http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (20): 5914-5921, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.5914.5921

New Method of Sparse Visual Saliency Feature Extraction and Application in Unmanned Vehicle Environment Sensing

Du Ming-Fang, Wang Jun-Zheng, Li Jing and Cao Hai-Qing
Key Laboratory of Complex System Intelligent Control and Decision
Beijing Institute of Technology, Ministry of Education, 100081, Beijing, China

Abstract: A new method based on Hessian matrix threshold of finding local low-level saliency features is proposed in this study after the standard local invariant feature extraction algorithm SRUF (Speeded Up Robust Features) is analyzed. In this method, the number of saliency feature points can change with the change of Hessian threshold. The saliency feature points will become sparser when Hessian threshold becomes larger. When a certain extreme threshold which is defined as Hessian Threshold Node is reached, the retained discriminative feature points are remarkable stability characteristics, also make up the best sparse saliency features set. This method is applied in the unmanned vehicle environment sensing system to help extract the before going vehicle's saliency feature and realize object tracking and obstacle detection. Experiment results show that this is an quick and robust method to determine saliency feature points quantitatively and is very suitable for the occasion which has strong demand on real-time.

Key words: Hessian threshold, visual saliency feature, sparse SURF feature, object detection, unmanned vehicle

INTRODUCTION

In the image processing, analysis and understanding fields which are related to motion, such as robotic vision and cognition, object tracking, visual navigation, etc., there exists a strong demand on processing speed which poses challenges for the feature extraction methods. An ideal feature extraction method should satisfy the requirements of sparsity and stability simultaneously. Sparsity can guarantee the real-time property of image processing algorithms, while stability can reduce the error generated in the motion process. Taking object tracking as an example, when tracking feature points, if we extract too dense feature points, the algorithm speed will be much affected. If we extract too sparse feature points, the tracking will become unstable or even fail. How sparse should be the feature points extracted is still not solved. Saliency feature is a good conception previously proposed which satisfies the requirements relatively well. It generally describes in the following aspects: (1) Local low-level feature considerations; (2) Global considerations; (3) Visual organization rules; (4) High-level prior. In this study, we try to consider from the local low-level features and propose a method which can determine the sparse and stable feature points by using the Hessian threshold in SURF algorithm.

The proposed method is based on the local invariant features in multi-scale space. Applying the method to the unmanned vehicle we developed to realize the before going vehicle's feature detection and tracking. Experiment results show that the extracted saliency feature points are very stable and quick, so prove this is a new effective environment sensing method for autonomous mobile robot.

RELATED WORKS ON VISUAL SALIENCY FEATURE

The visual attention mechanism can help filter invalid interference information and simplify the image understanding process. Visual saliency is a kind of local contrast reflected in images. The sharper the contrast is, the stronger the saliency is. C. Koch and S. Ullman proposed a very influential and theoretical visual attention model based on saliency, the bio-inspired model (Koch and Ullman, 1985; Koch and Poggio, 1999). Itti and Koch proposed its computational model. The saliency features can be defined from these models. That is, fusing multiple low-level features into an overall saliency measurement. It is clear that saliency features are the best measurements for representing the essence of images. The visual salience map, which shows how the attention

Corresponding Author: Du Ming-Fang, Key Laboratory of Complex System Intelligent Control and Decision Beijing Institute of Technology, Ministry of Education, 100081, Beijing, China

is attracted, can be obtained from saliency features. However, how to quantitatively represent and acquire it is difficult.

Previous typical methods proposed include: R. Achanta *et al*., used (1) To express saliency feature (Radhakrishna *et al*., 2009; Zhai and Shah, 2006).

$$S(x,y) = \left\| I_{\mu} - I_{\omega hc}(x,y) \right\| \quad (1)$$

where, Iμ is the feature vector of the mean image, $I_{\omega hc}$ is the smoothed image by Gaussian filter. ‖•‖ represents 2-norm.

Y. Zhai, *et al*., used (2) to express the saliency feature 4.

$$SalS(I_k) = \sum_{\forall I_i \in I} \left\| I_k - I_i \right\| \quad (2)$$

where, the range of I_i is (0, 255). ‖•‖ represents color distance metric.

Prof. Itti's group in iLab, University of South California has carried out robotic visual localization research based on biological inspiration (Itti and Baldi, 2005; Siagian and Itti, 2008). Harel J. in Koch Lab proposed visual saliency detection based on graph in 2006 (Harel *et al*., 2007). Y. Zhang *et al*. proposed to use multiple methods, such as wavelet and motion estimation, etc., to extract sensitive features on multiple scales, then fuse these features to obtain an overall dynamic saliency feature. Finally the multiple-object tracking was realized by combining the dynamic saliency feature with particle filter algorithm (Zhang *et al*., 2008a, b; Chen *et al*., 2010a) proposed to use minimum error rate criteria to describe the saliency of feature and applied it to vehicle license plate localization and vehicle type recognition (Chen *et al*., 2010a, b). Currently, saliency features have been applied successfully in many fields, including image registration, object tracking, human detection and natural scene recognition, etc.

HESSIAN MATRIX

Brief introduction of SURF algorithm: SURF (Speeded Up Robust Features) was first proposed by H. Bay et al. in the conference of ECCV 2006 (Herbert *et at*., 2006), then in the journal of CVIU in 2008 (Herbert *et at*., 2008). It can be seen from Bay's experiments that the speed of SURF is 5-0 times faster than SIFT's with equal accuracy. As a high performance local feature descriptor, SURF is robust to rotation, scale changes, occlusion, illumination and view angle changes, as well as affine transformations. Of them, the rotation and scale invariant is supported by Hessian matrix. The features of SURF are: using integral image to convolution images; using Hessian matrix to detect feature values; using distribution-based descriptors to extract local features. The algorithm steps of SURF are as follows: construct Hessian matrix, construct Gaussian pyramidal scale space, identify extreme points according to non-maxima suppression, accurately localize feature points, determine the dominant direction of feature points and construct SURF local feature descriptor. Hessian matrix determinant approximation image is used in SURF algorithm. Therefore, the Hessian matrix is the core of SURF algorithm. What feature points are reserved, or what low-level saliency features are reserved, is related to image pyramid and Hessian matrix.

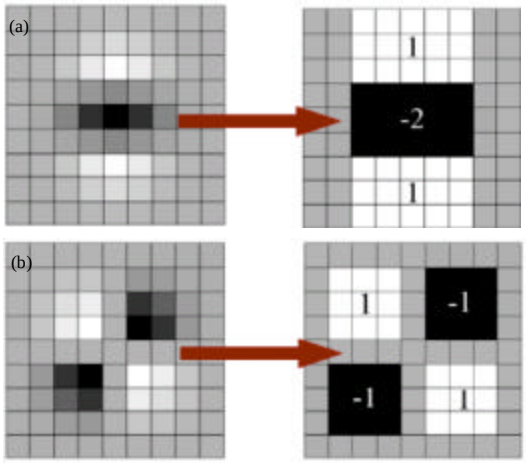

Fig. 1: Template simplification (Taking a 3×3 template as an example), (a) simplification of L_{yy} and (b) simplification of L_{xy}

Image pyramid: SURF detects feature points by searching extreme value in an image pyramid space. The image pyramid generating method which SURF uses is to continuously increase the size of the image kernal, thereby increase the size of box filter template, which indirectly establish the image pyramid. To speed up the algorithm, an approximate simplified template is constructed. The simplification method for the template along Y-coordinate is shown in Fig. 1a. The simplification method for second-order mixed derivative template is shown in Fig. 1b.

If the template box takes a value of $i^n \in \{1,-1-2\}$, the 4 corner points of the integral image are $\{p_1^n, p_2^n, p_3^n, p_4^n\}$ and the areas of the box are S_{xx}, S_{yy} and S_{xy}. Therefore, the responses of the box filters are:

$$D_{xx} = \frac{1}{S_{xx}} \sum_{n=1}^{3} i^n (p_4^n - p_2^n - p_3^n + p_1^n) \quad (3)$$

$$D_{yy} = \frac{1}{S_{yy}} \sum_{n=1}^{3} i^n (p_4^n - p_2^n - p_3^n + p_1^n) \quad (4)$$

$$D_{xy} = \frac{1}{S_{xy}} \sum_{n=1}^{4} i^n (p_4^n - p_2^n - p_3^n + p_1^n) \quad (5)$$

Using the simplified template to convolute the image, a speckle response image on a certain scale is obtained. By using templates on different scales, we can obtain the multi-scale speckle response image pyramid, which is shown in Fig. 2. That is the searching space for extreme values.

Hessian matrix: Consider a pixel x = (x, y) in image I. The Hessian matrix at this pixel with scale σ is defined as:

$$? (x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (6)$$

where, $L_{xx}(x, \sigma)$ is the convolution of the Gaussian second derivative:

$$\frac{\partial^2}{\partial x^2} g(\sigma)$$

And the image I at point x. The meanings of $L_{xy}(x, \sigma)$ and $L_{yy}(x, \sigma)$ can be derived using this method. The determinant of the Hessian matrix is:

$$\det(H) = L_{xx}L_{yy} - L_{xy}L_{xy} \quad (7)$$

Using the approximate values $L_{xx}(x, \sigma)$, $L_{xy}(x, \sigma)$, $L_{yy}(x, \sigma)$ replace D_{xx}, D_{yy}, D_{xy}, the Hessian's determinant can be described as follows:

Fig. 2: Image pyramid established by SURF

$$\det(H_{approx}) = D_{xx}D_{yy} - (\omega D_{xy})^2 \quad (8)$$

where, ω is the relative weight of the filter response, which balances the expression of Hessian's determinant. In practice, it is set to 0.9. SURF use the Hessian's determinant as the speckle response of each point in an image. When implementing the SURF algorithm, a Hessian threshold is set, which determines the number of reserved speckles.

UNMANED VEHICLE VISION SENSING EXPERIMENTS

Application background of unmanned vehicle: In environment sensing system of unmanned vehicle, vehicles in front, pedestrians and other moving objects can be seen as dynamic obstacles. Extracting real-time visual features to realize obstacle avoidance is a key technology problem. The correct choice of features is critical. In this study, C 30 vehicle produced by Beijing Automotive Group is modified as unmanned vehicle and as our research and experimental platform. The experiment video is captured from Beijing urban roads. The above algorithm is used to detect and track moving obstacles ahead. The experiment platform is shown as Fig. 3.

SURF feature extraction: CPU of the algorithms performance hardware platform is configured as Intel i5,

Fig. 3: Application background, (a) Unmanned Vehicle altered by C30 and (b) Vision field of the Unmanned Vehicle (From the city road of Beijing)

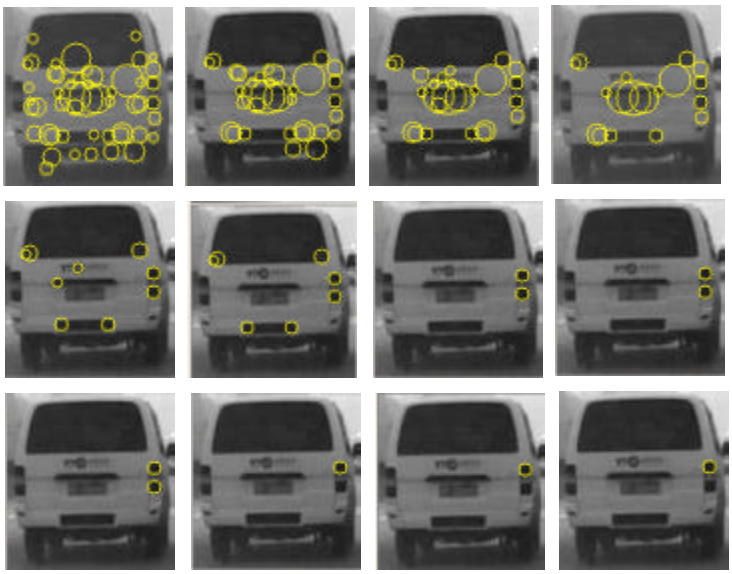

Fig. 4: Feature extraction results under different hessian thresholds (hessian thresholds and feature point numbers of the first to the last image are: (100, 51), (300,33), (600,24), (1000,18), (1500,9), (2000,7), (2500,2), (3000,2), (3500,2), (4000,1), (4500,1) and (5000,1))

clocked at 2 GHz.The SURF feature extraction results on sample image under different Hessian thresholds are shown in Fig. 4.

The feature extraction results under different Hessian thresholds reflect a truth, that is, the saliency features are closely related to the Hessian threshold. The higher the Hessian threshold is, the sparser the feature is. The reserved sparse features are most discriminative visual features in the low-level. It can be seen from the filtering effect of the Hessian threshold that the saliency features concentrate in the areas where intensities change drastically. For instance, the SURF features on the vehicle taillights in the sample image in Fig. 3 can be reserved when the Hessian threshold is larger than 2500. This also shows that, in cases which demands for high real-time requirement, we can first preprocess the image by detecting the edges or corners features, then extract the sparse saliency SURF features for tracking.

A new conception Hessian Threshold Node can be defined according to the experimental results. The Hessian threshold node is a Hessian threshold that makes the number of SURF feature points approach a constant small value, such as 2500 and 4000 in the above experimental vehicle image.

Feature matching based on Sparse saliency SURF: To test the effectiveness of the saliency sparse features extracted by the above method, we select different Hessian thresholds for object recognition. Saliency feature matching is used to search sample vehicle object in ROI of the road scene image (PNG format with 512*288 pixels, 222K). The experimental results are shown in Fig. 5.

The most discriminative feature points are the feature points finally reserved. These feature points can best ensure the recognition of objects in scenes. It is clear in Fig. 4 that, when 4500 is used for the Hessian thresholds, respectively, the locations of extracted SURF feature points in an image are the same as when 5000 is used. So the relative feature points are the most discriminative features.

Some typical datum of number of feature points and the recognition time under different Hessian thresholds are tabulated in Table 1.

Table 1 shows that when feature points are sparse enough, the feature is stable if the Hessian threshold is larger than a Hessian Threshold Node (e.g. 2500 and 4000 in Table 1). Increasing the Hessian threshold does not have much effect on detection results, that is, the results have no major changes.

To explain the characters and the meanings of the data in Table 1 clearly, curves showing the relation of data are drawn in Fig. 6 and Fig. 7.

It can be seen in Fig. 6 that the general trend is that the larger the Hessian threshold is, the faster the feature matching is and the faster the scene feature extraction is.

Fig. 5: Feature matching results under different hessian thresholds (hessian thresholds from the first to the last image are: 1500, 2000, 2500, 3000 , 3500, 4000, 4500, 5000)

Table 1: No. of feature points and recognition time con SUMPTION under different hessian thresholds

	Threshold								
Parameters	600	1500	2000	2500	3000	3500	4000	4500	5000
Osum	24	9	7	2	2	2	1	1	1
Ssum	143	58	40	27	23	17	13	9	9
Ts	29	26	25	26	28	24	24	20	19
Tp	17	6	4	2	1.8	1.3	0.9	0.7	0.62
Pair	46	18	14	4	4	4	2	2	2

When the threshold is exceeded, the change of time consumption is not significant and the time consumption is nearly stable.

The relation between the numbers of feature points, which are from the sample image and the scene image and the Hessian thresholds is shown in Fig. 7.

It can be seen in Fig. 7 that the larger the Hessian threshold is, the less the SURF feature points can be extracted. The number of feature points changes until the Hessian threshold reaches 4000 and it maintains a constant value of 2. Furthermore, the number of feature points of the sample object has the same change trend with that of the scene and the matching of point pairs is among these comparable points. The ideal matching, or 100% matching, should happen when the number of feature points of the object is exactly equal to that of the scene. It is exactly satisfied when the Hessian threshold reach its Nodes. Experimental results also proved that 100% matching can be achieved when the Hessian threshold is at its Nodes.

It can be seen from the experimental results that the Hessian threshold helps find the sparsest and most stable

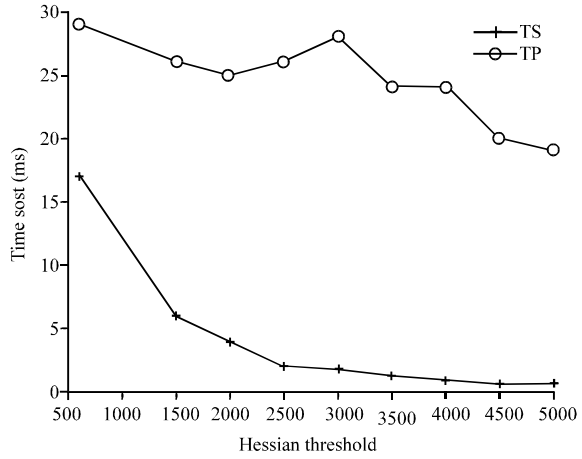

Fig. 6: Relation between Hessian threshold and recognition time

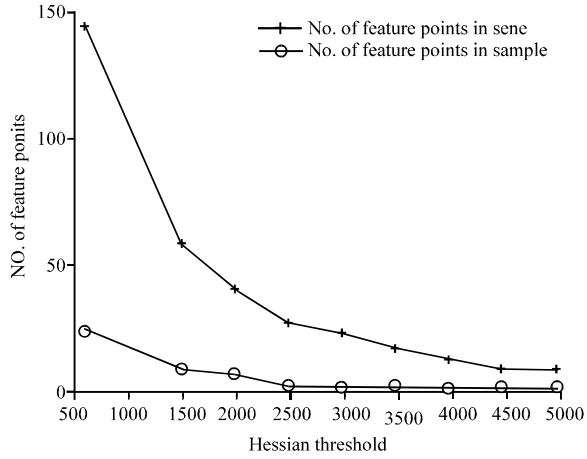

Fig. 7: Relation curve between feature numbers and Hessian thresholds

SURF features. Under this circumstance, the problem of matching error does not exist, because the number of feature points is few enough and the features are the most stable. It is a very valuable conclusion that the most salient low-level features can be determined by adjusting Hessian threshold Nodes, because finding saliency features itself is a difficult yet important work.

Robust matching experiments: The general feature robustness testing criteria are that whether the feature has scale invariance, rotational invariance, illumination invariance, affine invariance. This study only discusses scale changes, illumination changes two situations encountered in the target detection and tracking of unmanned vehicle. In fact, SURF algorithm itself has solved the problem of rotational invariance.

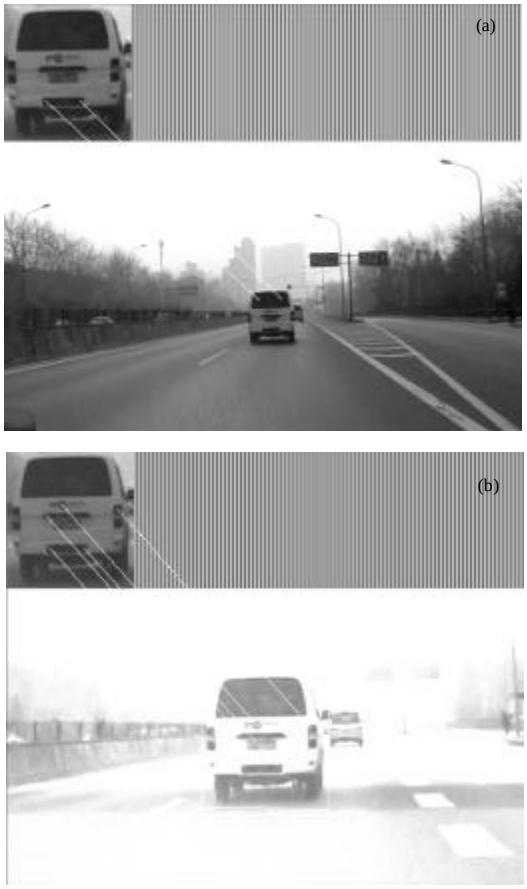

Fig. 8(a-b): Sparse saliency features matching when object size changed, (a) Object size changed greatly, Low illuminance and (b) Object size changed slightly, High illuminance

Still choosing the above sample image used in the above experiment but the target becomes smaller in the scene image (larger difference in size compared with the sample). This often occurs in the process of moving target tracking, in other words, the target size is a random variation. The saliency sparse features are adopted to match the sample and the object in scene with a Hessian threshold 1500 and the particle filter is selected to correct errors. The experiment result shows that the object can be tracked stably. The sparse features extraction result and the matching result are shown in Fig. 8.

Although sparse features help to significantly improve the system real-time but when the feature is too sparse, it will produce feature does not match and the consequences of missing the target.

It should be noted that described Hessian threshold for extracting significant features is a relative value, not

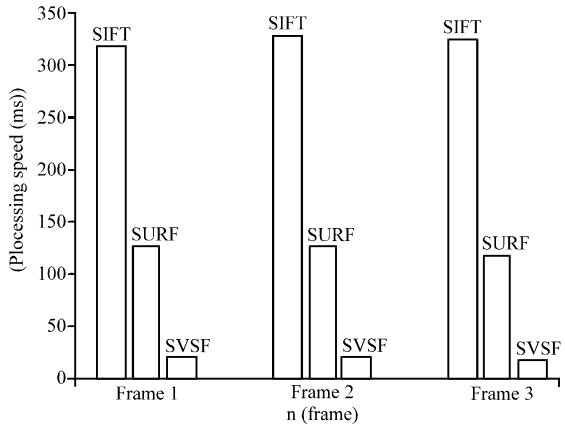

Fig. 9: Algorithm speed contrast

Table 2: No. of feature points and recognition time under different hessian thresholds

Algorithm	Time	Scale	Rotation	Blur	Illumination	Affine
SIFT	Common	Best	Best	Common	Common	Good
SURF	Good	Best	Good	Good	Best	Good
SVSF	Best	Best	Good	Good	Best	Good

absolute. In different scenes, or the target scale changes (compared with the sample) obviously, Hessian threshold will change and there is no fixed standard. So, in target detection and tracking applications, the real Hessian threshold is needed to adjust but not too often.

ALGORITHM EVALUATION

In Table 2, SVSF(Sparse Visual Saliency Feature) is compared with other robust feature extraction algorithms proposed in nearly years.

SIFT, SURF and SVSF in circumstances of scale changes all have quite good invariance effect. SURF and SVSF have the best matches result in large variations of brightness, have better effect than SIFT when image is blurring and have a somewhat poor effect in rotation invariance.

To the continuous 3 frames of unmanned vehicle visual sensing images, run the above three algorithms successively, the feature extraction and matching speeds are shown in Fig. 9.

So judging from the speed, SURF is about three times quick than SIFT and SVSF is about 6 times quick than SURF if the features are sparse as the original's 1/3.

In unmanned vehicle's visual target detection and tracking, it has not so high demand on the rotational invariance while has a strong demand on real-time and scale invariance. Because of the vibration in motion process, image will become blur. So the system has a demand on the robustness of feature extraction when image is blur. From the foregoing analysis the conclusion can be drawn: SVSF is best local invariant feature extraction algorithm for unmanned vehicles visual environment sensing.

CONCLUSION

A new method of low-level saliency feature extraction based on local invariant feature is proposed in this study. The effectiveness of the method is demonstrated by unmanned vehicle environment sensing experiments. The conclusion is also very valuable that the Hessian threshold nodes are in correspondence with the most salient and sparse feature set, which can be directly applied to the fields of robust object tracking and recognition, etc. In recent years, the Deformable Part Model by latent SVM learning method has abstracted the researchers' attentions all over the world (Pedro *et al*., 2008) and the author has been awarded "Lifetime Achievement" prize. This gives an idea for saliency feature extraction. So this paper's future work is to find the key nodes of Hessian threshold of the stable and sparse feature points by using machine learning method and automatically choosing proper key points of Hessian threshold to obtain the most reasonable saliency feature descriptor.

ACKNOWLEDGMENT

The authors would like to thank for the support by the National Natural Science Foundation of China "Research on multi-camera multi-target tracking algorithm based on compressed sensing" (61103157) and the Beijing Municipal Education Commission project SQKM201311417010 and PHR201107149.

REFERENCES

Bay, H., A. Ess, T. Tuytelaars and L. Van Gool, 2008. SURF: Speeded up robust features. Comput. Vision Image Understanding, 110: 346-359.

Chen, Z.X., C.Y. Liu and F.L. Chang, 2010. Vehicle Type Recognition Based on Biological Vision Salience. Comput.Sci., 37: 207-208.

Chen, Z.X., F.L. Chang and C.Y. Liu, 2010. Multi-features fusion license plates locating algorithm based on feature. Control Decision, 25: 1909-1912.

Felzenszwalb, P., D. McAllester and D. Ramanan, 2008. A discriminatively trained, multiscale, deformable part model. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, Anchorage, AK., 1-8.

Harel, J., C. Koch and P. Perona, 2007. Graph-based visual saliency. Adv. Neural. Info.Process. Syst., 19: 545-552.

Herbert, B., T. Tuytelaars and L.V. Gool, 2006. SURF: Speeded Up Robust Features. Comput.Vision, ECCV 2006 Lect. Not. Comput. Sci., 3951: 404-417.

Itti, L. and P. Baldi, 2005. Bayesian surprise attracts human attention. Adv. Neur. Info. Process Syst., 19: 547-554.

Koch, C. and S. Ullman, 1985. Shifts in selective visual attention: Towards the underlying neural circuitry. Human Neurobiol., 4: 219-227.

Koch, C. and T. Poggio, 1999. Predicting the Visual World: Silence is Golden. Nat. Neurosci., 2: 9-10.

Siagian, C. and L. Itti, 2008. Comparison of gist models in rapid scene categorization tasks. J. Vision, Vol. 8. 10.1167/8.6.734

Zhai, Y. and M. Shah, 2006. Visual attention detection in video sequences using spatiotemporal cues. Proceedings of the 14th Annual ACM International Conference on Multimedia, (MM 06), ACM, New York, USA., pp: 815-824.

Zhang, Y., Z.L. Zhang and Z.K. Shen, 2008. A Salient Feature Extraction Algorithm Fusing the Motion Characteristic of Objects. J. nat. univ. defense technol., 30: 109-115.

Zhang, Y., Z.L. Zhang and Z.K. Shen, 2008. The Images Tracking Algorithm Using Particle Filter Based on Dynamic Salient Features of Targets. acta electron. sinica. 36: 2306-2311.