Optik xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

An optical information processing-based idea for visual attention analysis

Jufeng Zhao[a,*], Xiumin Gao[a], Guang Lin[b], Daodang Wang[c]

[a] Institute of Electronic and Information, Hangzhou Dianzi University, Hangzhou, China
[b] State Key Lab of Modern Optical Instrumentation, Zhejiang University, Hangzhou 310027, China
[c] College of Metrology and Measurement Engineering, China Jiliang University, Hangzhou 310018, China

A R T I C L E I N F O

Article history:
Received 7 October 2015
Accepted 29 December 2015
Available online xxx

Keywords:
Visual attention
Saliency extraction
Optical system
Frequency selection

A B S T R A C T

One of the challenges in visual attention analysis is to generate full-resolution saliency map in real time. Traditional methods usually design algorithms and implement in computer, and they have faced a troublesome problem in balancing the resolution and computational speed. Optical computing is considered to be a practical and efficient method, so we build an optical information processing-based system to achieve fast visual saliency extraction to generate full resolution saliency maps with fine details of boundaries. The whole system mainly consists of a Display as the input image source, optical Fourier Lens and spatial filters, an imaging sensor. The key of our method is frequency selection with two precise pinholes which treated as band-pass spatial filters. And the saliency map is finally obtained using simple calculation. According to the experimental results, and comparing with other famous approach, the results demonstrate that our system efficiently produce full resolution saliency maps with good boundaries.

1. Introduction

Human Visual System (HVS) can simply figure out the interested area from complex scene. It is determined by our great eyes and brain. These interested area we called salient targets, which would be given high attention by HVS. The saliency detection ideas are inspired by this behavior of HVS. The saliency extraction technology is widely used in image compression [1–3], image segmentation [4], target detection [5], image enhancement [6] and so on. A lot of fields have considered saliency detection now.

Many researchers make efforts to saliency detection. Inspired by neuronal architecture, Itti and his partners tried to extract several features, including intensity, colors and orientations [7]. They created these feature maps. The saliency map was finally generated by fusing these feature maps based on winner-take-all strategy. Erdem et al. use covariance matrices of simple image features as meta-features for saliency estimation [8]. And they had shown that first-order statistics of features could be easily incorporated to the proposed approach to improve the performance. A context-aware saliency extraction model was discussed to detect image areas [9]. Their method combined four basic principles, which included local low-level considerations, global considerations, high-level factors and visual organization rules. Inspired by the way people watching videos, researchers mentioned a novel method based on candidate selection for video saliency estimation [10]. By restricting those salient locations to a carefully selected candidate set, both accuracy and computation speed were greatly improved. Well considering the framework of the Markovian idea, people proposed a graph-based visual saliency extraction method [11]. This approach made use of bottom-up visual saliency model. The further research should focus on the low resolution of saliency map. A graph-based multiscale saliency-detection algorithm was designed [12]. This method is learned by modeling eye movements as a random walk on a graph. Their experimental results had shown that this method could detect visual saliency precisely and reliably, though the speed is a little slow. Hou et al. mentioned a spectral residual method for saliency extraction, which considered the prior knowledge of the objects [13]. The main idea was learned by analyzing the log-spectrum of images. The method worked fast. The detected resolution of saliency map was so low.

According to all those methods, the first main problems lie at the computational time and the resolution of saliency map. The two seems contradicting. It is difficult to quickly obtain a full-resolution saliency map. Second, those methods are image post-processing ideas. People haven't considered the whole imaging chain. They only design algorithms and do optimization using computer. No one makes use of optical system to generate saliency map. The

* Corresponding author. Tel.: +86 571 86919029; fax: +86 571 86919029.
E-mail address: dabaozjf@hdu.edu.cn (J. Zhao).

http://dx.doi.org/10.1016/j.ijleo.2015.12.132
0030-4026/© 2016 Published by Elsevier GmbH.

two problems inspire us to develop a new approach for saliency extraction.

We are eager to build an optical information processing-based system to help realizing the visual attention analysis. We have made this true in Ref. [14], in which photographic film is needed as the input image, which make this difficult to apply in future. Here, we expect to use a display as the input source, which could be treated as natural scene. And the whole system is improved to the further application in real cases. With this design, the system is more competitive in visual attention analysis for the fast speed of optical computing. Using optical computing, our system could fast get saliency map with full resolution. The key point is the frequency selection, and we make use of appropriate band-pass filter to produce saliency map. Both simulation and experimental results demonstrate that our approach can generate full resolution saliency map with good boundaries in real-time.

2. Method

2.1. Frequency selection-based visual attention detection

In image visual attention analysis, researchers usually follow four rules [14]:

1. The large objects should be highlighted uniformly.
2. The boundaries of saliency map are supposed to be generated as well as possible.
3. Full resolution saliency map is desired.
4. The speed should be as fast as possible.

According to our previous analysis [14], the four rules could be satisfied with the frequency-selection idea. Here, we continue to create visual saliency map by extracting the frequency components from original scene.

By frequency selection idea, input the original image O, the visual saliency map S is determined by the following equation

$$S = \left| O_{\text{low}} - O_{\text{high}} \right| \tag{1}$$

O_{low} and O_{high} are two filtered results from image O. If f represents frequency, O_{low} is obtained by low frequency filter F_{low} using low cut-off frequency value f_{low},

$$F_{\text{low}}(f_{\text{low}}) = \begin{cases} 1 & f < f_{\text{low}} \\ 0 & f \geq f_{\text{low}} \end{cases} \tag{2}$$

Meanwhile, O_{high} is filtered from high frequency filter F_{high} using high cut-off frequency value f_{high},

$$F_{\text{high}}\left(f_{\text{high}}\right) = \begin{cases} 1 & f < f_{\text{high}} \\ 0 & f \geq f_{\text{high}} \end{cases} \tag{3}$$

Analyzed from the above description, the visual saliency map S could be rewritten as follows

$$S = \left| O_b \right| \tag{4}$$

where O_b could be called band-pass filtered result. In the above equations, $f_{\text{high}} > f_{\text{low}}$. Thus, S extracts the frequency components between f_{low} and f_{high} from O.

2.2. Optical information processing-based frequency selection

We are eager to build an optical system to realize the theory mentioned in this section. We have made this true in Ref. [14], while photographic film is needed as the input image, which make this difficult to apply in future. Here, we expect to use a display as the input source, which could be treated as natural scene. And the whole system is improved to the further application in real cases. With this design, the system is more competitive in visual attention analysis for the fast speed of optical computing.

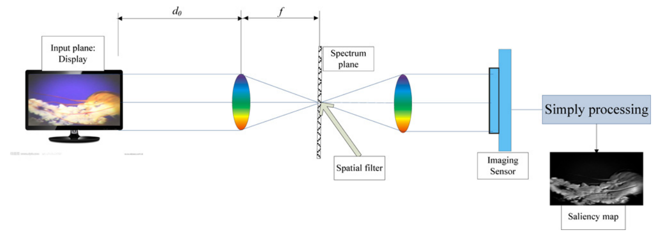

Fig. 1. The sketch of optical information processing-based system for saliency extraction.

The simple optical imaging set-up is shown in Fig. 1. In input plane, we set a display, which could give arbitrary image source to simulate natural scene. The following structure is reformed from 4F optical system with the length of f. The distance between the Display and the first Fourier lens is d_0, and $d_0 \gg f$. Since $d_0 \gg f$, we consider that the incident light of first lens is parallel the optical axis. With help of this lens, the image is transformed into spectrum. And we use spatial filter in spectrum plane to realize frequency selection. This spatial filter is a band-pass one, which has two parameters named radius R_1 and radius R_2, $R_1 < R_2$. Though the spectrum at this plane is different from that of the 4f system because of the phase curvature, but it doesn't affect our frequency selection. Our frequency selection only picks up frequency values, only changing amplitude not phase. With different radiuses of R_1 and R_2, different frequency components would be extracted. After finishing frequency selection, the spectrum is transformed by the second Fourier Lens. The image light finally comes to the image sensors, which produce a filtered image. And then, the visual saliency map is simply computed using Eq. (4).

2.3. Theoretical analysis for optical system

In Section 2.2, we have mentioned that the spectrum at Spectrum plane is different from that of the 4f system. It isn't a strictly Fourier transform because of the phase curvature. And it doesn't affect our frequency selection, as our operation only changing amplitude not phase. And next, we will explain this as follows.

For a 4f system, the object plane is the front focal plane. The image source O could generate a spectrum at the spectrum plane

$$U_f = \beta \cdot \text{FT}(O) \tag{5}$$

where β is a constant, which is determined by light wavelength, focal length and so on. In Eq. (5), FT(O) represents the Fourier transform of O.

And in our new system, The distance between the Display and the first Fourier lens is d_0. Then, the complex amplitude of spectrum could be expressed as

$$U'_f = \beta \cdot \exp\left[j\frac{k}{2f}\left(1 - \frac{d_0}{f}\right)\left(x^2 + y^2\right)\right] FT(O)$$

$$= \exp\left[j\frac{k}{2f}\left(x^2 + y^2\right)\right] U_f \tag{6}$$

where k is wavenumber, x and y represent coordinates at spectrum plane. Comparing these two equations, one can conclude that there only exist a phase difference between our system and 4f system. Since our frequency selection doesn't affect phase, we could realize our frequency selection with this system.

Please cite this article in press as: J. Zhao, et al., An optical information processing-based idea for visual attention analysis, Optik - Int. J. Light Electron Opt. (2015), http://dx.doi.org/10.1016/j.ijleo.2015.12.132