ICIC Express Letters
Part B: Applications
Volume **7**, Number **2**, February **2016**

ICIC International ©2016 ISSN 2185-2766
pp. 375–380

2D PRINCIPLE COMPONENT ANALYSIS BASED NOVEL MULTISCALE SALIENCY DETECTION METHOD

Shangbing Gao[1,2], Jingbo Zhou[1], Youdong Zhang[1]
Jun Zhou[2] and Jianxun Xue[1]

[1]Jiangsu Provincial Key Laboratory for Advanced Manufacturing Technology
Faculty of Computer Engineering
[2]The Key Laboratory for Traffic and Transportation Security of Jiangsu Province
Huaiyin Institute of Technology
No. 1, Meicheng East Road, Huai'an 223003, P. R. China
luxiaofen_2002@126.com

Received August 2015; accepted November 2015

Abstract. *In this paper, we propose a new multiscale saliency detection algorithm based on 2D principal component analysis. To measure saliency of pixels in a given image, we first segment the image into patches by a fixed scale and then use the principal component analysis to reduce the dimensions which throw out dimensions that are noises with respect to the saliency calculation. The contrast between a patch and other patches is computed based on the absolute contrast, bin contrast and the local contrast. Finally, we implement our algorithm through multiple scales that further decrease the saliency of background. Experimental results show that our method performs well in the benchmark dataset.*
Keywords: Saliency detection, Multiscale, 2D principle component analysis, Patch contrast

1. **Introduction.** Visual attention analysis has generally progressed on two fronts: bottom-up and top-down approaches. Bottom-up approach, which is data-driven and task-independent, is a perception processing for automatic salient region selection for images. There exist several computational models [1-8] for simulating human visual attention based on the bottom-up approaches.

In this paper, we also focus on the bottom-up approaches. In our proposed model, we first divide the input image into small image patches. In essence, our proposed algorithm first divides images into small image patches. Then it uses 2D principal component analysis (2DPCA [9]) to reduce the dimensionality of each patch. We exploit the absolute contrast, bin contrast and the local contrast by considering the differences between this patch and all other patches in the image. Furthermore, for a salient object, the part near the center of the input image is more salient than that far away from the center. To diminish this effect, we exploit the multiple scales instead of the central bias to decrease the saliency of background patches and improve the contrast between salient and non-salient regions.

The rest of this paper is organized as follows. In Section 2, we present the proposed saliency detection model based on 2DPCA. The experimental results are shown in Section 3. We end this paper by the conclusions in Section 4.

2. **Proposed Saliency Algorithm.** In this section, we will state the framework of our saliency detection method in detail. The steps of our algorithm are fourfold: representing the image patches, using 2DPCA to reduce dimensionality, computing each patch's saliency value and implementing our method to multiple scales. We will describe the details step by step in the following subsections.

2.1. **Image patches representation.** The first step of our algorithm is to divide each original input image into small image patches to gather local information. For simplicity, we take image patches from the original images without overlapping. Given an image A with dimension $H \times W$, non-overlapping patches with the size of $n \times n$ pixels are drawn from it. Generally speaking, the size of the patches located in the bottom and right boundary is smaller than the regular size. To make sure all patches have the same dimensions for feature extraction by PCA, we throw out the border regions for simplicity which do not have regular size. The total number of patches is $N = \lfloor H/n \rfloor \cdot \lfloor W/n \rfloor$. Denote the patch as A_i, $i = 1, 2, \ldots, N$. Then each patch is represented as a column vector x_i of pixel values. The length of the vector is $3n^2$ since the color space has three components. Finally, we get a sample matrix $X = [x_1, x_2, \ldots, x_N]$, where N is the total number of patches as stated above.

2.2. **2DPCA feature extraction.** Since saliency detection can be regarded as classification process of the image blocks, the classification result is salient patch and non-salient patch. While PCA method is mainly used for dimensionality reduction of high-dimensional data, and then in the low-dimensional subspace after projection, we get maximum between-class variance and minimum within-class variance. So we use PCA to select the appropriate subspace, and the within-class variance of the image blocks in the salient region and non-salient region is respectively smaller while the between-class distance is longer between the image blocks in the salient region and those in the non-salient region. Thus, we choose 2DPCA to compute significantly useful features and remove unwanted features, and then the contrast and distribution values of each patch are calculated in a low-dimensional subspace. The average of all N patches for image A is

$$\overline{A} = \frac{1}{N} \sum_{k=1}^{N} A_k \tag{1}$$

Total scatter matrix of image patches for row direction is:

$$G = \frac{1}{N} \sum_{k=1}^{N} \left(A_k - \overline{A}\right)^T \left(A_k - \overline{A}\right) \tag{2}$$

After solving the eigenvalue and eigenvector of G, on the basis of the contribution of principal component $\theta = \sum_{i=1}^{d} \lambda_i \Big/ \sum_{i=1}^{n} \lambda_i \geq 0.9$, we select the eigenvectors X_1, X_2, X_3, $\ldots$, X_d corresponding to the first d biggest eigenvalues λ_1, λ_2, λ_3, $\ldots$, λ_d to construct the optimal projection matrix X_{opt}, where $X_{opt} = [X_1, X_2, X_3, \ldots, X_d]$.

Total scatter matrix of image patches for column direction is:

$$G = \frac{1}{N} \sum_{k=1}^{N} \left(A_k - \overline{A}\right) \left(A_k - \overline{A}\right)^T \tag{3}$$

We get the optimal projection matrix $Y_{opt} = [Y_1, Y_2, Y_3, \ldots, Y_q]$ from the eigenvectors Y_1, Y_2, Y_3, $\ldots$, Y_q corresponding to the first q biggest eigenvalues γ_1, A_i, γ_3, $\ldots$, γ_q.

We directly project the image matrix A_i $(m \times n)$ onto both X_{opt} $(n \times d)$ and Y_{opt} $(m \times q)$ to obtain the recognition matrix Z_i $(q \times d)$:

$$Z_i = Y_{opt}^T A_i X_{opt} \tag{4}$$

2.3. **Patch contrast model.** Rareness reflects the property of distinctiveness that few pixels must be visually unique to inspire human visual concentration. It can be easily discovered from global and local contrasts. Absolute contrast and bin contrast are to cooperate in the proposed global contrast.

Absolute contrast preserves the color distinctiveness by measuring the L2-norm distance between the given patch A_i and the averaged patch from the image. The averaged patch is defined as $mean(A) = \frac{1}{N}\sum_{i=1}^{N} Z_i$. Thus, absolute contrast is defined as:

$$AC(A_i) = \|Z_i - mean(A)\| \tag{5}$$

In order to emphasize the uniqueness of the rare-color bins, their saliency should be larger than the saliency from common-color bins. Thus, we evaluate a bin's frequency to reflect its saliency. In detail, we separate the N patches into B clusters, $K_1, K_2, \ldots, K_B$, via Normalized Cut [10], and $|K_a|$ indicates its population where $1 \leq a \leq B$. The bin contrast for a patch A_i is defined as:

$$BC(A_i) = \exp\left(\frac{|K_a|}{\sum_{b=1}^{B}|K_b|}\right) \tag{6}$$

These two contrasts can preserve global saliency well, but they still cannot emphasize local contrasts owing to misunderstanding spatial property. Local contrast is thus considered further in the following.

The image block in the image itself does not determine its saliency, and its saliency is determined by the difference between its feature and those of the image blocks in the adjacent area. The more obvious the difference is, the more likely it is that it can cause human vision attention, and the contrast will be more distinct if the distance between the patch and the neighborhood is smaller. So we define the contrast model for the patch i in an image as a weighted sum of feature differences from other patches:

$$LC(A_i) = \sum_{j=1}^{N} w(x_i, x_j) \cdot d(Z_i, Z_j) \tag{7}$$

where $d(Z_i, Z_j) = \|Z_i - Z_j\|^2$ is the distance between the feature Z_i of patch i and Z_j of patch j, the Gaussian spatial weight is defined as $w(x_i, x_j) = (1/Z_x)\exp\left(\left(-\|x_i - x_j\|^2\right)/\sigma_x^2\right)$, and Z_x is the normalization factor. x_i and x_j are respectively the central positions of patch i and patch j.

The contrasts from local and global visual clues are integrated. We multiply AC, BC, and LC together for fusing them:

$$MC(A_i) = AC(A_i) \times BC(A_i) \times LC(A_i) \tag{8}$$

2.4. **Multiple scales extension.** Based on the observation that patches in background are likely to have similar patches at multiple scales, which is in contrast to more salient patches that could have similar patches at a few scales but not at all of them (salient object always smaller than the background), therefore, we wish to incorporate multiple scales to further decrease the saliency of background patches, improving the contrast between salient and non-salient regions.

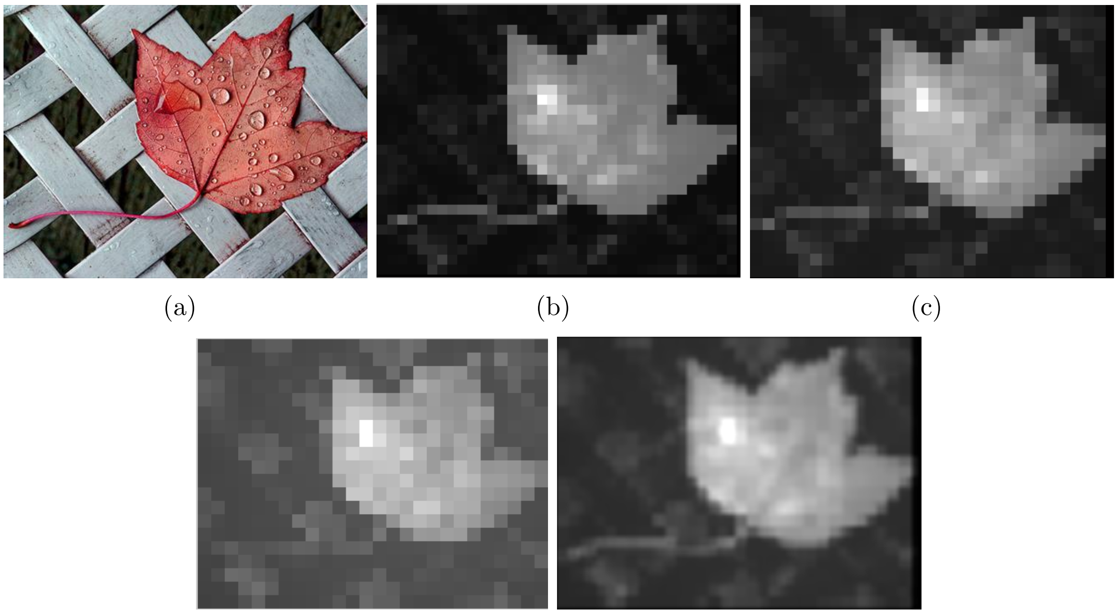
For a patch p_i of scale r, the saliency value according to Equation (8) is defined as:

$$S_i^r = 1 - \exp\left\{-\frac{1}{L}\sum_{k=1}^{L} MC(A_i^k)\right\} \tag{9}$$

Considering the scales $R_c = \{r_1, r_2, \ldots, r_M\}$, we use Equation (9) to calculate the saliency of patch i as $\{S_i^{r_1}, S_i^{r_2}, \ldots, S_i^{r_M}\}$. The final saliency is computed as:

$$S_i = \frac{1}{M}\sum_{r \in R_c} S_i^r \tag{10}$$

3. **Experiments.** We evaluate our method in two aspects: predicting human visual fixations and segmenting the salient object from natural images.



(a) (b) (c)

(d) (e)

Figure 1. The original images and its different saliency maps with different patch sizes: (a) original images, (b) saliency maps with the image patch size being 30×30; (c) saliency maps with the image patch size being 20×20; (d) saliency maps with the image patch size being 10×10; (e) final saliency maps which combines three results together

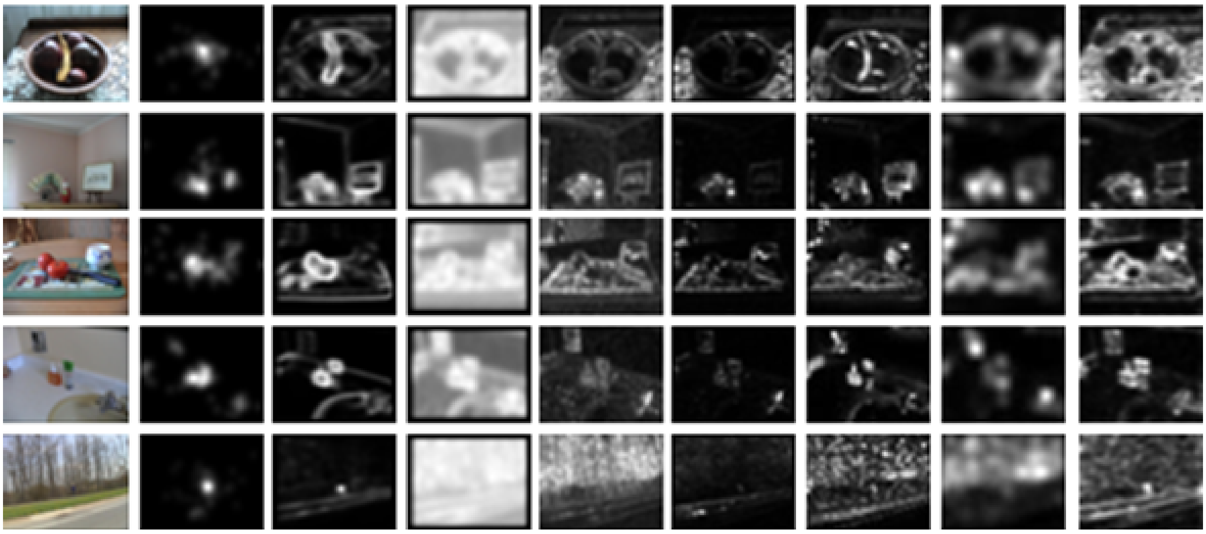
3.1. **Predicting human visual fixations.** In this subsection, we show several experimental results on detecting saliency in natural images. We used the image dataset and its fixation data collected by Bruce and Tsotsos [11] as a benchmark for comparison. This dataset contains eye fixation records from 20 subjects for a total of 120 images of size 681×511. We also computed the area receiver operating characteristic (ROC) curve [11], i.e., the area under the curve to quantitatively evaluate the algorithm performance. The final ROC area shown in Table 1 is the average value over 100 permutations. The mean and standard errors are also reported in Table 1. It is observed that our model outperforms all other methods in terms of ROC area.

Table 1. Performance in predicting human visual fixation data. SE means standard errors.

Attention Model	ROC (SE)
Itti et al. [1]	0.6146 (0.0008)
Bruce and Tsotsos [2]	0.6727 (0.0008)
SUN [3]	0.6682 (0.0008)
GBVS [4]	0.6818 (0.0007)
Duan et al. [5]	0.6837 (0.0008)
Our method	0.7042 (0.0007)

Some visual results of our algorithm are compared with the state-of-art methods in Figure 2. The comparison results show that the most salient locations on our saliency maps are more consistent with the human fixation density maps.

3.2. **Salient object segmentation.** We have evaluated the results of our approach on the publicly available database provided by Achanta et al. [6]. In order to comprehensively evaluate the accuracy of our method for salient object segmentation, we use the iterative



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Figure 2. Results on predicting human visual fixation data: (a) input images; (b) human fixations; (c) saliency map from the proposed model; (d) saliency map from Itti's model [1]; (e) saliency map from Bruce's model [2]; (f) saliency map from Zhang's model [3]; (g) saliency map from Harel's model [4]; (h) saliency map from Duan's model [5]; (i) saliency map from Achanta's model [6]

(a) (b) (c) (d) (e) (f) (g) (h) (i)

Figure 3. Results for a qualitative comparison between our method and the other six approaches: (a) original images; (b) ground truth; (c) saliency map of Itti's model [1]; (d) saliency map of Hou's model [8]; (e) saliency map of Ma's model [12]; (f) saliency map of Achanta's model [6]; (g) saliency map of Cheng's model [7]; (h) saliency map of Duan's model [5]; (i) saliency map of the proposed model

GrabCut [7] to obtain a binary mask for a given saliency map. Final saliency cut result is generated by this way as our binary mask to obtain the quantitative evaluation (see Figure 3).

4. **Conclusions.** We present a multiscale saliency detection algorithm based on 2DPCA to detect the saliency object in the color image. Our saliency algorithm is based on three elements: the absolute contrast, the bin contrast and the local contrast. We evaluate our

method on two publicly available data sets and compare our scheme with other models. The resulting saliency maps have a little improvement on the database with center bias mechanism. However, it is better suitable to salient object segmentation by preserving more fine details. In the future work, we will proceed to extend the proposed model and its corresponding algorithms to the video saliency detection.

Acknowledgment. This work is sponsored by the National Natural Science Foundation of China (NSFC) #61402192, Jiangsu Qing Lan Project, the Natural Science Foundation of the Jiangsu Higher Education Institutions of China #14KJB520006, 14KJB580002, 15KJB520004, the open fund for the Key Laboratory for traffic and transportation security of Jiangsu Province (TTS2015-05), the open fund of Jiangsu Provincial Key Laboratory for advanced manufacturing technology (HGAMTL-1401), the Science & Technology Fund of Huai'an under the Grant No. HAG2013059, HANZ2014006, HAG2014028.

REFERENCES

[1] L. Itti, C. Koch and E. Niebur, A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.20, no.11, pp.1254-1259, 1998.

[2] N. Bruce and J. Tsotsos, Saliency based on information maximization, *Advances in Neural Information Processing Systems*, pp.155-162, 2006.

[3] L. Zhang, M. Tong, T. Marks, H. Shan and G. Cottrell, SUN: A Bayesian framework for saliency using natural statistics, *Journal of Vision*, vol.8, no.32, pp.1-20, 2008.

[4] J. Harel, C. Koch and P. Perona, Graph-based visual saliency, *Advances in Neural Information Processing Systems*, pp.545-552, 2003.

[5] L. Duan, C. Wu, J. Miao, L. Qing and Y. Fu, Visual saliency detection by spatially weighted dissimilarity, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.21-23, 2011.

[6] R. Achanta, S. Hemami, F. Estrada and S. Süsstrunk, Frequency-tuned salient region detection, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1597-1604, 2009.

[7] M. Cheng, G. Zhang, N. Mitra, X. Huang and S. Hu, Global contrast based salient region detection, *IEEE International Conference on Computer Vision*, pp.409-416, 2011.

[8] X. Hou and L. Zhang, Saliency detection: A spectral residual approach, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-8, 2008.

[9] J. Yang, D. Zhang, A. F. Frangi and J. Yang, Two-dimensional PCA: A new approach to face representation and recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.26, no.1, pp.131-137, 2004.

[10] J. Shi and J. Malik, Normalized cuts and image segmentation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.22, no.8, pp.888-905, 2000.

[11] N. Bruce and J. Tsotsos, Saliency based on information maximization, *Advances in Neural Information Processing Systems*, pp.155-162, 2006.

[12] Y. Ma and H. Zhang, Contrast-based image attention analysis by using fuzzy growing, *The 11th ACM International Conference on Multimedia*, pp.374-381, 2003.