IEIE Transactions on Smart Processing and Computing

Visual Attention Detection By Adaptive Non-Local Filter

Dao Nam Anh[1]

[1] Department of Information Technology, Electric Power University, Hanoi, Vietnam

***** Corresponding Author: Dao Nam Anh

Received February 19, 2016; Revised February 22, 2016; Accepted February 25, 2016; Published February 29, 2016

** Extended from a Conference: Preliminary results of this paper were presented at the ICEIC 2016. This present paper has been accepted by the editorial board through the regular reviewing process that confirms the original contribution.*

***Abstract*:** Regarding global and local factors of a set of features, a given single image or multiple images is a common approach in image processing. This paper introduces an application of an adaptive version of non-local filter whose original version searches non-local similarity for removing noise. Since most images involve texture partner in both foreground and background, extraction of signified regions with texture is a challenging task. Aiming to the detection of visual attention regions for images with texture, we present the contrast analysis of image patches located in a whole image but not nearby with assistance of the adaptive filter for estimation of non-local divergence. The method allows extraction of signified regions with texture of images of wild life. Experimental results for a benchmark demonstrate the ability of the proposed method to deal with the mentioned challenge.

***Keywords*:** Adaptive non-local means filter; saliency; dissimilarity

1. Introduction

Patch similarity observation is fundamental image patches analysis which is generally capable for various application, such as denoising, tone mapping, edge editing and abstraction [1]. The patch analysis is followed by edge-keeping filters which are designed specifically for removing noise and preserving edge at the same time. The approach has also been proven successful by the use of observations in a neighborhood of a pixel of the interest, for example, Yaroslavsky's filter [2], the bilateral filter [3] and Susan filter [4] to distinguish edges from noise. Hence the filters allow to keep edges and eliminate noise.

The edge-keeping fitters can be done empirically, providing high certainty, but only for the cases where the patch similarity are well distributed locally, which permits observation in neighborhood to be rational. Due to the existence of the sparse distribution of similarity, the prevailing local approach in the field is turning toward combining it with non-local imaging modality and realizing patch similarity observation from the whole image.

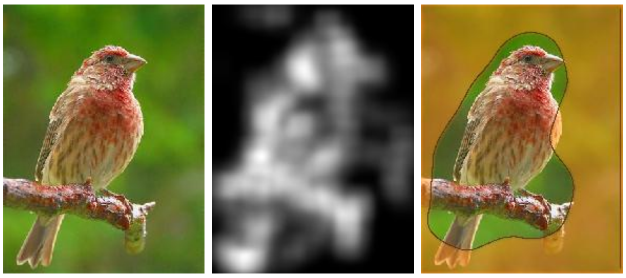
Among several techniques checking similarity in all places the non-local mean (NLM) filter is based similarity measure of patches with the weighted Euclidean distance [5]. Notable examples of the NLM are denoising by adaptive nonlocal regularization with Bregmanized methods [6].

The observation for similarity is analytical for edge-based functions and not suitable for region-based detection of visual attention. Hence we introduce a new version of NLM filter which is suitable for this mission. The article proposes a novel saliency method where the adaptive non-local mean filter focuses on observation of divergence of patches from the whole image.

Very often, the natural photos involve the texture of landscape, and wild animals, whose outlook is very similar to the environment to survive. Thus texture can be seen in both foreground and background posing challenge for the computation of visual attention. The adaptive NLM filter is able for patches analysis as a basic, employing the local symmetry factor as an additional constraint to the filter, in order to obtain better estimation of visual attention.

In this studies, we test the adaptive NLM filter on a set of natural images of a benchmark image set [7] providing user templates for quality metrics estimation. The contributions of our approach include two aspects:

- A novel adaptive non-local mean filter for observing divergence of patches. The method analyzes patches in spatial distance checking how the patches are distinct;
- A method for detection of visual attention using the



(a) Input (b) Salient map (c) Region in visual attention

Fig. 1. An example of using the adaptive NLM filter to detect visual attention.

adaptive NLM filter which is assisted by symmetry analysis. The experimental results of method are analyzed and discussed.

The adaptive NLM filter is implemented for extraction of regions in high visual attention. It applies adaptive Gaussian weights for measurement of intensity difference while keeping particular Gaussian weights for spatial distance.

Let's take a look at an example illustrated in Fig. 1 for getting brief concept of the algorithm. Input image initially blurred is showed in Fig. 1(a) Interest points found by gradient based filter are illustrated in Fig. 1(b) and visual attention region detected by the adaptive NLM filter is presented in Fig. 1(c).

The rest of this paper is organized as follows. The next section describes related works. Section 3 outlines the scientific background, of a new saliency detection algorithm taking into account the interest points and the dissimilarity analysis. The performance evaluation section includes the evaluations and discussion on the experiment results. Finally, we conclude this paper in the last section.

2. Related WorkS

Extensive efforts have been devoted to the research of visual attention to obtain consistent functionality which can be implemented for image segmentation, object detection and other media applications. We present briefly some works relevant to our method.

2.1 Saliency Detection

A method of self-resemblance is proposed in [8] where matrix cosine is used for measure of the likeness of a pixel to its surroundings. The method produces a saliency map whose each pixel is assigned a statistical likelihood of visual attention of a feature matrix. Our method is focused in feature distinction instead of feature similarity, and the method's measure checks the distinction between a patch around a pixel and its surroundings.

To allow combining top-down information with bottom-up visual features, the proposed method in [9] involves natural image statistics of a Bayesian framework for collection of natural images providing an explanation for search asymmetries observed in humans. Our method performs visual attention search considering interest points that is derived in symmetry.

Conditional random field (CRF) is exploited for saliency measure in [10] with a statistical framework for local feature contrast in illumination, color, and motion information. Our method focuses on non-local contrast but not local in color and illumination by the adaptive non-local filter.

2.2 Saliency and Interest Points

There is investigation on incorporating region of visual attention and interest points. The interest point detectors are suggested in [11] to select those features which are most appropriate or salient. Other method for feature point detection by saliency is noted in [12] where saliency is computed by a global optimization process constrained by volumetric conditions and points are discovered as the extrema of the saliency response. Our method is to select salient region suitable for interest points in contrast.

The relationship between the interest point and the visual attention is outlined in [13] reporting the performance of different interest point models in predicting the visual fixation. The work assumed that the region with many interest points has more confidence to be the fixation of human eyes. Agreed with this notion our method realizes visual attention search regarding to the interest points. Though, the way to search is distinct by applying non-local search of patches divergence.

A spatio-temporal attention detection framework is presented in [14] for detecting both attention regions and interesting actions in video sequences. Interesting regions and points in images are established and then combined. Our method searches interesting regions which is based on spatial distribution of interest points.

Designing interest operators based on human eye movement statistics is studied in [15]. Their saliency measures are reported well associated with contrast which is assumed to be one of the most visual attention features in human low-level vision. Our method is focused on contrast but on the level of patches with non-local search for patch divergence. The method is outlined in next section.

3. The Proposed Method

3.1 The Interest Points

The points of interest in images can specify salient features so they are capable of guiding detection of salient regions. The distribution of interest points may illustrate visual focus. With consideration of assumption of symmetry of most naturel object, interest points detection can be performed with symmetry investigation [16]. Starting with the gradient of second order $g_2(x)$ of a pixel x, its pair of neighborhood pixels $p^+(x)$, $p^-(x)$ in distance w which are affected by the pixel x can be identified:

$$p^{\pm}(x) := x \pm round\left(w\frac{g_2(x)}{\|g_2(x)\|} \right), g_2(x) = \frac{d^2u}{d^2x} \quad (1)$$

The orientation o is updated with $p(x)$ and the magnitude m of the affected pixels are accumulated by value of the gradient $g_2(x)$:

$$o(p^{\pm}(x)) := o(p^{\pm}(x)) + 1 \quad (2)$$

$$m(p^{+}(x)) := m(p^{+}(x)) + \|g_2(x)\|$$
$$m(p^{-}(x)) := m(p^{-}(x)) - \|g_2(x)\| \quad (3)$$

The symmetry $s(x)$ for a pixel is evaluated by the orientation $o(x)$ and magnitude $m(x)$ with scaling factor k for magnitude normalization:

$$s(x) = \tfrac{1}{c_1(x)} \int_{\|t\|=w} \|o(x+t)\|^k \, m(x+t)dt \quad (4)$$

$$c_1(x) = \int_{\|t\|=w} \|o(x+t)\|^k \, dt \quad (5)$$

The measure uses the Gaussian weights of spatial distance and the orientation weight to aggregate affect magnitude for each pixel. Thus local maxima of the estimation of the symmetry s can be the fundament to identify interest points. The whole calculation is based on the second order gradient and its affect analysis on neighbor patches. The estimation of symmetry now is implemented for further analysis of salient region.

3.2 The Adaptive Non-Local Mean Filter

The non-local search in large meaning represents the search for similarity of patches located in whole images while local search scans the similarity in narrow neighborhood. In denoising, the non-local mean filter [5] analyzes not only the intensity $u(x)$ in a single point but the geometrical configuration in a whole neighborhood $N(x)$.

$$nlsim(x,y) = E\|u(N(x)) - u(N(y))\|^2_{2,\sigma}$$
$$nlsim(x,y) = \tfrac{1}{c_2} \int_{t<=\tau} G_\sigma(t)\|u(x+t) - u(y+t)\|dt \quad (6)$$
$$c_2 = \int_{t<=\tau} G_\sigma(t)dt; nsim \in [0,1] \quad (7)$$

The similarity measure applied the Gaussian weight G with deviation σ. Involving weighted average of all the pixels in the image space Ω for each pixel the filter produces smoothing effect with a parameter h:

$$f(x) = \tfrac{1}{c_3(x)} \int_\Omega \exp(-\frac{nlsim(x,y)}{h^2})u(y)dy \quad (8)$$

$$c_3(x) = \int_\Omega \exp(-\frac{nlsim(x,y)}{h^2})dy \quad (9)$$

The similarity measure in the NLM is capable for noise remove and making image smooth. In order to detect visual attention regional distinction is need to be estimated. As the non-local similarity can be get with (5), the non-local dissimilarity is identified by (10) in the adaptive form of (5):

$$nldis(x,y) = 1 - nsim(x,y) \quad (10)$$

It is preferable to take symmetry based interest points estimation $s(x)$ to analyze the dissimilarity by (11). Hence the adaptive non-local mean filter by (12) arrives the measurement for salient region:

$$nldis(x,y) = 1 - E\|s(N(x)) - s(N(y))\|^2_{2,\sigma}$$
$$v(x) = \bar{f}(x) = \tfrac{1}{\bar{c}(x)} \int_\Omega \exp(-\frac{nldis(x,y)}{h^2})s(y)dy \quad (11)$$
$$\bar{c}(x) = \int_\Omega \exp(-\frac{nldis(x,y)}{h^2})dy \quad (12)$$

From the analysis of symmetry, similarity, and the dissimilarity we obtain a model for visual attention detection with corresponding consistent formula of the adaptive non local filter. The analytical method is then implemented to a novel saliency algorithm, described in next section.

3.3 The Algorithm

The concept and pseudo code for the detection of visual attention combining information of interest points and dissimilarity estimation by the adaptive non-local mean is presented in Figs. 2 and 3. The algorithm consists

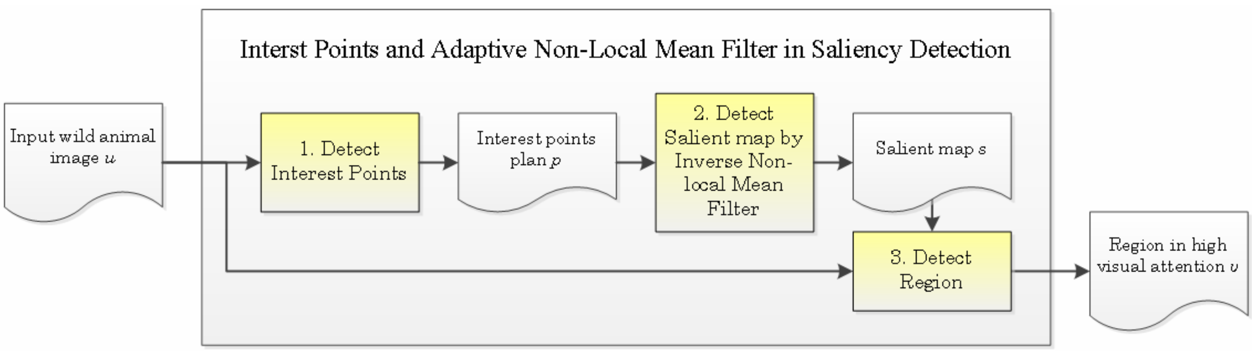


Fig. 2. Concept of detection of salient region with the adaptive non-local mean filter.

```
Algorithm 1: Interest Points and Adaptive
Non-Local Mean Filter in Saliency Detection
(PAMF)
Input: image u, parameters σ,h,w
Output: salient map v
Begin
  1. //A. Interest points detection
  2. Calculate its second gradient g2(x);
  3. Define affected neighborhood pixels for each pixels x;
  4. Get orientation and magnitude o and m;
  5. Calculate symmetry-based plan of interest points s(x);
  6. //B. Visual attention detection
  7. Estimate non-local dissimilarity nldis(x);
  8. Run the adaptive non-local filter for the saliency map v;
  9. //C. Particular manipulation
  10. Visual attention region extraction;
End
```

Fig. 3. Algorithm 1.

of two major stages: interest points plan estimation and the salient map calculation. The third stage is for further image manipulation tasks like region extraction and selective region enhancement.

To demonstrate results of each stage, an example is shown in Fig. 4. Image in Fig. 4(a) is *u(x)*, which has interest points in Fig. 4(b). The adaptive non-local mean filter produces the map *v(x)* in Fig. 4(c). Line 10 of pseudo code extracts region in high visual attention illustrated in Fig. 4(d). It's possible to improve contrast for the region, demonstrated in Fig 4(e). Thought the wild animal image of the example contains texture in both foreground and background, the marked region in Fig 4(d) shows suitable estimation of saliency.

The stage of estimation for interest points takes two double loops performing for *n* time with the size of image by *n,* multiplied to the size of patch *w*, results computing complexity $O(wn)$. The adaptive non-local mean filter like its straight vision performs search of dissimilarity of patches located in whole images, acquires time complexity $O(wn^2)$. Hence the total complexity is $O(wn^2)$.

4. Performance Evaluation

600 of wild animal images are selected from a general saliency benchmark [7] were used as test cases for the algorithm. Most of the original images are textured in both foreground and background. The symmetry are applicable for most of foreground objects in the actual collection of animals. Though this is relevant also for background. The test cases are tested with mentioned related methods [8-10] for performance analysis.

A number of examples from the test is demonstrated in Table 1. The names of methods in the first column reflect the results of the method applied for input images shown in the first row. A window is built from salient map of each method for checking similarity with manual segmentation templates which are provided by the benchmark [7]. These cases present wild animals in textured images where both foreground and background can be textures.

The performance of the new algorithm was demonstrated on a range of wild animal images from a benchmark and compared with three remarkable methods from the literature. Each benchmark image has its three manual segmentation templates for quality estimation.

The templates are used for metrics calculation including Mean Square Error (MSE) and averaged Sum of Absolute Differences (SAD).

$$MSE(v_1, v_2) = \frac{1}{numel(\Omega)} \int_\Omega \left(v_1(x) - v_2(x)\right)^2 dx \qquad (13)$$

$$SAD(v_1, v_2) = \frac{1}{numel(\Omega)} \int_\Omega \left|v_1(x) - v_2(x)\right| dx \qquad (14)$$

Table 2 presents the performance with MSE and SAD metrics. The best score are marked in bold. Proposed PAMF achieved the best MSE score and averaged SAD, while Self-resemblance method [8] gained the best SAD.

The experiments were performed with two parameters. The deviation parameter σ is for managing Gaussian weights, $\sigma=.5$ in the test. The lower values of σ make the weights flat. The *h* parameter regulates degree of the adaptive non-local mean filtering. Its value was set the same value of *w*- the size of working frame of filter. Large value of *w* follows extensive computing time as we have $O(wn)$ as discussed in the beginning of the section. The parameter should not too small to cover texture feature for patches dissimilarity analysis. With the image width of 250 pixels the frame size is set to $w=5$ in our test. This parameter can be regulated depending the average size of texture.

Thus quality of points affects to the salient map. Considering symmetry in detecting points, the method may lead to inconsistent output when the symmetry rule is not applicable to the input image, or it appeared strongly in the background objects. Example shown in Fig. 5 illustrates the case of making noise of points which is followed by

(a) Input image *u* (b) Interst points (c) Saliency map *v* (d) Extracted region (e) Selective enhancement

Fig. 4. An example of Algorithm 1.

Table 1. Saliency Detection For Wild Animal Images.

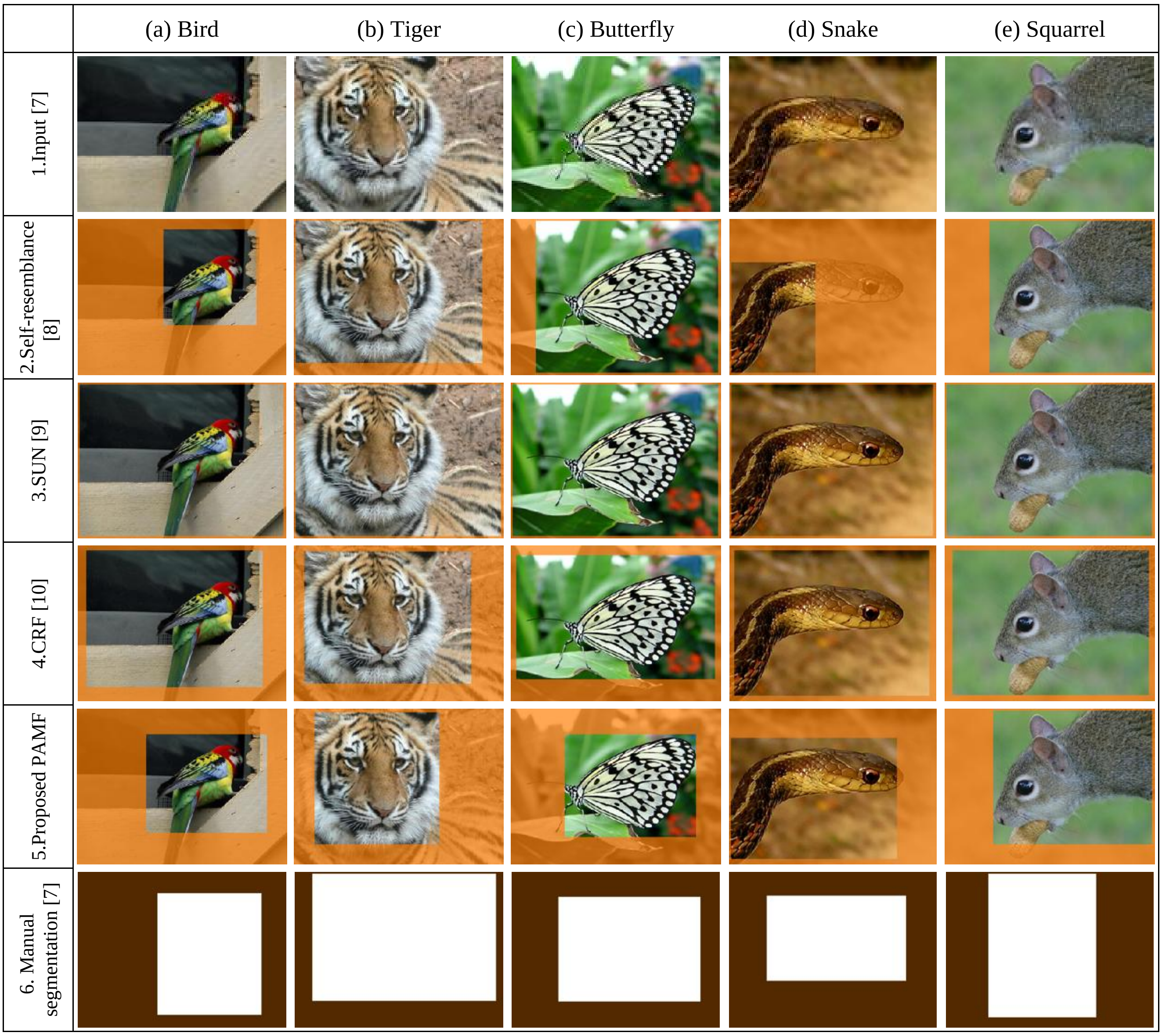
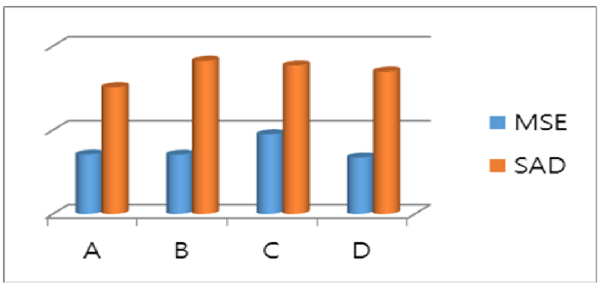


Table 2. Performance for Collection of Wild Animal Images.

Method	MSE	SAD
A. Self-resemblance [8]	0.1414	**0.3019**
B. Bayesian framework for saliency using natural statistics [9]	0.1410	0.3648
C. Segmenting Salient Objects from Images anVideos [10]	0.1895	0.3536
D. Proposed PAMF	**0.1342**	0.3390

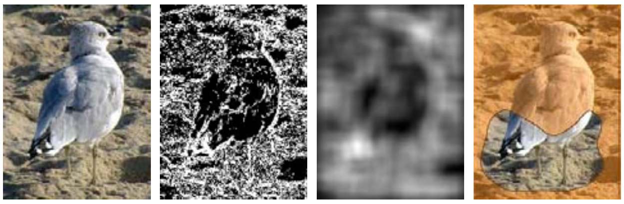


uncertain salient region.

This suggests that the distribution of interest points should be considered and regulated before taking the dissimilarity analysis by the adaptive non-local mean filter. Again, the complexity of the filter is not linear. The filter can be studied for another version to promote the performance. This is reserved for our further study for improvement.

5. Conclusion

In this paper a novel saliency detection method has



(a) Input (b) Interest poitns (c) Salient map (d) Extracted region

Fig. 5. Case of sparse points.

been proposed. The algorithm is robust against texture appeared in wild animal images. By employing a suitable interest point detector, the resulting estimates are precise and efficient for textured images. The adaptive non-local filter is potent and stable for dissimilarity analysis. Furthermore, the available filter provides essential tools for many different types of images, particularly in regard to object detection often seen in natural images.

Acknowledgement

We wish to thank Buades, A. A., Coll, B., Morel, J.M for their comprehensive work on non-local mean filter and its applications. Special thanks G. Loy and A. Zelinsky for the paper and codes on interest points. Author thanks anonymous referees for their valuable comments.

References

[1] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, “Edge-preserving decompositions for multi-scale tone and detail manipulation,” ACM Trans. Graph., vol. 27, no. 3, Aug. 2008, Art. ID 67. Article (CrossRef Link)

[2] Yaroslavsky, L.P.: Digital Picture Processing-an Introduction. Springer, Heidelberg (1985) Article (CrossRef Link)

[3] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in Proc. 6th ICCV, Bombay, India, Jan. 1998, pp. 836-846. Article (CrossRef Link)

[4] Smith, S.M., Brady, J.M.: SUSAN-a new approach to low level image processing. International Journal of Computer Vision 23, pp. 45-78 (1997). Article (CrossRef Link)

[5] Buades, A., Coll, B., Morel, J.M.: On image denoising methods. SIAM Multiscale Modeling and Simulation 4, pp. 490-530 (2005). Article (CrossRef Link)

[6] Xiaoqun Zhang, Martin Burger, Xavier Bresson, and Stanley Osher, Bregmanized Nonlocal Regularization for Deconvolution and Sparse Reconstruction. SIAM J. Imaging Sci., 3(3), pp. 253-276. Article (CrossRef Link)

[7] Tie Liu, J Sun, NN Zheng, Xoou Tang and Heung-Yeung Shum. Learning to detect a salient object. CVPR 2007. Article (CrossRef Link)

[8] Hae Jong Seo and Peyman Milanfar, Nonparametric Bottom-Up Saliency, Detection by Self-Resemblance, Hae Jong Seo and Peyman Milanfar, Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. Article (CrossRef Link)

[9] Lingyun Zhang, Matthew H. Tong, Tim K. Marks, Honghao Shan & Garrison W. Cottrell (2008). SUN: A Bayesian framework for saliency using natural statistics. Journal of Vision, 8(7):32, pp. 1-20, Article (CrossRef Link)

[10] Esa Rahtu, Juho Kannala, Mikko Salo, Janne Heikkilä, Segmenting Salient Objects from Images and Videos, Volume 6315 of the series Lecture Notes in Computer Science pp. 366-379, Article (CrossRef Link)

[11] D. Hall, B. Leibe, and B. Schiele, Saliency of Interest Points under Scale Changes, British Machine Vision Conference 2002, Article (CrossRef Link)

[12] Rapantzikos, K., Sch. of Electr. & Comput. Eng., Nat. Tech. Univ. of Athens, Athens, Greece, Avrithis, Y., Kollias, S. Dense saliency-based spatiotemporal feature points for action recognition. Computer Vision and Pattern Recognition, 2009. Article (CrossRef Link)

[13] Xiang Zhang, Inst. of Digital Media, Peking Univ., Beijing, China, Shiqi Wang, Siwei Ma, Wen Gao, A study on interest point guided visual saliency, Picture Coding Symposium (PCS), 2015. Article (CrossRef Link)

[14] Yun Zhai and Mubarak Shah. 2006. Visual attention detection in video sequences using spatiotemporal cues. In Proceedings of the 14th ACM international conference on Multimedia (MM '06). ACM, New York, NY, USA, pp. 815-824. Article (CrossRef Link)

[15] Kienzle W, Franz MO, Schölkopf B and Wichmann FA (May-2009) Center-surround patterns emerge as optimal predictors for human saccade targets Journal of Vision 9(5:7) pp. 1-15. Article (CrossRef Link)

[16] G. Loy and A. Zelinsky. Fast Radial Symmetry for Detecting Points of Interest. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003. Article (CrossRef Link)

Dao Nam Anh received his B.S. in Applied Mathematics and Ph.D. in Physics-Mathematics from University of Moldova, in 1987 and 1992, respectively. He was involved in various international software projects. He is currently teaching at Electric Power University. His research interests include Intellectual Intelligence, Image Processing and Pattern Recognition, Machine Vision.