LETTER
Segmentation of Depth-of-Field Images Based on the Response of ICA Filters

André CAVALCANTE†, Allan Kardec BARROS††, Nonmembers, Yoshinori TAKEUCHI†, and Noboru OHNISHI†, Members

SUMMARY In this letter, a new approach to segment depth-of-field (DoF) images is proposed. The methodology is based on a two-stage model of visual neuron. The first stage is a retinal filtering by means of luminance normalizing non-linearity. The second stage is a V1-like filtering using filters estimated by independent component analysis (ICA). Segmented image is generated by the response activity of the neuron measured in terms of kurtosis. Results demonstrate that the model can discriminate image parts in different levels of depth-of-field. Comparison with other methodologies and limitations of the proposed methodology are also presented.

key words: image segmentation, depth-of-field, kurtosis

1. Introduction

Depth-of-field is a photographic technique which allows us to control the image sharpness of different areas of a visual scene. Segmentation of focused objects is generally achieved by firstly computing a map which highlights structural complexity in the image [1], [2]. In referenced works, this map is then post-processed or integrated with other clues such as color information to improve results.

In this respect, there are two important issues. The first issue is the degree of noise of the map computed at the first stage. The second one is the smoothness of the map at highlighting structured and structureless image parts without suffering from over-segmentation. The proposed methodology is able to build a structural complexity map which fairly address these issues. Specifically, the proposed map reflects the structural complexity in the image based on the response activity of a simulated vision neuron.

2. Methods

The proposed methodology is based on the two-stage model of visual neuron represented in Fig. 1. The map proposed in this letter to describe structural complexity is constructed upon the response activity of the neuron. The following subsections describe each part of the methodology in detail.

Luminance normalization. The input image is firstly divided in neighborhoods of $N \times N$ pixels around each pixel $I(i,j)$. Each neighborhood is denoted by a column vector $x_{ij}$. This vector is formed by reading pixels intensities in a column-wise fashion, i.e., from top to down and left to right in the neighborhood. For each neighborhood, the logarithm of the pixel intensities is taken, i.e.,

$$x_{ij}' = \log^* (x_{ij}),$$

(1)
where $\log'(a) = 0$ for $a < 1$ and $\log'(a) = \log(a)$ for $a \geq 1$. This non-linear transformation reduces large differences between luminance intensities in different parts of the image. This effect can be observed in Fig. 2.

**ICA filters.** The FastICA algorithm [3] is used to learn a set of filters represented by the column vectors $w_1, w_2, \ldots, w_n$. The individual responses of filters $w_k$ to the log transformed neighborhood $x_{ij}'$ are calculated as

$$y_{ij}' = w_k^T \cdot x_{ij}'$$

where $T$ is the transpose operator. Here, filters are learned from an ensemble of natural images. Details about training data and learning parameters are given in the section Results.

**Activity map.** The activity map $A$ is computed based on the kurtosis of the response vectors $y_{ij} = (y_{ij}, y_{ij}, \ldots, y_{ij})$, i.e.,

$$A(i, j) = \frac{1}{n} \cdot \sum_{k=1}^{n} (\bar{y}_{ij}' - \bar{y}_j')^4$$

$$\frac{1}{n} \cdot \sum_{k=1}^{n} (\bar{y}_{ij}' - \bar{y}_j')^2$$

where $\bar{y}_j' = \frac{1}{n} \cdot \sum_{k=1}^{n} y_{ij}'$.

Once the activity map $A$ has been calculated, morphological reconstruction [4] is used to fill holes inside the map. This process is given as follows:

1. **Define**

   $$J_0(i, j) = \begin{cases} A^C(i, j) & \text{if } (i, j) \text{ is on the map borders}, \\ M & \text{otherwise}, \end{cases}$$

   where $A^C(i, j) = -A(i, j)$ and $M < A^C(i, j) \forall i, j$. Here, it is used $M = \min\{A^C(i, j)\} - 1$.

2. **While** $J_{z+1} \neq J_z$ **do**

   $$J_{z+1}(i, j) = \min[[J_z \oplus B](i, j), A^C(i, j)],$$

   where $z$ represents the iteration index, $[J \oplus B](i, j) = \max\{J(i - i', j - j')\}$ is the dilation of matrix $J$ by a flat structuring element $B$. Here, it is used a $3 \times 3$ flat element so that $N_B$ is an 8-connectivity neighborhood.

   **end of while**

3. **Return** $J_{z+1}$.

   After this procedure is executed, the desired map with holes filled is given by $J_{z+1}$.

![Fig. 2](image)

**Fig. 2** Luminance normalization. (Left) Original image. (Right) log transformed image.

3. **Results**

The natural scenes dataset was obtained from the McGill Calibrated Colour Image Database [5]. This database consists of TIFF formatted non-compressed images. The resolution of image files is $576 \times 768$ pixels. In order to build a dataset as general as possible, 150 scenes were selected from natural image categories such as forests, landscapes, and natural objects such as animals. Images containing man-made objects were not used.

In order to estimate the filter population, 150,000 image patches of $15 \times 15$ pixels were extracted in a non-overlapping fashion from our dataset. This set of image patches were used as the learning input for the FastICA algorithm. For this algorithm, the number of iterations was set to 200 and the working non-linearity was the hyperbolic tangent. Dimension reduction have not been used here since it is not known exactly how this process relates to neural adaptation. A total of 224 filters and a DC component were learned. Examples of the learned FastICA filters used in this analysis are shown in Fig. 3 A. In order to provide a quantitative analysis of these filters, they were fitted by real-valued Gabor functions. The preferred parameters of these functions are presented in Fig. 3 B. In the polar plot, each filter is represented by a gray-colored circle whose orientation and distance from the plot origin represents respectively the preferred spatial orientation (degrees) and spatial frequency of the filter (cycles/pixel). The gray intensity of each circle represents the filter’s half-amplitude spatial frequency bandwidth (octaves). The associated graymap along with the histogram of bandwidth values are given in below.

![Fig. 3](image)

**Fig. 3** Characteristics of the population of ICA filters. (A) Examples of filters were learning using the FastICA algorithm. (B) Filter parameters are resumed in the polar plot and the bandwidth histogram. Each gray-colored circle in the polar plot represents one of the 224 filters learned by the FastICA. The distance of the circle from the origin represents the preferred spatial frequency of the filter and it is given in cycles/pixel. The orientation of the circle represents the spatial orientation of the filter and it is given in degrees. The gray intensity of the circle represents the half-amplitude spatial frequency bandwidth. The related graymap along with the histogram of bandwidth values is show below the polar plot.
the polar plot.

The polar plot shows that the majority of FastICA filters are located on the highest frequency part of the Fourier spectrum. The filters are not uniformly distributed along spatial orientations which are visibly more concentrated at oblique orientations. The histogram of half-amplitude spatial frequency shows that the majority of filters have bandwidths between 0.2 and 0.5 octaves.

**Segmentation results.** We have compared the performance of the proposed method with other two methods. The first method is the higher-order-statistics map (HOS) used by the researches [1] and [2]. This map is calculated based on the fourth-order moment of the neighborhoods $x_{ij}$, i.e.,

$$m^4(i, j) = \frac{1}{N^2} \cdot \sum_{k=1}^{N^2} (x_{ik} - \bar{x})^4,$$

(5)
where $x^{ij}_k$ represent the intensities of the pixels inside of the neighborhood $x_{ij}$ and $\bar{x}^{ij} = \frac{1}{N^2} \sum_{k=1}^{N^2} x^{ij}_k$. The HOS map is finally calculated as

$$HOS(i, j) = \min \left\{ 255, \frac{m^4(i, j)}{DSF} \right\},$$

where DSF is a down scaling factor found by heuristics (for these images, DSF= 4000). The second method is the map generated by calculating the power of the response vector $y^{ij}$, i.e., $P(i, j) = \sum_{k=1}^{n} (y^{ij}_k)^2$.

Figure 4 shows three figures with different characteristics of depth-of-field. These images were not included in the learning set of FastICA algorithm. Images were converted to grayscale. Images were processed in neighborhoods of $15 \times 15$ pixels. The first row in Fig. 4 shows the original test images. The second row shows the HOS maps. The third row, the map $P$ generated by calculating the response power. The last row shows results of the proposed methodology.

In Fig. 4, it is easy to see that the figure on the left column has three main areas in different degrees of focus. The segmentation result of the HOS map from [1] and [2] is noisy providing only an almost binary characterization of the scene. The map $P$ represents the energy at the frequency bands of the ICA filters and seems to only discriminate what is most in focus. The proposed method which uses the kurtosis of the response provides a smooth representation of the main three image parts. The image on the center column is simpler in which there is only one object and a strongly blurred background. In the image on the right, only the bottom part of the image is in perfect focus. The activity map proposed here is the only one which generates a smooth description of this blur effect.

In Fig. 5, the effect of morphological reconstruction for the proposed methodology is analyzed. On the left we have the original test image. This image has two highlighted issues. The first issue is that the image is corrupted by camera noise during the acquisition process. This is noticed by the spike in the upper part of the image. The second issue is the small object (“bird”) which is in focus but surrounded by less structured areas. The center and right images are generated using the proposed model without and with morphological reconstruction respectively.

This process has successfully removed small artifacts generated by camera noise. However, it has also removed small objects in focus which were surrounded by structure-less areas. This is not desirable since the size of objects in visual scenes can be largely influenced by the distance of the camera.

4. Conclusion

The goal of this work is to provide an improved map for applications related to segmentation of depth-of-field images. Here, additional information such as color data [2] and further post processing techniques such as region merging [1] are not used. The results demonstrate, however, that the proposed method is extremely competitive generating smooth and noiseless segmented images.

References

[1] C. Kim, “Segmenting a low-depth-of-field image using morphological filters and region merging,” IEEE Trans. Image Process., vol.14, no.10, pp.1503–1511, 2005.
[2] C. Kim, J. Park, J. Lee, and J.N. Hwang, “Fast extraction of objects of interest from images with low depth of field,” ETRI Journal, vol.29, no.353-362, 2007.
[3] A. Hyvärinen, “Fast and robust fixed-point algorithms for independent component analysis,” IEEE Trans. Neural Netw., vol.10, no.3, pp.626–634, 1999.
[4] L. Vincent, “Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms,” IEEE Trans. Image Process., vol.2, no.2, pp.176–201, 1993.
[5] A. Olmos and F.A.A. Kingdom, “Megaill calibrated colour image database,” http://tabby.vision.mcgill.ca, 2004.