

Avian Eye-Inspired Visual attention Approach to UAV Target Detection

Beiwei Zhang¹, Jiangtao Cao¹, , Honghai Liu²

1. School of Information and Control Engineering, Liaoning Shihua University, Fushun Liaoning, 113001, China;

2. University of Portsmouth, IEEE Senior Memeber

Abstract

Abstraction – The purpose of this paper is to propose an avian eye-inspired visual attention approach for unmanned air vehicles (UAV), with the objective of improving the capability of UAV target detection. A visual attention model inspired by the functional architecture of the avian visual system is proposed in this paper. The neuro memetic model employs multiple contrast features as primary channels inspired by the contrast sensitivity of avian retina. The global context strategy is also adopted for focus attention based on the visual mechanisms of tectofugal pathway of avian cortex. Our approach is evaluated through quantitative simulations in various environments. Experimental results demonstrated the feasibility and effectiveness of the proposed method: The proposed method is capable to guarantee efficiency of target localization in the complex environment.

Keywords Unmanned air vehicle (UAV), visual attention, target detection, cortex, contrast, context

Introduction

Unmanned aerial vehicles (UAV) have received considerable attentions in multiple applications. UAV has made it possible to perform dangerous and complicated tasks without the participation of man. However, the increasing requirement of UAV for civilian and military applications has led to more complex structure in both software and hardware systems [1-2]. Multiple scenarios have been proposed for UAV route planning, flight control and collision avoidance[3-5] One of the crucial technologies of UAV is the capability of target detection [6-7]. It's necessary to equip UAVs with precise target information for the execution of various tasks. Advancements in digital imaging technology have presented significant results with modified vision sensors integrated with effective image processing software in UAV. The higher flight velocity leads the visual detection on UAVs to be a more challenging task to some extent, considering the velocity effect on image quality [8]. Therefore, the modification of vision algorithms contributes a lot in the processes of UAV crucial target detection undoubtedly.

Multiple works have been done to address the issue of accurate target detection [9]. For instance, template matching is frequently adopted due to its simplicity. However, the adoption of template information cannot avoid the shortage on robustness. The process of matching is prone to be influenced by illumination, rotation and other factors. Visual attention mechanism aims to detect the salient or interested targets in the observed scene [10-12]. It is utilized for extracting useful information in multiple applications such as robot navigation, medical image processing and 3D scene understanding [13-14]. Visual attention has emerged as an efficient strategy to focus processing and analysis in restrained portions of

scenes. Therefore, multiple methods were proposed in succession aimed at improving the attention accuracy [15-17]. In recent decades, huge advancements have been made in the study on functioning principles of visual cortices. The scene-driven bottom-up attention models were established based on the feature integration theory. Itti et al. proposed a classical attention model utilizing multiple low level feature channels and the center-surround difference operator [18]. The combination of feature channels in Itti computational model is on the basis of the early primate visual system [19-21]. Hou X. et al. proposed the spectral residual model employing the phase strategy. Harel J. et al. implemented the Graph-Based Visual Saliency (GBVS) as a modification of the Itti model. All the data-driven methods aim at detecting salient parts automatically. However, the bottom-up mechanism cannot guarantee the extraction of regions of interest. The human visual system generally implements visual research tasks based on prior knowledge of targets. In the task-driven attention models, target features and scene layouts are considered as major resources for visual research. Gist is introduced as a brief presentation of images and frames. The gist feature provides sufficient contextual cues for scene discrimination and target recognition.

There is a considerable amount of visual attention models inspired by mammalian visual systems, especially the human visual system. More research has been concentrated on the study of bird vision as avian visual system possesses higher visual acuity and better capability on detecting some kinds of features. Large quantities of various neuron cells compose a complex network with certain connections in the avian visual system. This paper proposed the saliency detection method based on the contrast sensitivity of avian retina cells. A visual context-based focus selection strategy is utilized simulating the visual cortex network. It is to be noted that the proposed visual attention method does not aspire to replicate all the electrophysiological data of avian visual system. This bio-inspired model aims at implementing the attention paradigms of birds at the functional level. Series of experimental results examined the accuracy and robustness of our proposed model.

This paper is structured as follows. Section 2 reviews the structure and properties of avian visual system. Section 3 presents a formal description of our proposed visual attention model, while in Section 4, comparative experimental results are recorded to evaluate the performances of our proposed method. The final section contains our concluding remarks.

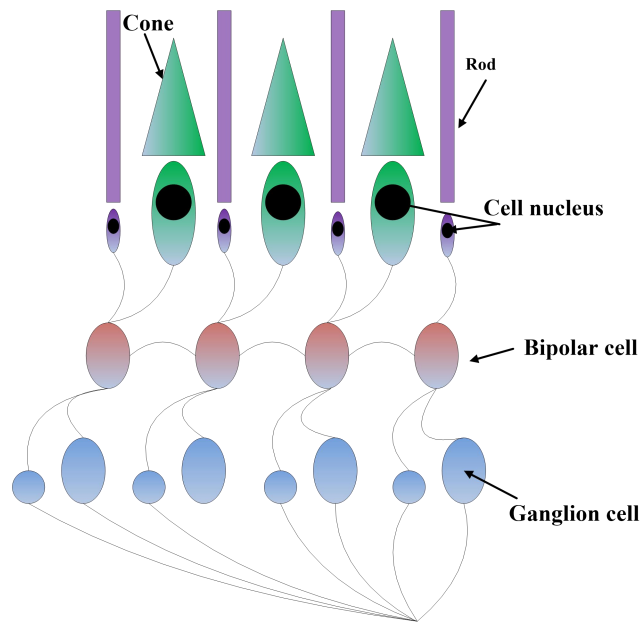
Avian Visual System

Retina

The basic retinal structure of vertebrates generally comprises five layers: photoreceptors, bipolar cells, amacrine cells, horizontal cells, and ganglion cells. Our work mainly concerns about the components and mechanisms of the photoreceptors. Similar to the mammalian retina, avian photoreceptors are classified into rods and cones. The sensitivity of dim illumination is determined by rods, while the cones contribute to the formation of daylight and colour vision. The

retina of birds presents some differences when compared with mammalian retina. For instance, the presence of double cones provides the avian retina with more intraretinal connections. The sensitivity to spectrum ranges from ultraviolet (320nm) to far red (650nm), which is wider than most mammalian retinal cells. Additionally, the existence of oil droplets narrows the bandwidth sensitivity of cones before light reaches the visual pigments.

Figure 1 Schematic drawing of avian retina



Receptive field

The study of visual neuron cell properties is generally based on receptive fields. The functions and characteristics of cells have been obtained by the research on various receptive fields in recent years. The receptive field of simple visual cells generally presents responses to local and primary visual features. It contributes to the detection of specific positions and lines within the visual field. The receptive field of complex cells mainly responds to high level visual features including line segments with different orientations. Both the simple and complex receptive fields are contained in the avian visual network. Therefore, birds are sensitive to low level features as well as special graphs and motion information. Contrast is considered to play an important role during the response of retinal neuron cells. The dependency of contrast in avian visual system is obviously higher than other information such as intensity. The contrast sensitivity and response frequency of visual system differ among various kinds of birds according to physiology searching results.

Visual cortex

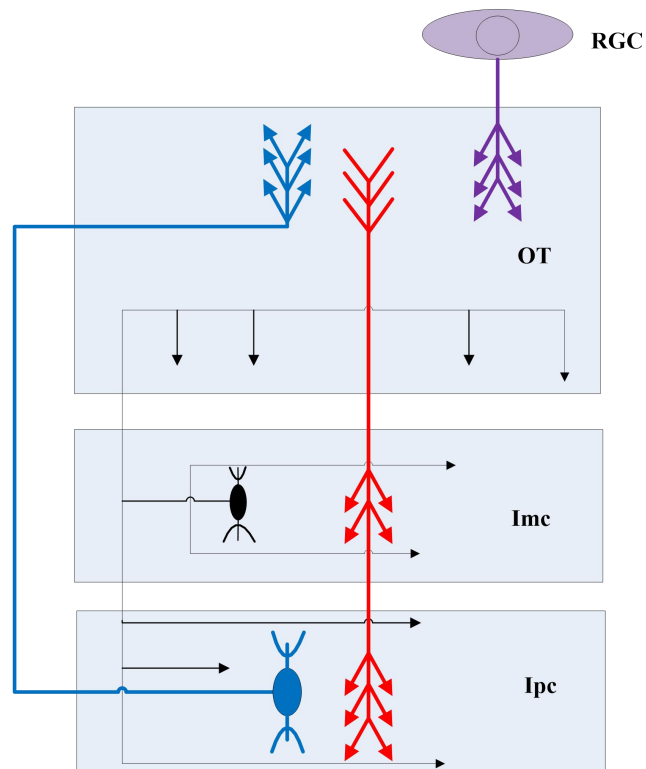
The avian visual system mainly consists of two pathways that transfer and process optical signals acquired from retina: the thalamofugal and the tectofugal pathway. Additionally, the accessory optic system participates in the processing of motion information. In the tectofugal pathway, the optic tectum obtains retinal information and transmits it to the nucleus

rotundus (nRt) on the same side. The ectostriatum acts as the termination of the tectofugal pathway within avian cortex. The physiological features of ectostriatum are similar to optic tectum cells. Most ectostriatum cells are equipped with the capability of selecting orientations. In addition, the ectostriatum is sensitive to the stimulation from small targets.

Tectal-isthmic loop

The tectal-isthmic loop plays an important role in the tectofugal pathway. Generally the nucleus isthmi obtains visual feature information from optic tectum and exports to deep layers of optic tectum, thus generating excitable or inhibitory response. It has been found that the tectal-isthmic loop is especially sensitive to moving targets. The visual field formed in nucleus isthmi is similar to that in the optic tectum. The tectal-isthmic loop is formulated through the interaction between the excitable pathway modulated by isthmi pars magnocellularis (Imc) and the inhibitory excitable pathway modulated by isthmi pars parvocellularis (Ipc). The avian visual system employs the tectal-isthmic loop to implement selective visual attention. Therefore, the winner-take-all network is built up between the optic tectum and nucleus isthmi. The schematic map of tectal-isthmic loop is shown in Fig. 2. Neurons in the optic tectum (OT) receive the input from the retina ganglion cells (RGC). Part of OT neurons project to the nucleus isthmi (red). The neurons in Imc employ Gamma-aminobutyric acid for information transmission and have large visual fields. The cholinergic Ipc neurons (blue) project topographically to the optic tectum.

Figure 2 Schematic drawing of tectal-isthmic loop



Model Development

Primary feature channels

The obtainment, filtering and transmission of optic signal occur during the primary phase of visual processing by avian retinal cells. The resolutions of retinal regions differ due to the characteristics of the retinal structures. Foveas of avian retina generate vision with high resolution. The resolution capability decreases as the cells approach the retinal edge. Contrast sensitivity reflects the ability of avian visual system to distinguish subtle changes in lines and edges. Cells with higher contrast sensitivity are provided with stronger ability to distinguish features. According to the neuroscience research on visual sampling characteristics, the central regions present higher resolution and the surround regions present lower resolution during the imaging of retina. The unique part of scenes generally attracts more attention when processed with neurons in multiple layers. Therefore, the output of visual feature detection can be designed as the contrast of features.

Regions with higher luminance contrast and texture contrast tend to capture more attention of avian visual system compared with surrounding regions. Therefore, the contrast features can be employed as the primary element for saliency detection. The observation windows can be built centered with each pixel. The degree of visual stimulations is determined by the relative intensity rather than absolute intensity according to the avian retina properties. The luminance contrast of pixels can be represented as follows in accordance with the Weber-Fechner rule [22].

$$s = c \lg \frac{I}{I_0} \quad (1)$$

where c is the proportional parameter, and I_0 is the threshold of absolute intensity. Additionally, the texture contrast is calculated by the intensity variance of pixels as

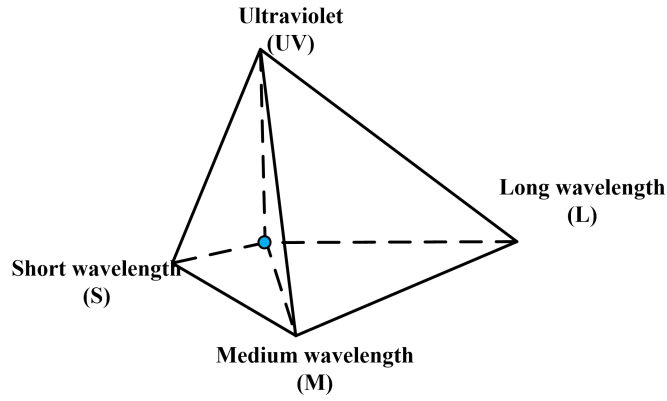
$$(T_{CM}(x, y)) = \left[\frac{1}{N-1} \sum_{n=1}^N (I_n - \frac{1}{N} \sum_{n=1}^N I_n)^2 \right]^{\frac{1}{2}} \quad (2)$$

The colour contrast sensitivity depends on the oil droplet located in the inner segment of retinal cone cells. All the visual signals pass through the oil droplets before reaching the pigments. Generally the oil droplets containing carotenoids are regarded as filters for spectral information. Contrasts are adopted during the coding for the activation of retinal photoreceptors. The signal types of different photoreceptors are supposed to be similar. Three types of retinal cones (L: Long wavelength cone; M: Medium wavelength cone; S: Short wavelength cone.) obtain the incoming light and generate receptor channels accordingly. The transformation from CIE colour space to LMS cone space can be implemented as

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = K \begin{bmatrix} L^* \\ a \\ b \end{bmatrix}, \quad K = \begin{pmatrix} 0.61209 & 2.14231 & 0.12966 \\ 0.38791 & -1.14231 & -0.08217 \\ 0 & 0 & 1.21183 \end{pmatrix} \quad (3)$$

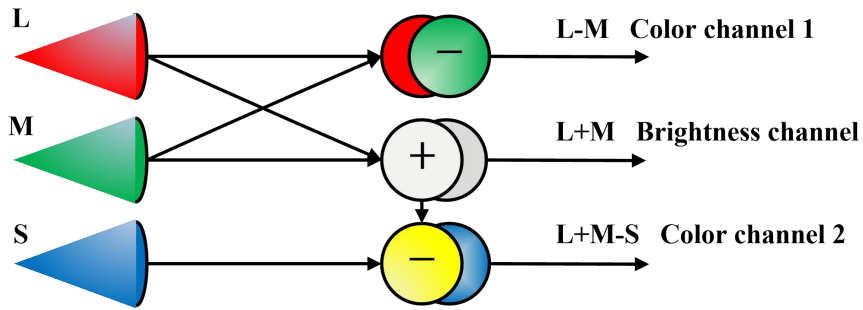
where K is the transformation matrix. The three cones (LMS) compose tetrahedron structure together with ultraviolet spectrum as is illustrated in Fig. 3. This paper merely employs LMS to build the cone colour space.

Figure 3 Tetrahedron structure of avian retina colour space



The cone space is composed of three channels. The brightness channel (L+M) combines the signals of L and M for the processing of intensity information. The other two are (L-M) opponent channel (red-green channel) and (S-(L+M)) opponent channel (yellow-blue) channel respectively. The retinal coupling relationship of colours is illustrated in Fig. 4.

Figure 4 Colour coupling of avian retina



The colour space of avian visual system is equipped with three dimensions corresponding to the colour channels described above. A same-brightness xy plane is formulated by the red-green axis and blue-yellow axis. Each point on the xy plane shares the same intensity and different hues and saturations. Any colour in the avian colour space can be represented using a three-dimensional vector. The hue value is measured with two angles that are similar to longitude and latitude respectively. Brightness is represented using the angle between the azimuth angle and the xy plane ranging from -180° to 180° . The norm of colour vectors represents saturation values ranging from 0 to 1.

The colour difference between two points $Y_1 = (H_1, S_1, I_1)^T$ and $Y_2 = (H_2, S_2, I_2)^T$ is

$$\Delta_{HSI}(Y_1, Y_2) = \sqrt{(\Delta_I)^2 + (\Delta_C)^2} \quad (4)$$

where $\Delta_I = |I_1 - I_2|$, $\Delta_C = \sqrt{S_1^2 + S_2^2 - 2S_1S_2 \cos \theta}$, $\theta = \begin{cases} |H_1 - H_2|; & \text{if } |H_1 - H_2| \leq \pi \\ 2\pi - |H_1 - H_2|; & \text{if } |H_1 - H_2| > \pi \end{cases}$

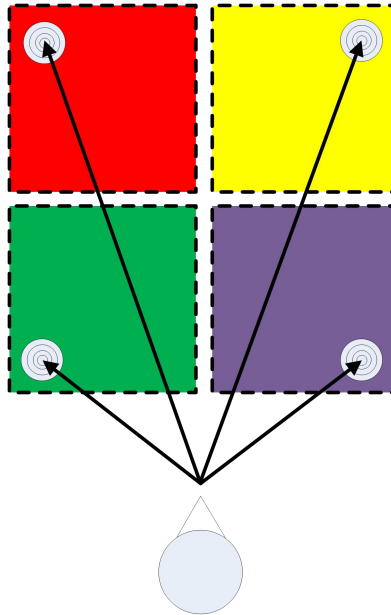
The colour contrast between two points is

$$C_{CM}(x, y) = \frac{1}{N-1} \left[\sum_{n=1}^{N-1} \Delta_{HSI}(Y(x, y), Y_n) \right] \quad (5)$$

Contextual guidance

The targets are generally considered as the basic units for visual searching tasks of birds. However, recently some researchers believe that visual recognition begins from the global scene rather than concentrating on local regions. The ethological researches have demonstrated that birds are capable of locating the targets on the basis of visual context. In the experiment conducted by Wasserman et al.[23], pigeons were placed in a conditioning chamber located in a dark room with white noise. The pigeons were taught to peck a target in four possible positions on the screen in the pre-training phase. The possible positions are corresponded to four different colour contexts or photo contexts as backgrounds as is shown in Fig. 5. The pigeons are capable to estimate the position of target based on the colour or photo context with relatively high accuracy. This result verifies that visual context acts as the guidance of attention and estimates the position of interested targets.

Figure 5 Experiment of visual context of avian eyes



The saliency map is obtained through the integration of avian primary visual features. The primary features are presented as a 8-dimensional vector as

$$F(x, y) = [L(x, y), a(x, y), b(x, y), I_{CM}(x, y), T_{CM}(x, y), C_{CM}(x, y), x, y]^T \quad (6)$$

where $L(x, y)$, $a(x, y)$ and $b(x, y)$ is the three constituents of pixel (x, y) in Lab colour space. After the extraction of retina contrast features, a winner-take-all visual context model is constructed inspired by the optic tectum-nucleus isthmi network in avian cortex. The context model consists of the saliency map generation and focus transition. The saliency value of each point can be estimated by calculating the covariance between the feature vectors of the current point and its neighbourhood [24]. The covariance matrices similar regions in the image resemble each other, while the covariance of dissimilar parts is distinct. Region R in the feature map can be described using the covariance matrix as

$$C_R = \frac{1}{n-1} \sum_{i=1}^n (f_i - \mu)(f_i - \mu)^T \quad (7)$$

where $\{f_i\}_{i=1, \dots, n}$ presents the feature value of i -th feature point in region R , and μ is the mean value of all the feature points. The covariance comparison between two regions can be implemented using the Euclidean distance.

$$\rho(C_1, C_2) = \sqrt{\sum_{i=1}^n \ln^2 \lambda_i(C_1, C_2)} \quad (8)$$

where $\lambda_i(C_1, C_2)$ is the normalized feature value of the i -th point within region R . The diagonal elements in covariance matrices contain integrated information of different visual features, while the other elements describe the correlation between the features. The saliency value can be achieved based on the comparison of feature covariance introduced above. The dissimilarity of region i and j is

$$d(R_i, R_j) = \frac{\rho(C_i, C_j)}{1 + \|x_i - x_j\|} \quad (9)$$

where x_i and x_j are the feature matrices of region R_i and R_j respectively. m regions that are the most similar with R_i are selected to calculate the saliency as

$$S(R_i) = \frac{1}{m} \sum_{j=1}^m d(R_i, R_j) \quad (10)$$

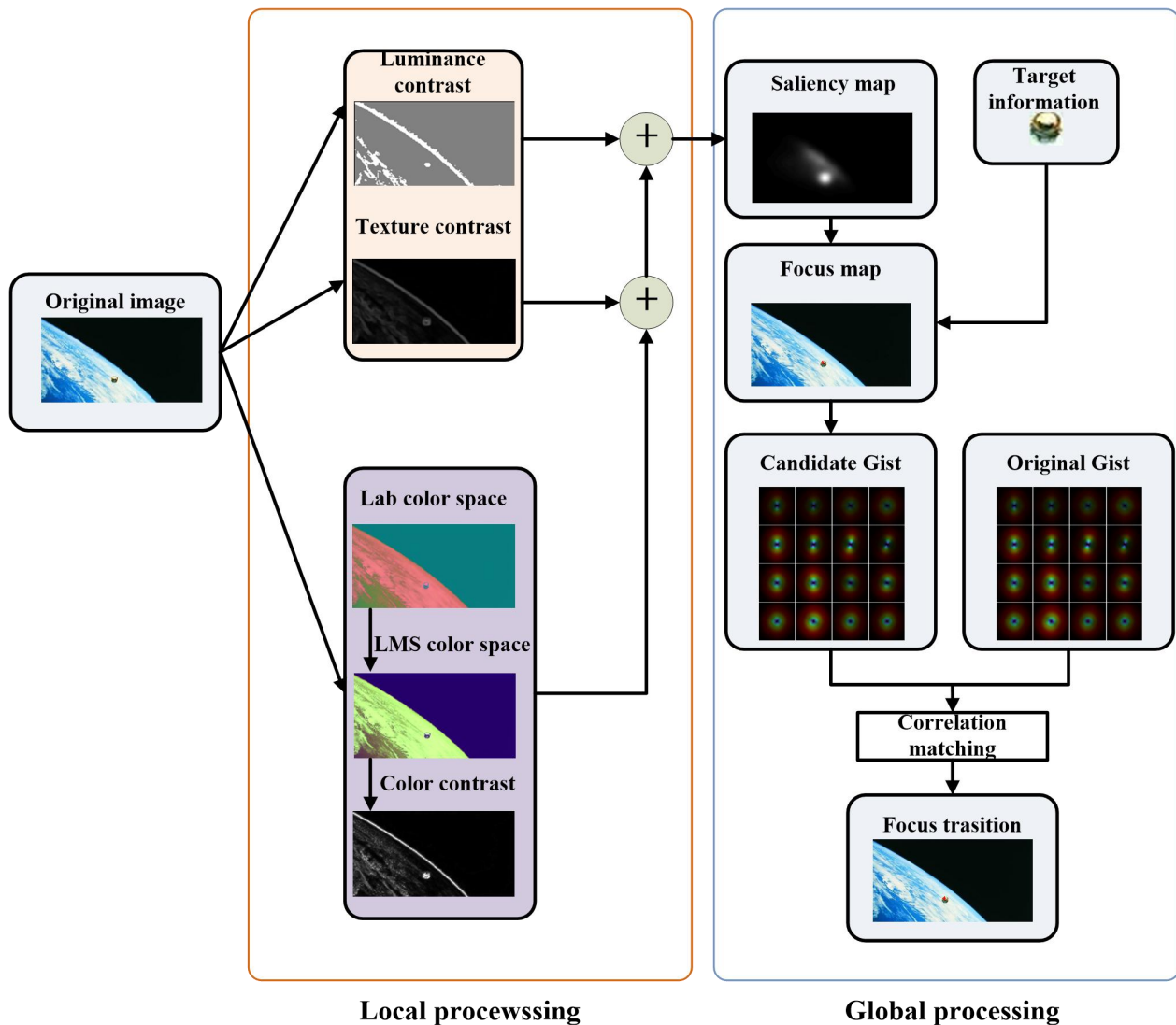
The final saliency map is obtained by covariance operation on each image region. The simple weight assignment for feature integration in Itti model cannot meet the requirement in various environments. The adopted feature covariance strategy fit the pattern of non-linear information integration of complex cells in avian cortex better.

The transition of focus from saliency regions to interest regions can be implemented by the cooperation and competition among information in avian visual cortex. After obtaining the initial saliency map, the focus of attention eagle

guided by visual information between competition and cooperation, to achieve the focus of attention for the transfer (the transfer of interest from a significant area to area) .

The Gist feature proposed by Torralba is adopted as global context for focus transition. The semantic information in scenes can be described using Gist feature. The scene is described using a group of Gist features rather than a single global feature. The Gist feature extraction utilizes a group of Gabor filters with multiple scales and directions for image resolution. The filtered results are projected to the grid structure for mean response calculation. The Gist feature can be obtained by the dimension reduction of mean response vectors. The proposed visual context matching strategy exchanges the window content with template. The Gist feature is adopted to measure the similarity of the original image and the exchanged image. This implicates that the similarity of sub-image and template is measured in the scene context. The flow chart of our proposed visual attention mechanism is presented in Fig. 6.

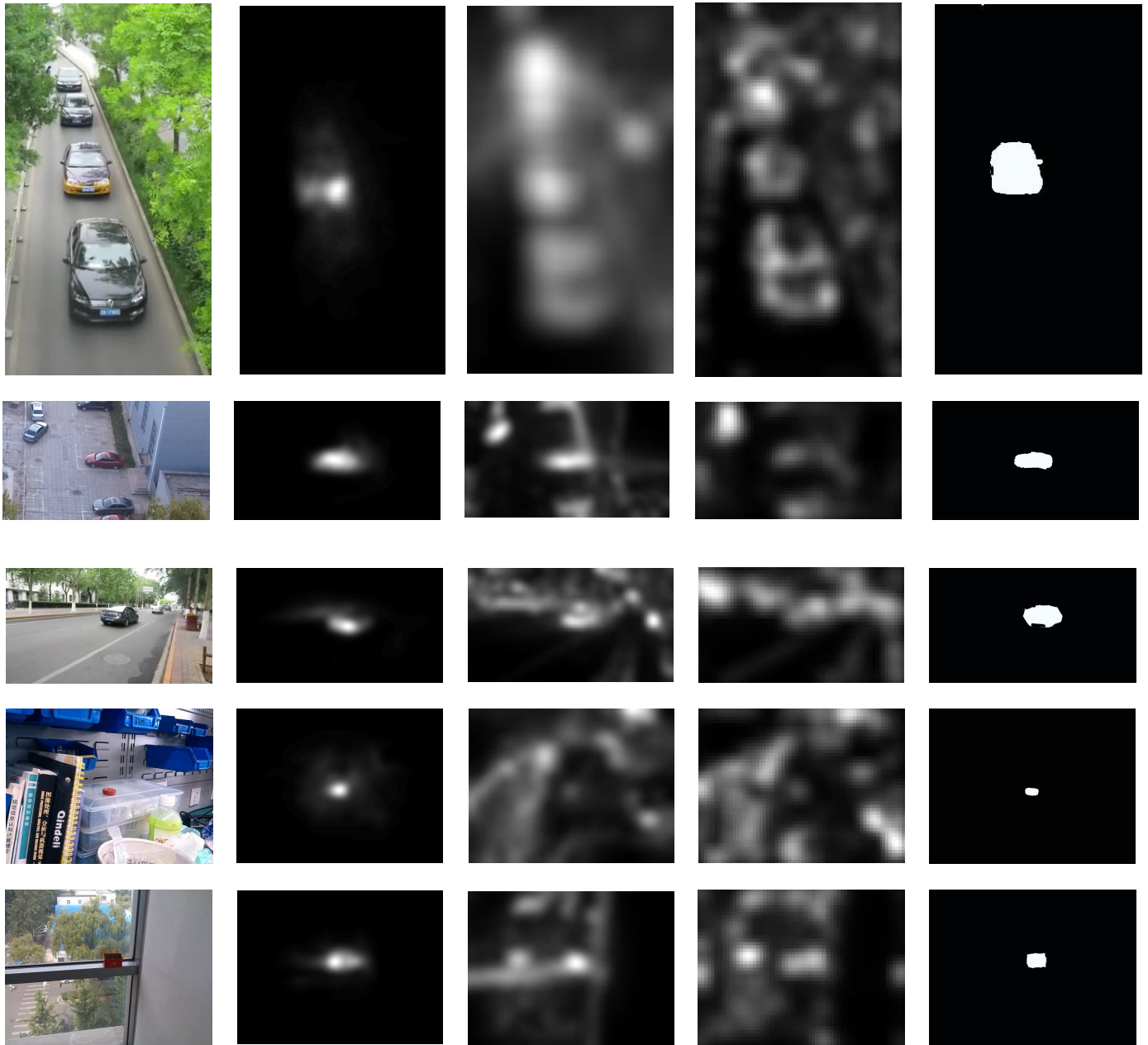
Figure 6 Flow chart of the proposed visual attention mechanism



Results

Series of experiments have been conducted to verify the feasibility and accuracy of the proposed approach in this work. The performances of saliency detection and FOA shift were validated respectively. The experimental results compared with two classical attention models for saliency detection are recorded in Fig. 7.

Figure 7 Original image and results of edge detection algorithms



(a)

(b)

(c)

(d)

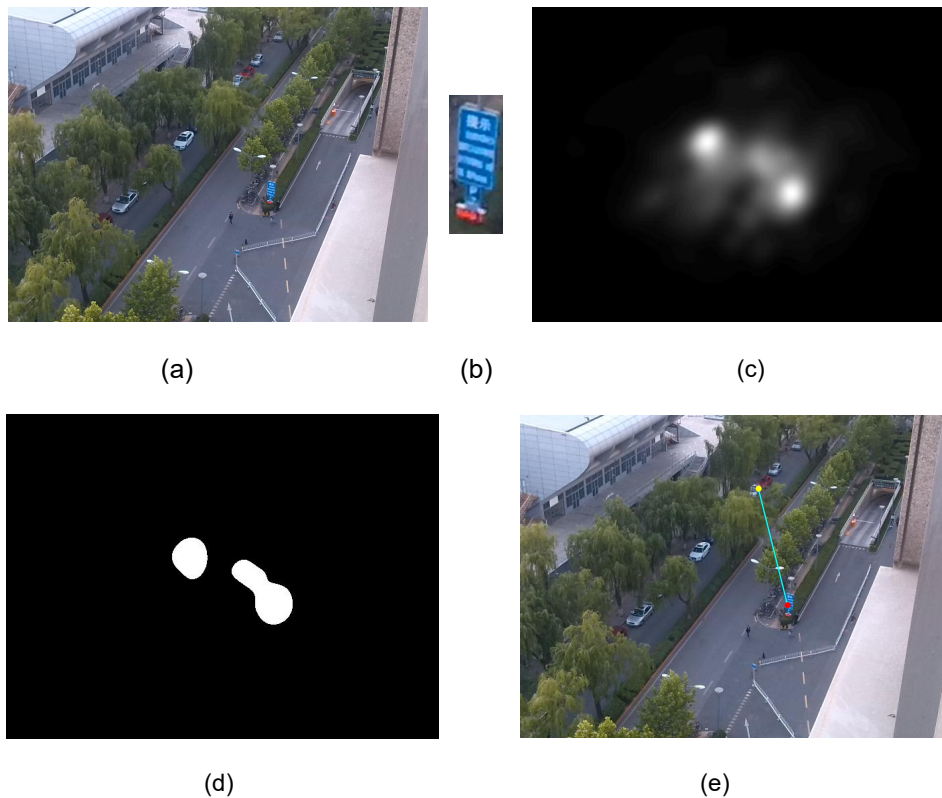
(e)

Notes: (a) Input. (b), (c) and (d) are saliency map results of visual attention using our method, Itti model and SR model respectively. (e) Ground truth

We selected three outdoor scenes and two indoor scenes for the target detection tests. Targets and backgrounds were labelled 1 and 0 respectively in the ground truths. For each image, the ground truth presents regions where typical viewers focus in the scene. The saliency results in Itti and SR model cannot remove the background disturbance effectively. Our method adopted the visual context strategy using feature covariance comparison. Therefore, the proposed avian eye-based visual attention model makes better performances on removing background disturbance.

Tests on focus of attention (FOA) transmission were implemented based on the saliency detection conducted above. The threshold of region segmentation is 0.1. The size of gist feature for each image is 512×1 . Results of FOA shift tests are shown in Fig. 8.

Figure 8 Evolution curves of our algorithm.



Notes: FOA shift tests (a) Original image. (b) Template. (c) Saliency map. (d) Saliency segmentation.

(e) FOA shift

The experimental results shown in Fig.8 (e) simulated the avian eye movement from salient regions to the real target based on the contextual guidance. The blue billboard illustrated in Fig. 8(b) is supposed to be the real target. However, the white car is considered as the most salient part in the scene due to its uniqueness compared with the neighbourhood. Combination images are built by embedding the template into the salient regions in turn. The FOA shift is implemented according to the correlation between gist features of the original image and the combination image. The results have demonstrated the effectiveness of our method on UAV target detection.

Conclusion

In this paper, the avian eye-inspired visual attention model is established for target detection of UAV. The contrast sensitivity of avian retina is simulated for the low-level feature detection. Saliency composition and focus transition are implemented on the basis of ethological researching results.

The visual attention model can help find the target region fast and accurately, thus reducing the amount of information and make the following processes easier and simpler. The experimental results show that our proposed method is a feasible and effective way in target edge detection. Our future work will focus on how to apply the proposed approach to solve other real-world problems, such as, mobile robots, industry production line, and intelligent transportation.

Acknowledgements

This work was supported by Natural Science Foundation of China(NSFC) under grant #61203021.

References

- [1] Pradeep, S. (2002), "Nonlinear control of unmanned combat aircraft during take-off", AIAA paper 2002-0250, in AIAA 40th Aerospace Sciences Meeting and Exhibit, Reno, NY, 2002.
- [2] Zhang, Y. P., Duan, H. B. (2013), ".A directional control system forUCAV automatic takeoff roll," Aircraft Engineering and Aerospace Technology, Vol. 85, No. 1, pp.46-61.
- [3] Zhang, X. Y., Duan, H. B., Yu, Y. X. (2010) ,"Receding horizon control for multi-UAVs close formation control based on differential evolution", SCIENCE CHINA Information Sciences(Science in China Series F), Vol.53, No.2, pp. 223-235.
- [4] Xu, C. F., Duan, H. B. (2010), "Artificial bee colony (ABC) optimized edge potential function (EPF) approach to target recognition for low-altitude aircraft", Pattern Recognition Letters, Vol. 31, No. 13, pp. 1759–1772.
- [5] Zhao, Z. Y., Lu, G. S. (2012),"Receding horizon control for cooperative search of multi-UAVs based on differential evolution", International Journal of Intelligent Computing and Cybernetics, Vol. 5, No.1, pp. 145-158.
- [6] Bertuccelli, L. F., Cummings, M. L. (2010), "Operator choice modeling for collaborative UAV visual search tasks", IEEE Transactions on System, Man, and Cybernetics—Part A: Systems and Humans, Vol. 42, No. 5, pp. 1088–1099.
- [7] Deng, Y. M., Duan, H. B. (2014), "Biological edge detection forUCAV via improved artificial bee colony and visual attention," Aircraft Engineering and Aerospace Technology, Vol. 86, No. 2, pp.138–146.
- [8] Lin, F., Dong, X. X., Chen, B. M., Lum, K. Y., Lee, T. H. (2012), "A robust real-time embedded vision system on an unmanned rotorcraft for ground target following", IEEE Transactions on Industrial Electronics, Vol. 59, No. 2, pp. 1038–1049.
- [9] Tian, J., Yu, W. Y., Xie. S. L. (2008), "An ant colony optimization algorithm for image edge detection", IEEE Congress on Evolutionary Computation (CEC), Hong Kong, China, pp. 751-756
- [10] Jiwhan Kim, Dongyoon Han, Yu-Wing Tai, Junmo Kim, Salient region detection via high-dimensional color transform, in: CVPR, 2014.
- [11] X. Li , Y. Li, C. Shen, A. Dick, A. van den Hengel, Contextual hypergraph modeling for salient object detection, in: ICCV, 2013.
- [12] H.-H. Yeh, et al., Salient object detection via local saliency estimation and global homogeneity refinement, Pattern Recognit. 47 (2014) 1740–1750.
- [13] Li, Z. C., Qin. S. Y., Itti, L. (2011), "Visual attention guided bit allocation in video compression", Image and Vision Computing, Vol. 29, No. 1, pp. 1-14.
- [14] Siagian, C., Itti, L. (2007), "Rapid biologically-inspired scene classification using features shared with visual attention", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No. 2, pp. 300-312.
- [15] Cong Jia, Jinqing Qi, Xiaohui Li, Huchuan Lu. Saliency detection via a unified generative and discriminative model. Neurocomputing, Volume 173, Part 2, 15 January 2016, Pages 406-417.
- [16] Se-Ho Lee, Je-Won Kang, Chang-Su Kim. Compressed domain video saliency detection using global and local spatiotemporal features. Journal of Visual Communication and Image Representation, Volume 35, February 2016, Pages 169-183.

- [17] Zhe Chen, Xin Wang, Zhen Sun, Zhijian Wang. Motion saliency detection using a temporal fourier transform. *Optics & Laser Technology*, Volume 80, June 2016, Pages 1-15.
- [18] Itti, L., Koch, C., Niebur, E. (1998), "A model of saliency-based visual attention for rapid scene analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 11, pp. 1254-1259.
- [19] Itti, L., Gold, C., Koch, C. (2001), "Visual attention and target detection in cluttered natural scenes", *Optical Engineering*, Vol. 40, No. 9, pp. 1784-1793.
- [20] Parikh, N., Itti, L., Weiland, J. (2010), "Saliency-based image processing for retinal prostheses", *Journal of Neural Engineering*, Vol. 7, No. 1, pp. 1741-2560.
- [21] Itti, L., Koch, C. (2001), Computational modelling of visual attention. *Nature reviews neuroscience*, Vol. 2, No. 3, pp. 194-203.
- [22] Zhang, J., Sun, J., Liu, J., Yang, C., Yan, H. (2010). "Visual attention model based on multi-scale local contrast of low-level features", 2010 IEEE 10th International Conference on In Signal Processing (ICSP), IEEE, pp. 902-905.
- [23] Wasserman, E. A., Teng, Y. J., Castro L. (2014), "Pigeons exhibit contextual cueing to both simple and complex backgrounds", *Behavioural Processes*, Vol. 104, pp. 44-52.
- [24] Erdem E., Erdem A. (2013) "Visual saliency estimation by nonlinearly integrating features using region covariances", *Journal of vision*, Vol. 13, No. 4, pp. 1-20.