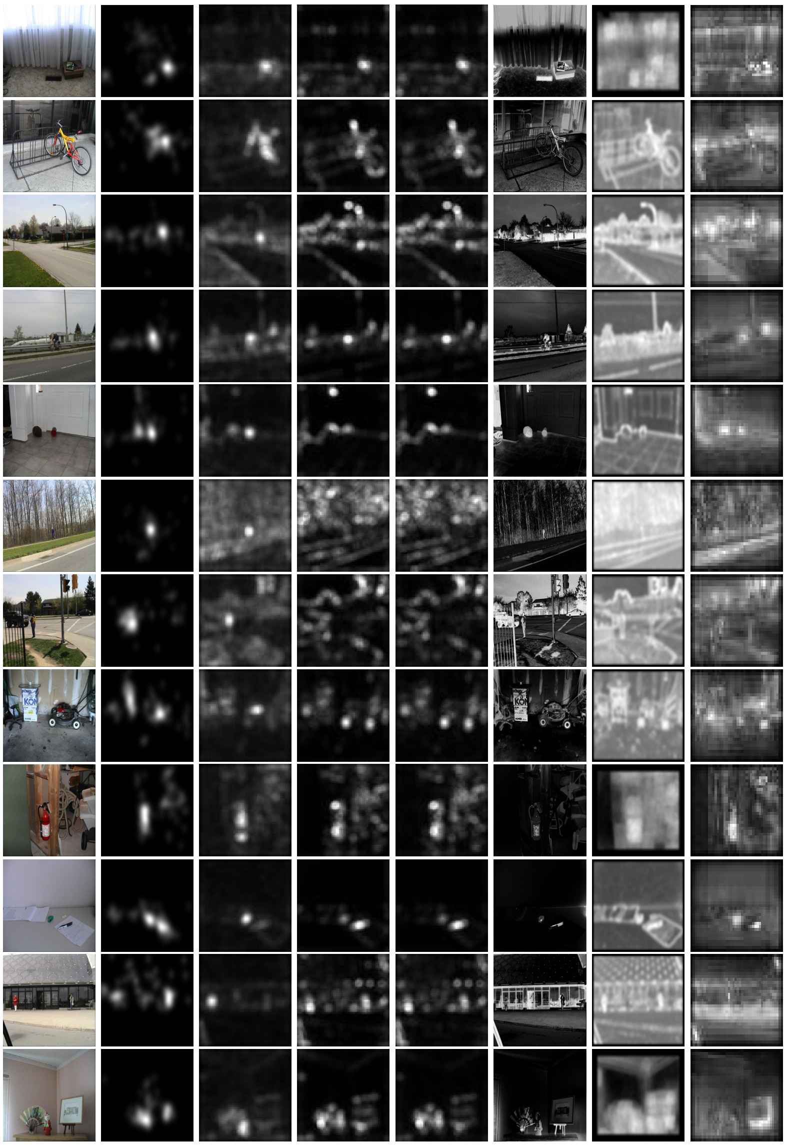

Fig. 8. Results for qualitative comparisons based on Bruce eye fixation data [4]. (From left to right: original image, eye fixation density map, saliency maps resulted from ER, SR, PFT, AC, AIM, STB.)

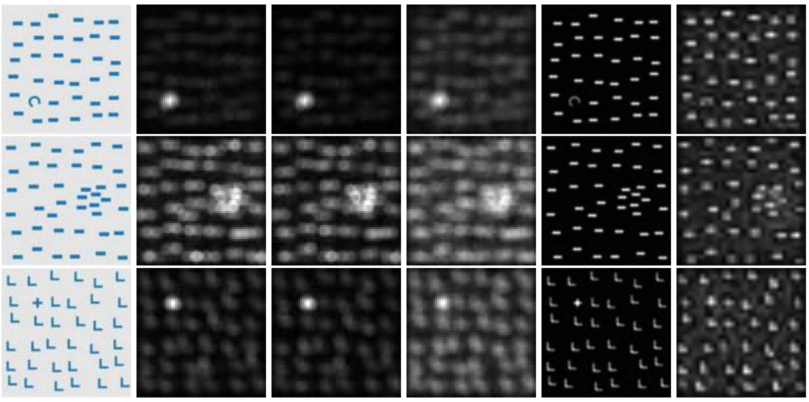

Fig. 9. Examples of saliency detection on psychological patterns. From Left to right: input image, results of SR, PFT, ER, AC and STB.

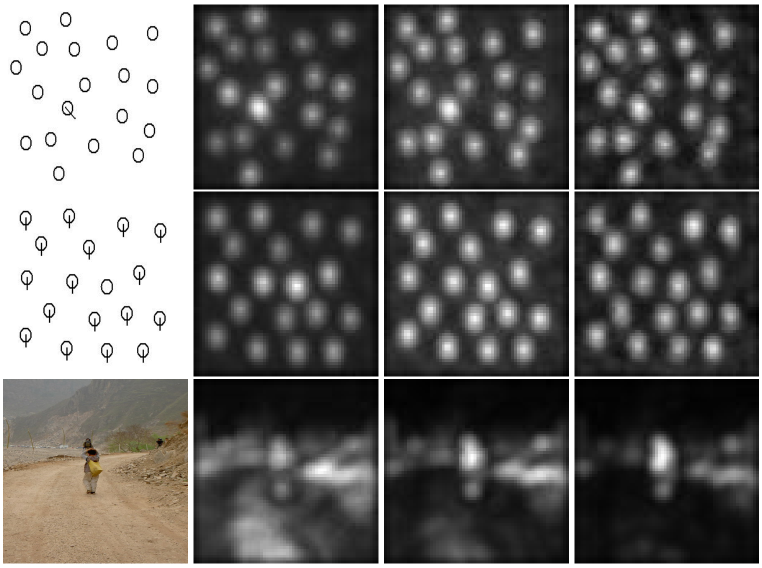

Fig. 10. Saliency maps of three groups of test images from source images at different sizes. The first two groups are two visual asymmetry images. The last one is a natural image. For each group, from left to right: 1) Original image; 2) Saliency map derived from source image with size of 256 × 256; 3) size of 128 × 128; 4) size of 64 × 64.

psychological evaluation worth mentioning is the visual asymmetrical patterns shown in Fig. 10, which means that people is more likely to focus on the object with some features absent from the distracters [17]. We will explain it in the section 4.2.

Table 3. Average time-costs of ER and five state-of-the-art methods

Methods	Average Cost-Time(s)	Standard Deviation
ER	0.0817	0.0073
SR	0.0491	0.0054
PFT	0.0445	0.0026
AC	0.6186	0.0296
AIM	1.8759	0.0115
STB	0.2846	0.0125

4.2 Time Cost and Discussion

We also compare the computational speed of these models. The software environment is based on MATLAB2009a. The computer used for the evaluation is an Intel Core 2 Duo E7500 2.93GHz and 4GB of DDR2 memory. The dataset collected by Bruce et al. [4] is large enough to be used as our test set because we compare their cost times relatively. It is not difficult to see from Table. 3 that the methods in the spectrum domain have faster speed than those methods in the spatial domain.

Here we study the influence of source image's scale on the performance of ER. One example to demonstrate the influence of image's scale is the visual asymmetries stated in the psychological experiments section. As source images' scale become larger, ER becomes consistent with visual asymmetry in which the contrast of a "Q" versus many "O"s is brighter than that of a "O" versus many "Q"s. The down-sampling step used by ER results in losing the detail of symbol “Q” so that the difference between symbols “Q” and “O” is not discernible.

For natural images, we take the image containing a road with many details shown in Fig. 10 for example. Both the child and the road will be detected when the scale is large. When the source image is downsampled to a proper scale, the details in the road are weakened and only the child will be detected. It indicates that better performance can be obtained if the scale factor can be adjusted adaptively.

5 Conclusion

In this paper, we study the intrinsic mechanism of two saliency detection methods (SR [11] and PFT [12]) in the spectrum domain. We point out that the essence of spectrum normalization is the re-allocation of energy. With spectrum normalization, the regular parts of an image are relatively weakened and the objects become prominent. We propose a saliency detection model based on Energy Re-allocation mechanism (ER). In ER, the energy is re-allocated by the combination of color energy normalization, spectrum normalization and channel energy normalization. Quantitative and qualitative experiments indicate that ER is better than other five state-of-the-art methods. In the future, we will study how to adaptively select the optimal source image's scale to further improve the quality of saliency detection.

Acknowledgements

This work was supported in part by the NFSC (No. 60975044), 973 program (No. 2010CB327900), Shanghai Leading Academic Discipline Project No. B114, and the Shanghai Committee of Science and Technology, China. Grant No. 10DZ2272800.

References

1. Hou, X., Zhang, L.: Dynamic Visual Attention: Searching for Coding Length Increments. In: NIPS, vol. 20 (2008)
2. Wang, Y.S., Tai, C.W., Sorkine, O., Lee, T.Y.: Optimized Scale-and-Stretch for Image Resizing. SIGGRAPH ASIA (2008)
3. Itti, L., Koch, C., Niebur, E.: A model of Saliency-based Visual Attention for Rapid Scene Analysis. TPAMI 20, 1254–1259 (1998)
4. Bruce, N., Tsotsos, J.: Saliency Based on Information Maximization. In: NIPS, vol. 18 (2006)
5. Gao, D., Vasconcelos, N.: Bottom-up Saliency is A Discriminant Process. In: ICCV (2007)
6. Gao, D., Vasconcelos, N.: An Experimental Comparison of Three Guiding Principles for The Detection Salient Image Locations: Stability, Complexity, and Discrimination. In: CVPR workshops (2005)
7. Lowe, D.G.: Object Recognition from Local Scale-invariant Features. In: ICCV (1999)
8. Sebe, N., Lew, M.S.: Comparing Salient Point Detectors. In: ICME (2001)
9. Kadir, T., Brady, M.I.: Scale, Saliency and Image Description. IJCV 45, 83–105 (2001)
10. Achanta, R., Hemami, S., Estrada, F., Süsstrunk, S.: Frequency-tuned Salient Region Detection. In: CVPR (2009)
11. Hou, X., Zhang, L.: Saliency Detection: A Spectral Residual Approach. In: CVPR (2007)
12. Guo, C., Ma, Q., Zhang, L.: Spatio-temporal Saliency Detection Using Phase Spectrum of Quaternion Fourier Transform. In: CVPR (2008)
13. Field, D.J.: Relations Between the Statistics of Natural Images and The Response Properties of Cortical Cells. JOSA A 4, 2379–2394 (1987)
14. Field, D.J.: What is the Goal of Sensory Coding? Source Neural Computation archive 6, 559–601 (1994)
15. He, D.-C., Wang, L.: Texture Unit, Texture Spectrum, and Texture Analysis. TGRS 28, 509–512 (1990)
16. Oppenheim, A.V., Lim, J.S.: The Importance of Phase in Signals. Proceedings of the IEEE 69, 529–541 (1981)
17. Wolfe, J.M.: Asymmetries in Visual Search: An Introduction. Perception and psychophysics 63, 381–389 (2001)