

Robust detection algorithm with triple constraints for cooperative target based on spectral residual

Shuai Hao^{1,*}, Yongmei Cheng², and Xu Ma¹

1. School of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710054, China;

2. College of Automation, Northwestern Polytechnical University, Xi'an 710072, China

Abstract: The accurate detection of cooperative targets plays a key and foundational role in unmanned aerial vehicle (UAV) landing autonomously. The standard method based on fixed threshold is too susceptible to both illumination variations and interference. To overcome issues above, a robust detection algorithm with triple constraints for cooperative targets based on spectral residual (TCSR) is proposed. Firstly, by designing an asymmetric cooperative target, which comprises red background, green H and triangle target, the captured original image is converted into a Lab color space, whose saliency map is yielded by constructing the spectral residual. Then, the triple constraints are developed according to the prior knowledge of the cooperative target. Finally, the salient region in saliency map is considered as the cooperative target, and it meets the triple constraints. Experimental results in complex environments show that the proposed TCSR outperforms the standard methods in higher detection accuracy and lower false alarm rate.

Keywords: saliency detection, cooperative target, spectral residual, triple constraints.

DOI: 10.21629/JSEE.2017.04.04

1. Introduction

Computer vision, as a cheap, passive and information-rich source, has been used for different purposes such as collision avoidance [1], aerial refueling [2], spacecraft rendezvous and docking [3] and autonomous landing [4]. In this paper, we focus on how to guide unmanned aerial vehicle (UAV) landing on aircraft carrier autonomously by using vision. This approach firstly uses camera to acquire environment images. Then, we calculate the relative

position and attitude between the cooperative target and UAV by applying image analysis techniques and transformation between different coordinate systems. Finally, it conducts UAV and flight guidance during the landing process by using the calculated flight parameters [5].

It is a critical step to detect cooperative target accurately and robustly when UAV lands on the cooperative target by using vision. Researchers usually tend to design the cooperative target whose color is different from the background. Then they can use the color information of the cooperative target to segment it from the background easily.

The conventional detection method of the cooperative target is the fixed threshold method [6–9]. According to the prior color information, a fixed color threshold is chosen to segment the image. Those image areas which meet the threshold condition are considered to be the candidate cooperative targets, vice versa. Then other prior information of the cooperative target, such as the shape and size information, is used to determine which candidate cooperative target is our desired one. However, the method of setting a fixed color threshold is susceptible to illumination variations and interference. It is not robust and easily makes mistakes in the following image processing. To address this problem, Bi et al. [10] designed an RGB three-channel filter and Hao et al. [11] proposed an adaptive threshold method. The experimental results showed that their methods were superior than the fixed threshold method. However the accuracy of their methods is still dependent on the color threshold selection, and they are not tested in complex environments, such as under the environment of illumination variations. Thus the robustness of their methods has not yet been verified.

In the principle of cooperative target design, it is required that the designed cooperative targets have unambiguous and easy detection characteristics. In other words, the cooperative target should have higher saliency than the background. In computer vision analysis, the saliency

Manuscript received August 29, 2016.

*Corresponding author.

This work was supported by the National Natural Science Foundation of China (61135001), the Scientific Research Program of Shaanxi Provincial Department of Education (16JK1499), the Doctoral Fund of Xi'an University of Science and Technology (2015QDJ007), the Cultivation of Xi'an University of Science and Technology (2014015) and the Ministry of Education Key Laboratory of Information Fusion Technology (LIFT2015-G-1).

is also called as visual attention mechanism. Then if the saliency of the cooperative target is higher than the surrounding environment, we can detect the cooperative target accurately even if there are illumination variations and the interference in the captured image.

Visual saliency has recently drawn great interests because of its usefulness in many applications, such as target detection [12,13], image segmentation [14], target recognition [15] and target tracking [16,17]. Viewed from the information processing perspective, saliency detection methods can be generally divided into two categories [18–21]: top-down (target-oriented) methods and bottom-up (data-driven) methods. Top-down models are slow, attentive, task-dependent and based on prior knowledge. These methods are considered to be a supervised process which utilizes the prior information from the training data, i.e., Yang et al. [22] proposed a saliency model that jointly learned a conditional random field and a discriminative dictionary. However, Cui [23] pointed out that the detection accuracy of the algorithms was mainly decided by the completeness of the selected training samples and they were usually time-consuming. However, the samples are usually not sufficient for training in practical applications. Therefore, considering the complexity of the landing environment and the requirement of real-time, top-down models would not work well. In contrast, bottom-up models are rapid, scene-dependent and based on mathematical methods, which are based on kinds of low-level visual feature and are usually thought to be an unsupervised process. The representative models are Itti [24], graph-based visual saliency (GBVS) [25] and spectral residual (SR) [26], etc. Although the bottom-up models are effective for highlighting the informative regions of images, they may miss the targets which we are interested in due to lack of the top-down prior knowledge.

In general, it is an effective way to identify the cooperative target based on visual attention mechanism. However, the existed models cannot simultaneously meet the detection accuracy, robustness and real-time. In order to solve this problem, a robust detection algorithm with triple constraints for the cooperative target based on spectral residual (TCSR) is proposed, which is robust or insensitive to illumination variations and interference. First, the original image is transformed into a Lab color space. Subsequently, the saliency map of a channel in the Lab color space is obtained by using the spectral residual algorithm. Then, triple constraints are developed according to the prior knowledge of the cooperative target. Finally, the salient region in the saliency map which meets the constraints is the cooperative target. Experiment results in complex environments show that the proposed method has higher detection accu-

racy and lower false alarm rate compared with other three state-of-the-art methods. In this paper, there are two contributions. First, prior knowledge of the cooperative target is integrated into the bottom-up model in the proposed model. It not only has the high detection accuracy and robustness, but also meets the real-time requirement. Second, we provide a fresh thought of detecting the interesting target in complex environments.

The reminder of this paper is structured as follows. Section 2 describes the proposed method with more details. Experimental results and the performance analysis are reported in Section 3. Finally, Section 4 concludes the paper and presents future work.

2. Proposed method

The schematic diagram of the proposed method is shown in Fig. 1.

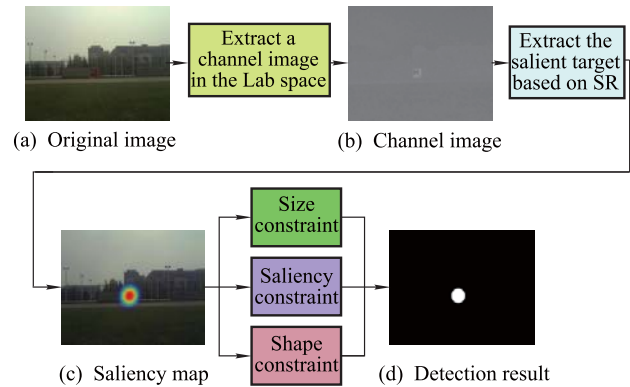


Fig. 1 Schematic diagram of the proposed method

2.1 Cooperative target model

A color cooperative target is designed in this paper, which comprises the red background, a green H and a triangle target, as shown in Fig. 2.



Fig. 2 Cooperative target

2.2 Saliency map generation in a channel

In the Lab color space, a channel represents a color change from magenta to green. In order to enhance the saliency of the cooperative target, the original image is firstly transformed into the Lab color space. Its conversion formula

[27] is as follows:

$$\begin{cases} X = 0.49R + 0.31G + 0.2B \\ Y = 0.177R + 0.812G + 0.011B \\ Z = 0.01G + 0.99B \end{cases} \quad (1)$$

$$\begin{cases} L = 116f(Y) - 16 \\ a = 500 \left[f\left(\frac{X}{0.982}\right) - f(Y) \right] \\ b = 200 \left[f(Y) - f\left(\frac{Z}{1.183}\right) \right] \end{cases} \quad (2)$$

$$f(\varepsilon) = \begin{cases} 7.787\varepsilon + 0.138, & \varepsilon \leq 0.008\ 856 \\ \varepsilon^{1/3}, & \varepsilon > 0.008\ 856 \end{cases} \quad (3)$$

where ε is the independent variable, R , G and B denote the gray value in red, green and blue respectively, X , Y and Z are intermediate variables, L , a and b respectively denote the corresponding gray value of the three channels in the Lab color space. The saliency map of the channel calculated steps as follows:

Step 1 The input image $I(x, y)$ is converted to Lab color according to (1), (2) and (3), and a channel image in the Lab space is recorded as $I_a(x, y)$.

Step 2 $I_a(x, y)$ is transformed into frequency domain $F(u, v)$ through a two-dimensional discrete Fourier transform. Then the amplitude spectrum $A(f)$ and phase spectrum $p(f)$ can be obtained, and the calculating formulas are

$$A(f) = |F[I_a(u, v)]| \quad (4)$$

$$P(f) = \varphi(F[I_a(u, v)]) \quad (5)$$

where $F[\bullet]$ denotes the two-dimensional discrete Fourier transform, $|\bullet|$ denotes amplitude, $\varphi(\bullet)$ denotes phase. Then we take the logarithm of $A(f)$ and $L(f)$ can be obtained as follows:

$$L(f) = \ln(A(f)). \quad (6)$$

Step 3 The spectrum residual $R(f)$ is defined as

$$R(f) = L(f) - \mathbf{h}_n(f) * L(f) \quad (7)$$

where $\mathbf{h}_n(f)$ is an $n \times n$ matrix defined by

$$\mathbf{h}_n(f) = \frac{1}{n^2} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}. \quad (8)$$

Step 4 The saliency map of a channel is given as

$$S(x, y) = g(x) * F^{-1}[\exp(R(f) + P(f))]^2 \quad (9)$$

where $g(x)$ is a Gaussian filter and $\sigma = 5$, F^{-1} denotes the inverse Fourier transform.

2.3 Triple constraints strategy

After obtaining $S(x, y)$, it can be inferred that the higher the value is, the more possible the cooperative target is. Thus we consider those areas in $S(x, y)$ with high saliency value as candidate cooperative targets. Due to the influence of illumination variations and interference, some background image areas show high saliency and they are needed to be further identified. Therefore, according to the prior knowledge of the cooperative target, a triple constraints strategy is developed.

2.3.1 Size constraint

The airborne camera is adjusted to down view when capturing images. The height of the UAV from the ground could be gotten by an airborne altimeter, such as radio altimeter, which is considered to be the distance between airborne camera and the plane of the cooperative target. On the premise that we have known the size of the cooperative target and the distance between the airborne camera and the plane of the cooperative target, the occupied area of the cooperative target in the image, which is denoted as A , can be calculated based on the principle of pinhole imaging. Assume the camera resolution is $m \times n$, focus is f , horizontal field of view is α , vertical field of view is β , length and width of cooperative target are l and w respectively, and the distance between airborne camera and plane of cooperative target is d . Based on the principle of pinhole imaging, pixel numbers of the cooperative target along the horizontal (denoted as u) and vertical (denoted as v) coordinates in the image coordinate are defined as

$$u = \frac{wm}{2d \tan\left(\frac{\alpha}{2}\right)} \quad (10)$$

$$v = \frac{ln}{2d \tan\left(\frac{\beta}{2}\right)} \quad (11)$$

and the circumradius of the cooperative target occupied pixel number (denotes as r) are defined as

$$r = \sqrt{u^2 + v^2}/2. \quad (12)$$

2.3.2 Saliency constraint

It can be assumed that the cooperative target image region has the higher saliency than other image regions. However, due to the complexity of the landing environment, there may be other regions which show strong saliency in the saliency map. In this paper, the first three maximum areas in $S(x, y)$ are considered to be the candidate cooperative targets. And the coordinate positions of three candidate cooperative targets are denoted as $M_i(x, y)$ ($i = 1, 2, 3$). Then we truncate three circular regions in $S(x, y)$, where

$M_i(x, y)$ is the center, r is the radius and the saliency map of the three circular regions is denoted as $S_i(x, y)$ ($i = 1, 2, 3$). We calculate the mean values of the three circular regions, arranged in descending order and recorded as N_i ($i = 1, 2, 3$).

2.3.3 Shape constraint

As seen from Fig. 1, the cooperative target shows the obvious shape characteristics. Considering that Hu moments have good real-time property and are suitable to describe the target's shape characteristics, we use them to determine which candidate cooperative target in N_i is our desired one. In order to improve the real-time of the proposed algorithm, we just calculate the first three Hu moments. Considering that Hu moments may be negative and have different orders of magnitude in the moments, we take the absolute value of Hu moments and take the logarithm of them.

Euclidean distance of Hu moments between the candidate cooperative target and the real one is defined as

$$d = \sqrt{(\phi_1 - \varphi_1)^2 + (\phi_2 - \varphi_2)^2 + (\phi_3 - \varphi_3)^2} \quad (13)$$

where ϕ_1 , ϕ_2 and ϕ_3 denote the first three Hu moments of the candidate cooperative target, φ_1 , φ_2 and φ_3 denote the first three Hu moments of the real cooperative target.

When the calculated d of candidate cooperative target satisfies the threshold condition, we can assume that it is our desired cooperative target.

In addition, in order to ensure the integrity of the cooperative target structure when detecting, radius of the candidate cooperative target is selected a little bigger than the calculated value.

3. Experiments and analysis

To demonstrate the effectiveness and robustness of the proposed method, the proposed method is tested in the three types of complex environments, and the proposed method is compared with three existing representative models, including the Itti, GBVS and SR.

3.1 Weak illumination

The size of the cooperative target used in the experiment is $2 \text{ m} \times 2 \text{ m}$. When the camera is 35 m away from the cooperative target, we acquire the original image under the weak illumination condition, as is shown in Fig. 3(a). Saliency maps are generated respectively by GBVS, Itti, SR and the proposed method in this paper (abbreviated TCSR) are shown in Figs. 3(b)–(e). In saliency maps, image colors are the more “red”, the higher saliency. Then we respectively extract the salient target and the results are shown in Figs. 3(g)–(j).

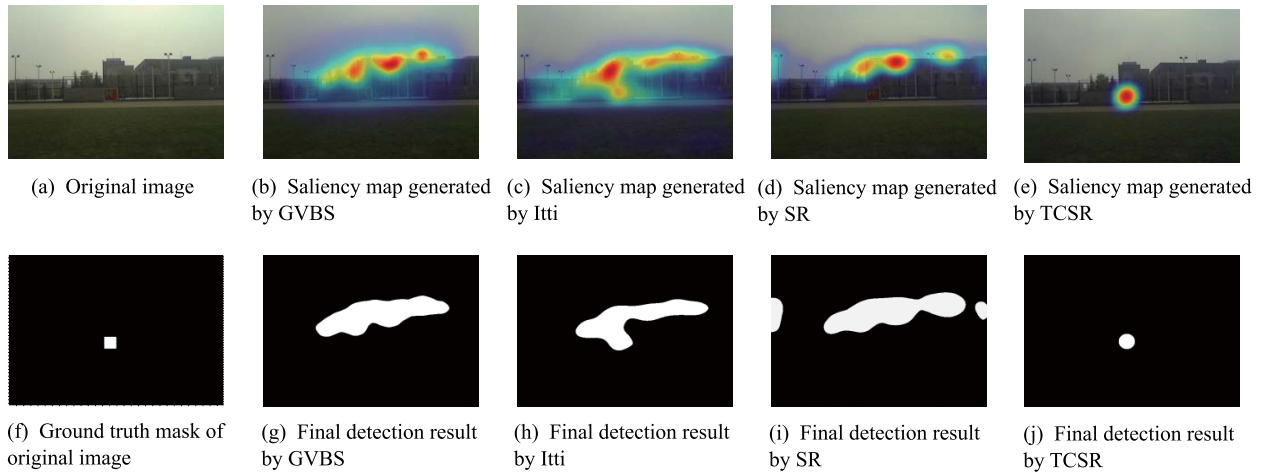


Fig. 3 Detection results of different methods under weak illumination conditions

It can be seen from Fig. 3 that the proposed method can extract the cooperative target accurately. However, it should be noted that in order to ensure the integrity of the detected cooperative target, we use a circular area to extract the cooperative target. Compared to the actual target area, the cooperative target in the saliency map is slightly larger than the actual one.

3.2 Strong light

The size of the cooperative target used in the experiment is

$5.5 \text{ cm} \times 5.5 \text{ cm}$ and it is placed on the aircraft carrier model. When the camera is 1.5 m away from the cooperative target, we acquire the original image under strong light.

Fig. 4(a) shows the original image of the cooperative target under the strong illumination condition. Figs. 4(b)–(e) are the saliency maps generated by GBVS, Itti, SR and TCSR respectively. Fig. 4(f) is the corresponding ground truth mask of the original image. The final detection re-

sults of above four methods are shown in Figs. 4(g)–(j). As can be seen from Fig. 4, similar to the strong illumi-

nation condition, the cooperative target shows the highest saliency in the proposed model.

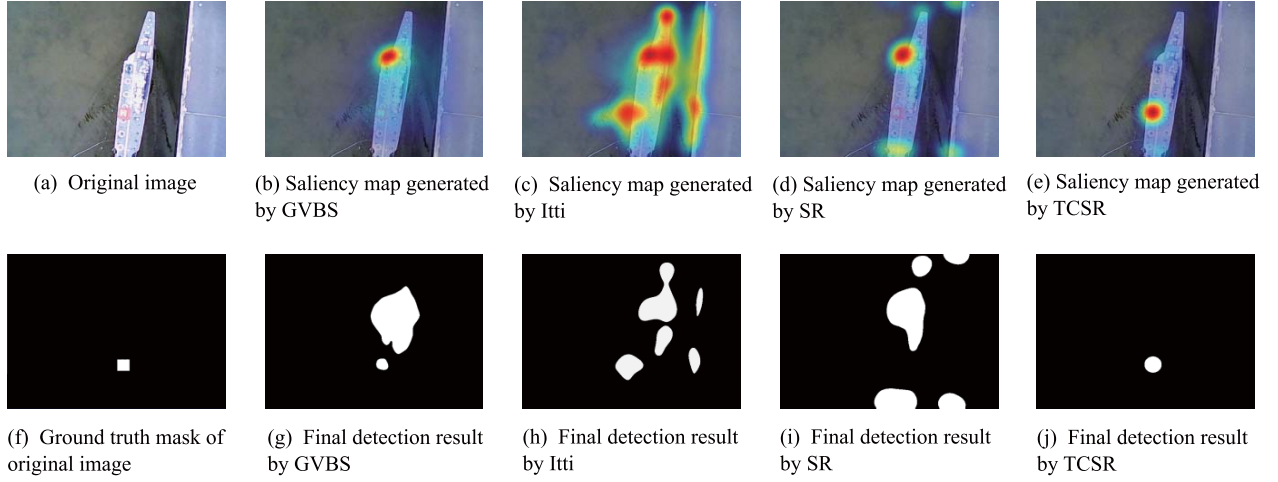


Fig. 4 Detection results of different methods under strong light conditions

3.3 Presence of interference

In order to verify the robustness of the proposed method, a red background which has the same size with the coope-

rative target is added based on the weak illumination condition, and related experimental results are shown in Fig. 5.

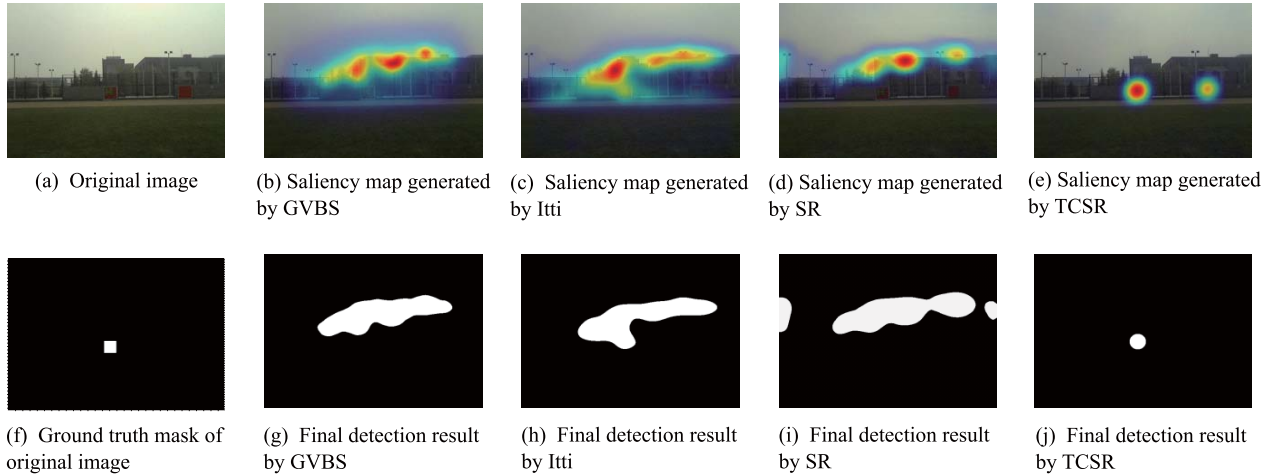


Fig. 5 Detection results of different methods in the presence of interference conditions

As seen from Fig. 5, in the presence of interference conditions, although two part image regions show the higher salient in the proposed model, we can accurately extract the cooperative target with the triple constraints strategy.

As seen from the cooperative target detection experiments under three complex environments, the proposed method has strong robustness and high accuracy.

3.4 Performance comparison

In order to evaluate the detection performance, we adopt the commonly used performance measures, i.e., false

alarm rate (FAR), precision and recall, which are defined as follows:

$$\text{FAR} = E \left(\prod_k (1 - O_k(x)) S_k(x) \right) \quad (14)$$

$$\text{Precision} = \sum_k O_k(x) S_k(x) / \sum_k S_k(x) \quad (15)$$

$$\text{Recall} = \sum_k O_k(x) S_k(x) / \sum_k O_k(x) \quad (16)$$

where $O_k(x)$ denotes the k th binarized image of a real target, $S_k(x)$ denotes the saliency map after binarization,

$\prod_k(\bullet)$ denotes even multiply. One hundred cooperative target images which are obtained in complex environments are selected to be tested. The performance comparison of four methods is shown in Fig. 6.

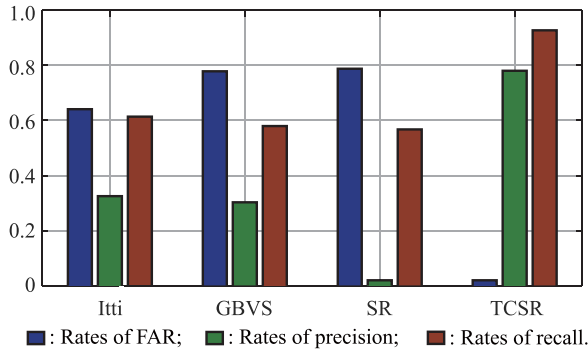


Fig. 6 Performance comparison of different algorithms

In Fig. 6, compared with the other three algorithms, the propose method has the lowest FAR, the highest precision and recall. We have also tested the average running time of the four algorithms on a computer with Inter Core 2 Duo 2.13 GHz CPU and 2 GB RAM, as reported in Table 1.

Table 1 Average running time of different methods on source images with sizes of 640×480

Algorithm	Time/s	Code
Itti	0.424 4	MATLAB
GBVS	1.312 7	MATLAB
SR	0.083 6	MATLAB
TCSR	0.352 1	MATLAB

In Table 1, the SR algorithm has the best real-time, and the real-time performance of our proposed algorithm is better than the other two algorithms. Overall considering the detection accuracy and real-time, the proposed has the best real-time, and the real-time performance has more advantages and is more attractive. Experimental results under complex environments show that the proposed method not only has strong robustness and high accuracy, but also has a good real-time property.

4. Conclusions

In this paper, a novel cooperative target detection algorithm is proposed. In the proposed algorithm, the original image is first converted into the Lab color space and a channel saliency map is generated based on SR, then a triple constraint strategy is developed according to the prior knowledge of the cooperative target. Finally, the image region in the saliency map which meets the triple strategy is considered to be the desired cooperative target. Experimental results in complex environments demonstrate that the proposed method can outperform other three state-of-the-art

methods. This research provides a new idea for how to detect the cooperative target accurately and robustly in complex environments by using vision.

In future work, we intend to incorporate the structure information of the cooperative target to further improve the detection accuracy.

References

- [1] A. Mcfadyen, L. Mejias, P. Corke, et al. Aircraft collision avoidance using spherical visual predictive control and single point features. *Proc. of the IEEE International Conference on Intelligent Robots and Systems*, 2013: 50–56.
- [2] C. Martinez, T. Richardson, P. Thomas, et al. A vision-based strategy for autonomous aerial refueling tasks. *Robotics and Autonomous Systems*, 2013, 61(8): 876–895.
- [3] B. Qiao, S. Tang, K. Ma, et al. Relative position and attitude estimation of spacecrafts based on dual quaternion for rendezvous and docking. *Acta Astronautica*, 2013, 91: 237–244.
- [4] S. Yang, S. A. Scherer, A. Zell. An onboard monocular vision system for autonomous takeoff, hovering and landing of a micro aerial vehicle. *Journal of Intelligent & Robotic Systems*, 2013, 69: 499–515.
- [5] C. J. Wu, W. H. Tsai. An omni-vision based localization method for automatic helicopter landing assistance on standard helipads. *Proc. of the IEEE International Conference on Computer and Automation Engineering*, 2010: 327–332.
- [6] S. Saripalli, G. S. Sukhatme. Landing a helicopter on a moving target. *Proc. of the IEEE International Conference on Robotics and Automation*, 2007: 2030–2035.
- [7] J. H. Choi, W. S. Lee, H. Bang. Helicopter guidance for vision-based tracking and landing on a moving ground target. *Proc. of the IEEE International Conference on Control, Automation and Systems*, 2011: 867–872.
- [8] D. Lee, T. Ryan, H. J. Kim. Autonomous landing of a VTOL UAV on a moving platform using image-based visual serving. *Proc. of the IEEE International Conference on Robotics and Automation*, 2012: 971–976.
- [9] X. X. Wang, Y. Liang, Q. Pan, et al. Measurement random latency probability identification. *IEEE Trans. on Automatic Control*, 2016, 61(12): 4210–4216.
- [10] Y. Bi, H. Duan. Implementation of autonomous visual tracking and landing for a low-cost quadrotor. *Optik-International Journal for Light and Electron Optics*, 2013, 124(18): 3296–3300.
- [11] S. Hao, Y. Cheng, X. Ma, et al. Robust corner precise detection algorithm for visual landing navigation of UAV. *Systems Engineering and Electronics*, 2013, 35(6): 1262–1267. (in Chinese)
- [12] J. Han, Y. Ma, B. Zhou, et al. A robust infrared small target detection algorithm based on human visual system. *IEEE Geoscience and Remote Sensing Letters*, 2014, 11(12): 2168–2172.
- [13] M. M. Cheng, N. J. Mitra, X. L. Huang, et al. Global contrast based salient region detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 569–582.
- [14] C. Jung, C. Kim. A unified spectral-domain approach for saliency detection and its application to automatic object segmentation. *IEEE Trans. on Image Processing*, 2012, 21(3): 1272–1283.
- [15] S. Lee, J. Oh, J. Park, et al. A 345 mW heterogeneous many-core processor with an intelligent inference engine for robust object recognition. *IEEE Journal of Solid-State Circuits*, 2011, 46(1): 42–51.

- [16] Y. Su, Q. Zhao, L. Zhao, et al. Abrupt motion tracking using a visual saliency embedded particle filter. *Pattern Recognition*, 2014, 47(5): 1826–1834.
- [17] V. Mahadevan, N. Vasconcelos. Biologically inspired object tracking using center-surround saliency mechanisms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2013, 35(2): 541–554.
- [18] Y. Xie, H. Lu, M. H. Yang. Bayesian saliency via low and mid level cues. *IEEE Trans. on Image Processing*, 2013, 22(5): 1689–1698.
- [19] L. Zhang, L. Yang, T. Luo. Unified saliency detection model using color and texture features. *Plos One*, 2016, 11(2): 1987–1989.
- [20] J. Yu, J. Tian. Saliency detection using midlevel visual cues. *Optics Letters*, 2012, 37(23): 4994–4996.
- [21] X. X. Wang, Y. Liang, Q. Pan, et al. Design and implementation of Gaussian filter for nonlinear system with randomly delayed measurements and correlated noises. *Applied Mathematics and Computation*, 2014, 232(1): 1011–1024.
- [22] J. Yang, M. H. Yang. Top-down visual saliency via joint CRF and dictionary learning. *Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2012: 2296–2303.
- [23] X. Cui, Y. Tian, L. Ma. Top-down visual saliency detection in optical satellite images based on local adaptive regression kernel. *Journal of Multimedia*, 2014, 9(1): 173–180.
- [24] L. Itti, C. Koch, E. Niebure. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1998, 20(11): 1254–1259.
- [25] J. Harel, C. Koch, P. Perona. Graph-based visual saliency. *Proc. of the Advances in Neural Information Processing Systems*, 2006: 545–552.
- [26] X. Hou, L. Zhang. Saliency detection: a spectral residual approach. *Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2007: 1–8.
- [27] X. M. Xiao, Z. J. Min, J. M. Yu. Color image segmentation

based on HIS and LAB color space. *Journal of Guangxi University: Natural Science Edition*, 2011, 36(6): 976–980.

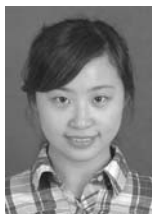
Biographies



Shuai Hao was born in 1986. He received his Ph.D. degree from School of Automation, Northwestern Polytechnical University in 2014. He is currently a lecturer in Xi'an University of Science and Technology. So far he has published more than 20 papers, in which two papers were indexed by SCI and ten papers were indexed by EI. His research interests are vision navigation and digital image processing.
E-mail: hsh000@163.com



Yongmei Cheng was born in 1960. She received her Ph.D. degree from School of Automation, Northwestern Polytechnical University in 2001. She now is a professor in School of Automation, Northwestern Polytechnical University. So far she has published more than 100 papers, in which 20 papers were indexed by SCI and 40 papers were indexed by EI. Her research interests are vision navigation, digital image processing and information fusion.
E-mail: chengym@nwpu.edu.cn



Xu Ma was born in 1985. She received her Ph.D. degree from School of Automation, Northwestern Polytechnical University in 2017. She is currently a lecturer in Xi'an University of Science and Technology. So far she has published more than 15 papers, in which two papers were indexed by SCI and eight papers were indexed by EI. Her research interests are vision navigation and digital image processing.
E-mail: maxucat@163.com