This full text paper was peer-reviewed at the direction of IEEE Instrumentation and Measurement Society prior to the acceptance and publication.

Localize Car Door Handles with Image Segmentation and Saliency Detection

Fan Zhang, Mengchao Hao, Muyang Liu , Jinyu Yang
School of Instrumental Science and Opto-electronics Engineering
Beihang University
Beijing, China
zhangfan17@buaa.edu.cn, liangjiren@buaa.edu.cn, muyangliu@buaa.edu.cn, yangjinyu@buaa.edu.cn

***Abstract*—Importance of little objects in cars such as door handles is obvious, both in daily lives and in industrial manufacture. However, since the lack of the distinctive appearance and feature, obtaining the location of them is still remaining a challenge. This paper proposes an effective approach for the detection of the door handles of cars.**

The method innovatively combines frequency and spatial domains' algorithms to detect the location of little objects of cars in a relatively large image. To illustrate the method more concisely, our method localizes door handles through image segmentation and visual saliency detection. First, by segmenting the image we can remove the unnecessary area to improve the speed and accuracy of our approach. After finding the region of interest, our approach uses a visual saliency detection algorithm named Spectral Residual Approach which can get the location of door handles accurately.

At last, the approach is tested by different kinds of images of vehicles. The results of the experiments show that our approach is obvious and practical.

Keywords—Hough Transformation; Image segmentation; Little Landmarks; Saliency Detection; Spectral Residual Model

I. INTRODUCTION

Door handle of a car is usually the first thing we manipulate when interacting with the car and the world is full of objects like these tiny but useful things which we call little landmarks. These little landmarks are barely visible in an image, and it is difficult for us to detect their locations. Automatically localizing door handles of a car in images is challenging, since they do not have a distinctive appearance of their own [1]. However, these landmarks are visual attention objects and they are largely defined by their context. For most cars, the position of their door handles is similar. They are placed in the middle area between car wheels and car windows and the wheels and windows have their own distinctive appearances. Almost all car wheels are round and the bottom of most car windows, which connect the windows with the car body, is linear. These are two main characteristics and they are easily recognized in an image through image processing skills.

When a person looks at an image, he/she may pay much visual attention on a small region while paying less attention on other regions according to the human visual system [2,3]. Visual attention is an important method to perceive the world around humans and it has been widely studied in many fields. Saliency detection is an effective way to find the visual attention region in images with computation on their feathers. Recently, saliency detection has been widely used in object detection, object recognition and image/video compression [4-8].

We describe a method to localize the door handles of a car which combining image segmentation and visual saliency detection. We handle the tiny landmarks localization problems with visual attention detection because we consider tiny landmarks as saliency objects. For most car images, the background has some visual attention objects and we need to remove them before saliency detection, so our method segments the middle area which only has door handles and car body from the original image firstly and then we use Spectral Residual Saliency Model [9] to generate saliency map of this little area in order to localize the target tiny objects.

II. RELATRD WORK

Landmark localization has been well researched in the past few years especially in the field of human pose estimation [10] and bird part localization [11,12]. However, the above-mentioned methods are aiming to detect objects which are large and distinctive [13]. And there is not so much work existing for localizing little landmarks. Since little landmarks are largely defined by their context, therefore the method of using context to localize little landmarks has been studied recently [14, 15].

On the other hand, saliency detection has been studied since 1980s, and now a great number of methods proposed which can be classified into two categories: biologically inspired [16, 17] and learning based models [18]. Most saliency detection methods are biologically inspired which are developed based on the understanding of human visual systems. The most representative work on saliency detection is Itti's model [19], which combines center-surrounded features of color, intensity and orientation together. Most of the detection models focus on summarizing the properties of target objects. However, general properties shared by various categories of objects are not likely to exist. Hou et al. [9] proposed a method named Spectrum Residual Model and they pose this problem in an alternative way to explore the properties of the backgrounds.

978-1-5386-1620-8/17/$31.00 ©2017 IEEE

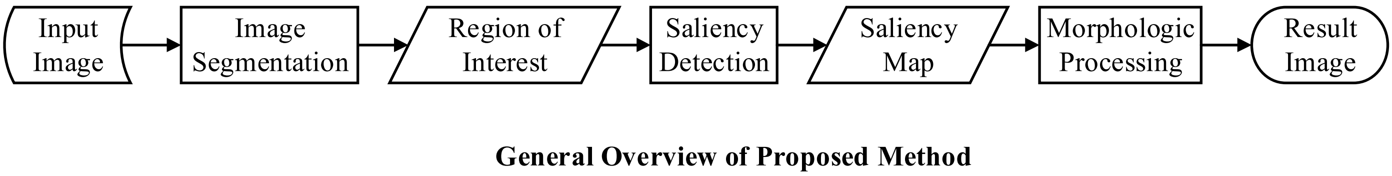


Fig. 1. General Overview

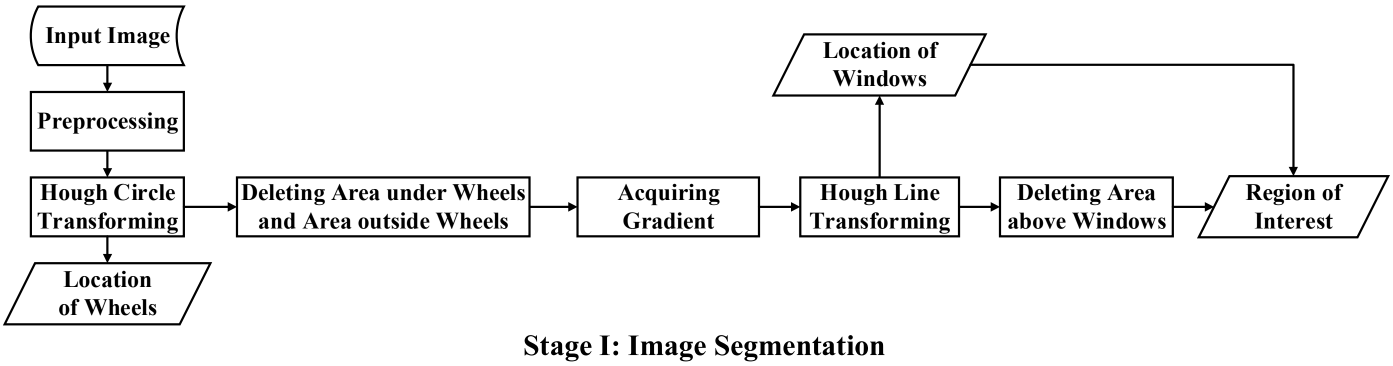


Fig. 2. Steps of the Image Segmentation stage

Singh [20] proposed a method using full convolution networks (FCNs) to find a sequence of latent landmarks which are useful for finding the little landmark. Inspired by his method and Hou's saliency model, we proposed our method which combines the image segmentation and saliency detection ways to localize little landmarks. Instead of using a method related to Convolution Neural Network [21-24], a simple image segmentation algorithm is used here to detect the context of a car and then the Spectrum Residual Model is applied to find car door handles, which performs well in the experiment.

III. Proposed Method

Our method is mainly divided into three main stages. Fig. 1 provides an overview of our method.

A. Stage 1: Image Segmentation

We use a way to segment image with its context. Inspired by Zhu's stereo-matching algorithm [25, 26], we classify the objects in an image into two different kinds. One is distinct object which occupies a large area with remarkable shape features and color features and we call it latent landmark. The other is a little object which we call it target landmark. In most cases, the size of latent landmarks is larger than target landmarks so it is convenient for us to localize them first.

In a car, we consider car wheels and car windows as latent landmarks while considering car door handles as target landmarks. We divided this stage into several stages and Fig. 2 presents all steps of this stage.

1) Step1: Average filter

Since we need to use image difference in the following steps so we need to use average filter $h_n(x,y)$ to remove noise from the image $I(x,y)$. The result image $M(x,y)$ can be approximated by convoluting the input image:

$$M(x,y) = h_n(x,y) * I(x,y) \tag{1}$$

Where $h_n(x,y)$ is an $n \times n$ matrix defined by

$$h_n(x,y) = \frac{1}{n^2}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \tag{2}$$

in our method, n equals 3.

2) Step2: Localize car wheel

All car wheels are round and they can be described by circles. A circle can be described by 3 parameters which are the coordinate of the center of a circle and the length of radius.

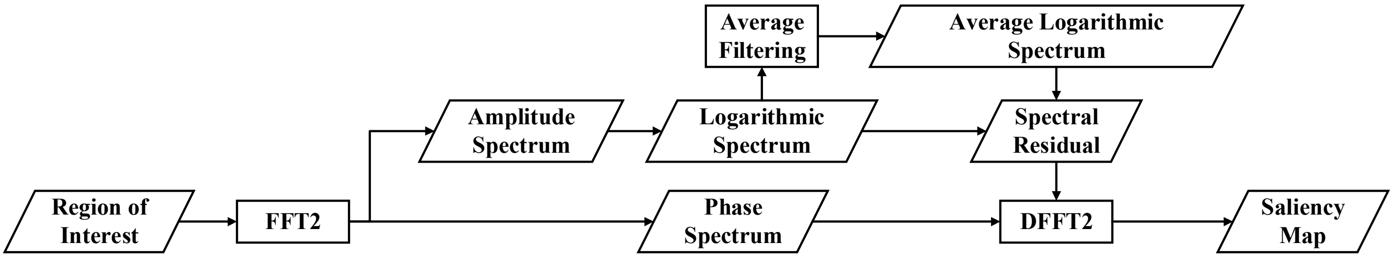
$$\text{Circle} = \begin{cases} x & \text{Center Abscissa} \\ y & \text{Center Ordinate} \\ r & \text{Radius} \end{cases} \tag{3}$$

Hough Gradient transformation is used here to detect circles. Firstly, Canny Edge Detecting Method is utilized to get the edge image. For each non-zero point in the edge image, we calculate the gradient of these points in the image and label them. After that, drawing lines along the gradient direction and the opposite direction then calculate the number of points which pass through the line is counted in the accumulator. Larger the number of points in the accumulator, more likely it is the center of the circle. Location of the car wheels is detected in this way. In the end, we draw a line above the car wheels as a segmentation line in order to remove information beneath this line.

3) Step3: Localize car windows

For most cars, the bottom of car windows is linear and has a distinct appearance from the car body. We use these features to extract the bottom line of the car window with edge detecting algorithm.

First, we utilize Sobel Gradient Operator to calculate the gradient in the vertical direction of the image so that we can obtain the horizontal edge lines. Fig. 3 shows the result of this



Stage II: Saliency Detection

Fig. 4. Steps of the Saliency Detection stage

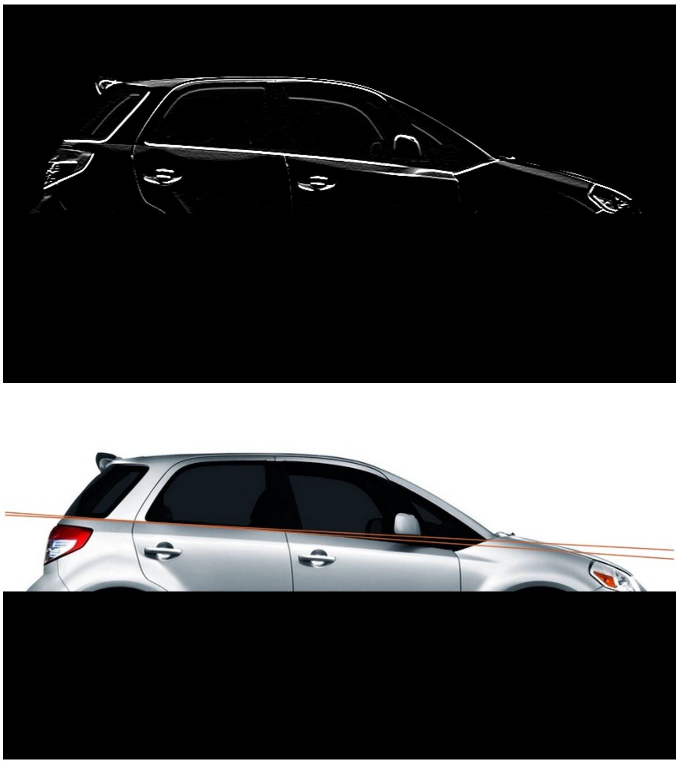


Fig. 3. Result of the Localize car windows

step. It is obvious that the bottom line is distinctive. Same as the last step, then we utilize Hough Line Transform to get the position of the bottom line.

4) Step4: Find the range of interest (ROI)

After localizing the location of car wheels and the bottom line of car windows accurately, we use the coordinate of car wheels and bottom line to remove pixels locates above the bottom line of car windows and pixels under the segmentation line so that we can get the middle area (ROI).

B. Stage2: Saliency Detection, Spectral Residual Model

After getting the ROI which only has the car handles and car body, we use saliency detection method to generate a saliency map. According to efficient coding perspective theory [27], the image information H(Image) is divided into two parts:

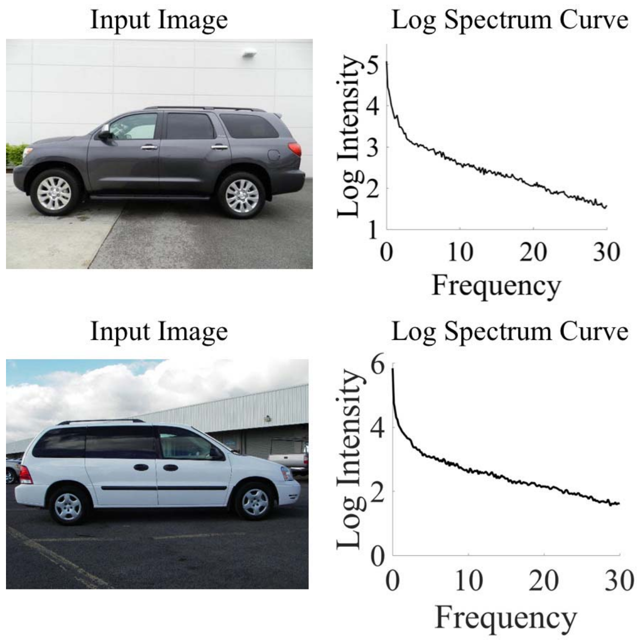


Fig. 5. Examples of Log Spectrum Curve

$$\mathrm{H}(\mathrm{Image}) = \mathrm{H}(\mathrm{Innovation}) + \mathrm{H}(\mathrm{Prior\ Knowledge}) \quad (4)$$

H(Innovation) denotes the novel part, and H(Prior Knowledge) is the redundant information that should be suppressed by a coding system.

Spectral Residual Model is used to process an image in the frequency domain. When we analyze an image in the frequency domain, the 1/f law [28] suggests that the amplitude of the average Fourier spectrum of a natural image obeys a distribution:

$$\mathrm{E}\{\mathrm{A}(f)\} \propto 1/f \quad (5)$$

which means the log intensity decreases as the frequency going up [29]. Fig. 5 presents examples of Log spectrum curve. This law shows that the log spectrum of different images has similar trends. Therefore, when we think about an average log spectrum of up to 100 images, it becomes smooth and continuous. Similarities imply redundancies, so for a single image, in its log spectrum, what deserves our attention is the information that jumps out of the smooth curves, and this is the saliency area.

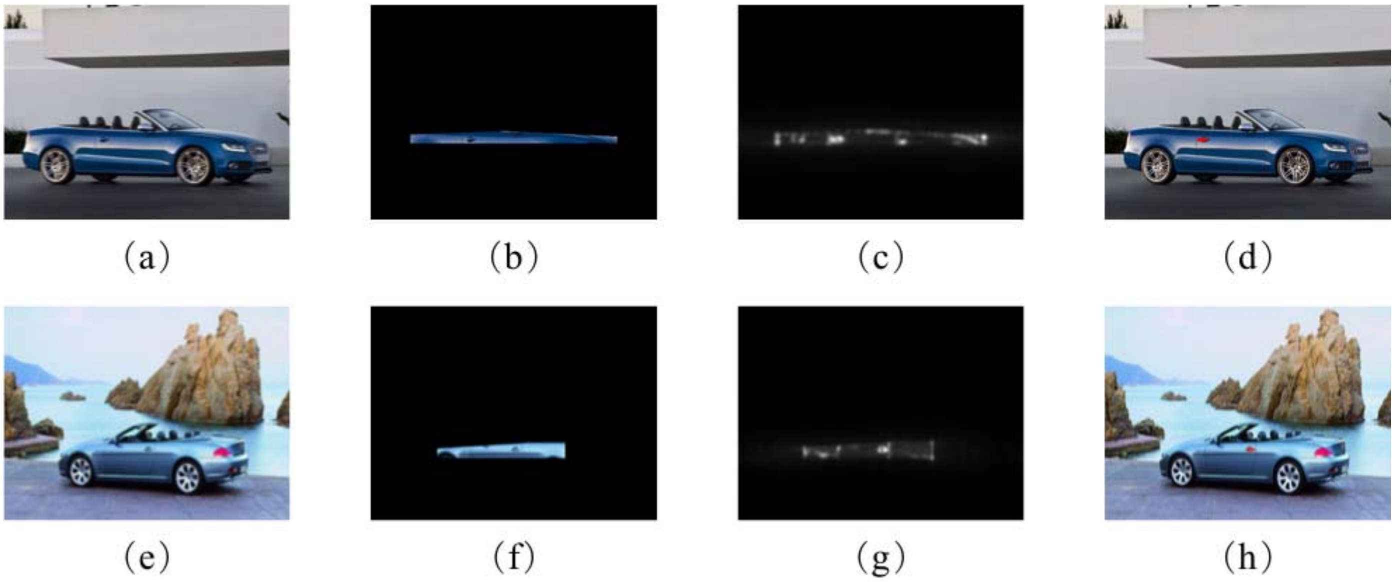


Fig. 6. Results of every stage

Therefore, given the ROI image $F(x,y)$, we do the following steps to get a saliency map. All steps of this stage are presented in Fig. 4.

1) Step1:

The two-dimensional Fourier transform is performed on a given image and we get the amplitude spectrum $A(f)$ and phase spectrum $P(f)$:

$$A(f) = \Re(\mathcal{F}[I(x)]) \tag{6}$$

$$P(f) = \Im(\mathcal{F}[I(x)]) \tag{7}$$

2) Step2:

Adopting the log transformation to amplitude spectrum to get a log spectrum representation $L(f)$:

$$L(f)=\log(A(f)) \tag{8}$$

3) Step3:

Adopting a local average filter $h_n(f)$ to take the average of log spectrum and we define this as the average log spectrum. And spectral residual $R(f)$ can be obtained by:

$$R(f) = L(f) - A(f) \tag{9}$$

4) Step4:

The spectral residual contains the innovation of an image. Using Inverse Fourier Transform, we can obtain the output image in spatial domain. And in order to have better visual effects, smoothing the output image with a Gaussian filter (31×31) and we can finally get a saliency map

$$S(x) = g(x) * \mathcal{F}^{-1}\left[\exp(R(f)+P(f))\right]^2 \tag{10}$$

C. *Stage3: Morphologic Processing*

After obtaining the saliency map of the ROI image, we need to do some additional work to get the position of car door handles. First, we use simple threshold segmentation to detect objects in a saliency map. Given $D(x)$ of the saliency map, the object map $O(x)$ obtained:

$$O(x)=\begin{cases}1 & \text{if } D(x) > \text{threshold} \\ 0 & \text{otherwise}\end{cases} \tag{11}$$

During our test, we set $\text{threshold}=E(S(x))\times2.5$, where $E(S(x))$ is the average intensity of the saliency map. The selection of threshold is a trade-off problem and its value could be changed among different saliency images.

However, the edge of ROI, some other objects such as car fuel tank cap, will also become saliency objects and these false objects will cause some misrecognition in some case. In most cases, car door handles cover the largest area in the saliency map. Therefore, we use morphological erode operator to erode the binary saliency map and then we can obtain the final position of car door handles.

IV. EXPERIMENTAL RESULT

In order to verify the effect of the proposed method, we tested our method on the Stanford Cars dataset [30], which contains cars of nearly all classes. We select some images which are taken in front of car door handles as testing images. All the experiments selected C++ environment as the implementation language. Xcode 8.3.3 was used as the development environment. CPU with Intel Core i5, 2.60GHz and 8G memory size.

Fig.6 shows two series of results of every stage in our method. Fig. 3 (a) and Fig. 3 (e) are the input image of the experiment, Fig. 3 (b) and Fig. 3 (f) show the ROI after segmentation, Fig. 3 (c) and Fig. 3 (g) are the saliency map smoothed with a Gaussian filter and Fig. 3 (d) and Fig. 3 (h) show the final location of the door handle. It is obvious that the regions of interest cover only car door handles and car body, thus the saliency maps consider door handles as the most salient area. The final result images present the practicality of our method in detecting little objects of cars.

However, since we use traditional algorithm to localize car wheels and windows, therefore for some images which cars have a strange position, our method failed to detect them. In addition to that, the graffiti and inkjet in cars may cause mistakes in saliency detection so that our method will show a wrong result.

V. FURTHER WORK

We developed our method by finding the shape of car wheels and windows in traditional ways, so this method is largely constrained by the photo position of a car. Our method performs poor due to this reason in some images, so the potential work is to train better classifier to detect car wheels and windows instead of traditional image processing algorithm.

Furthermore, we use saliency detection model based on the frequency domain, it performs well in most cases whereas it presents non-sense results sometimes. Moreover, we can apply our method into the measuring system of machine vision [31, 32]. Due to the deep learning network, we can apply more efficient and accurate model in our method to improve the practicality of our method.

VI. CONCLUSION

This paper describes a novel approach to localize little landmarks of cars such as door handles by image segmentation and saliency detection. Different from other methods which aim at detecting objects, we focus on context instead of object itself so the major part of our work is to using image processing ways to detect car windows and car wheels in an image. Our approach combines algorithms both in spatial and frequency domain. We segment the input image into region of interest which has only car body and car door handles, after which considering door handles as saliency objects and use spectral residual model to generate saliency map which focuses on the position of car door handles.

We evaluate the performance of our method in some car datasets and the result shows that the door handles of most cars are saliency objects so they can be localized via our methods, which confirms the practicability of our method.

ACKNOWLEDGMENT

This research is funded by the National Natural Science Foundation of China (NSFC) under grant No.61375025. The authors would like to express our appreciations to the reviewers for their thorough review and very helpful comments, which help improving this paper.

REFERENCES

[1] Egeth, H. E., Virzi, R. A., & Garbart, H., “Searching for conjunctively defined targets,” J Exp Psychol Hum Percept Perform, vol. 10, no. 1, pp. 32-39, 1984.

[2] Borji, A., Sihite, D. N., & Itti, L., “What stands out in a scene? a study of human explicit saliency judgment,” Vision Research, vol. 91, no. 15, pp. 62-77,2013.

[3] Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S., & Zhao, Q.,” Predicting human gaze beyond pixels,” Journal of Vision, vol. 14, no. 1, pp. 97-97, 2014.

[4] Shiping Zhu and Chunyan Zhang, "A fast algorithm of intra prediction modes pruning for HEVC based on decision trees," Multimedia Tools and Applications, 2017, doi: 10.1007/s11042-016-4056-0.

[5] Shiping Zhu, Yangshuan Hou, Zaikuo Wang and Kamel Belloulata, "Fractal video sequences coding with region-based functionality," Applied Mathematical Modelling, vol. 36, no. 11, pp. 5633-5641, 2012.

[6] Deng, X., Xu, M., Jiang, L., Sun, X., & Wang, Z., “Subjective-driven complexity control approach for hevc,” IEEE Transactions on Circuits & Systems for Video Technology, vol. 26, no. 1, pp. 91-106, 2016.

[7] Shiping Zhu, Shupei Zhang, and Chenhao Ran, “An improved inter-frame prediction algorithm for video coding based on fractal and H.264,” IEEE Access, 2017, doi: 10.1109/ACCESS.2017.2745538.

[8] Shiping Zhu and Ziyao Xu, "Spatiotemporal visual saliency guided perceptual high efficiency video coding with neural network," Neurocomputing, 2017, doi: 10.1016/j.neucom.2017.08.054.

[9] Hou, X., & Zhang, L., (2007). “Saliency detection: A spectral residual approach,” IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2007, pp.1-8, IEEE Computer Society, 2007.

[10] B. Sapp, A. Toshev, and B. Taskar. Cascaded models for articulated pose estimation. In ECCV, 2010.

[11] K. Xu, J. Ba, R. Kiros, A. Courville, R. Salakhutdinov,R. Zemel, and Y. Bengio. Show, attend and tell: Neural im- age caption generation with visual attention. arXiv preprint arXiv:1502.03044, 2015.

[12] Liu, J., Li, Y., & Belhumeur, P. N., “Part-Pair Representation for Part Localization,” Computer Vision – ECCV 2014, Springer International Publishing, pp. 456-471, 2014.

[13] Caicedo, J. C., & Lazebnik, S., “Active object localization with deep reinforcement learning,” IEEE International Conference on Computer Vision, pp. 2488-2496, 2015.

[14] Alexe, B., Heess, N., Teh, Y. W., & Ferrari, V., “Searching for objects driven by context,” Advances in Neural Information Processing Systems 25, pp. 890-898, 2012.

[15] Fu, H., Cao, X., & Tu, Z., “Cluster-based co-saliency detection,” IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, vol. 22, no. 10, pp. 3766, 2013.

[16] Ruderman, D. L., “The statistics of natural images,” Network Computation in Neural Systems, vol. 5, no. 4, pp. 517-548, 1994.

[17] Lin, Y., Tang, Y. Y., Fang, B., Shang, Z., Huang, Y., & Wang, S., “A visual-attention model using earth mover's distance-based saliency measurement and nonlinear feature combination,” IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 35, no. 2, pp. 314-328, 2012.

[18] Han, J., Zhang, D., Cheng, G., Guo, L., & Ren, J.,” Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning,” IEEE Transactions on Geoscience & Remote Sensing, vol. 53, no. 6, pp. 3325-3337, 2015.

[19] Itti, L., Koch, C., & Niebur, E., “A model of saliency-based visual attention for rapid scene analysis,” IEEE Computer Society, 1998.

[20] Singh, S., Hoiem, D., & Forsyth, D., “Learning to Localize Little Landmarks,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 260-269, 2016.

[21] Fergus, R., Perona, P., & Zisserman, A., “Object class recognition by unsupervised scale-invariant learning,” IEEE Computer Society Conference on Computer Vision & Pattern Recognition, vol. 2, pp.II-264-II-271, 2003.

[22] Juneja, M., Vedaldi, A., Jawahar, C. V., & Zisserman, A.," Blocks that shout: distinctive parts for scene classification," Computer Vision and Pattern Recognition, vol.9, pp.923-930, 2013.

[23] Mnih, V., Heess, N., Graves, A., & Kavukcuoglu, K., "Recurrent models of visual attention," vol. 3, pp. 2204-2212, 2014.

[24] Ba, J., Mnih, V., & Kavukcuoglu, K., "Multiple object recognition with visual attention". Computer Science, 2014.

[25] Shiping Zhu and Lina Yan, "Local stereo matching algorithm with efficient matching cost and adaptive guided image filter," Visual Computer, vol. 33, no. 9, pp. 1087-1102, 2017.

[26] Shiping Zhu, Ruidong Gao and Zheng Li, "Stereo matching algorithm with guided filter and modified dynamic programming," Multimedia Tools and Applications, vol. 76, no. 1, pp. 199-216, 2017.

[27] Barlow, H. B., "Possible principles underlying the transformation of sensory messages". Sensory Communication, vol. 1, 2012.

[28] Gluckman, J., "Higher order whitening of natural images," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 354-360, IEEE Computer Society, 2005.

[29] Yang, Y., & Ramanan, D., "Articulated human detection with flexible mixtures of parts," IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 35, no. 12, pp. 2878-2890, 2013.

[30] Krause, J., Stark, M., Jia, D., & Li, F. F., "3D object representations for fine-grained categorization," IEEE International Conference on Computer Vision Workshops, pp. 554-561, 2014.

[31] Shiping Zhu, Jiancheng Fang, Rui Zhou, Jianhui Zhao and Wenbo Yu, "A new noncontact flatness measuring system of large 2-D flat workpiece," IEEE Transactions on Instrumentation and Measurement, vol. 57, no. 12, pp. 2891-2904, 2008.

[32] Shiping Zhu and Yang Gao, "Noncontact 3-D coordinates measurement of cross- cutting feature points on the surface of a large-scale workpiece based on the machine vision method," IEEE Transactions on Instrumentation and Measurement, vol. 59, no. 7, pp. 1874-1887, 2010.