Machine Intelligence Research
www.mi-research.net

20(5), October 2023, 741-752
DOI: 10.1007/s11633-022-1342-3

Biological Eagle-eye Inspired Target Detection for Unmanned Aerial Vehicles Equipped with a Manipulator

Yi-Min Deng Si-Yuan Wang

School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China

Abstract: Inspired by eagle eye mechanisms, the structure and information processing characteristics of the eagle′s visual system are used for the target capture task of an unmanned aerial vehicle (UAV) with a mechanical arm. In this paper, a novel eagle-eye inspired multi-camera sensor and a saliency detection method are proposed. A combined camera system is built by simulating the double fovea structure on the eagle retina. A saliency target detection method based on the eagle midbrain inhibition mechanism is proposed by measuring the static saliency information and dynamic features. Thus, salient targets can be accurately detected through the collaborative work between different cameras of the proposed multi-camera sensor. Experimental results show that the eagle-eye inspired visual system is able to continuously detect targets in outdoor scenes and that the proposed algorithm has a strong inhibitory effect on moving background interference.

Keywords: Unmanned aerial vehicle (UAV), eagle eye, multi-camera sensor, target detection, saliency detection.

Citation: Y. M. Deng, S. Y. Wang. Biological eagle-eye inspired target detection for unmanned aerial vehicles equipped with a manipulator. *Machine Intelligence Research*, vol.20, no.5, pp.741–752, 2023. http://doi.org/10.1007/s11633-022-1342-3

1 Introduction

With in-depth research on unmanned aerial vehicle (UAV) technology, UAVs with robotic arms have gradually become a research hotspot[1]. The autonomy of an unmanned aerial vehicle with a manipulator depends on intelligent perception. Visual sensors are one of the mainstream perception devices. Designing image acquisition devices and image processing algorithms is a challenge. At present, many studies have been carried out on the target recognition technology of fixed platform manipulators[2, 3]. However, for the manipulator grasping on the UAV, the requirement of visual target detection is higher. Thomas et al.[4] implemented an indoor image-based autonomous grasping of micro aerial vehicles with a manipulator. However, the indoor background interference is low, and the lighting conditions are suitable. The detection and relative position measurement of outdoor targets is a challenge.

In all animals, the eagle′s visual system is second to none and has advantages such as high visual acuity[5]. The eagle has strong color perception and broad visual fields and is sensitive to movement information[6]. The process of eagle hunting prey is similar to the process of the UAV platform with a robotic arm grasping target. Therefore, the visual target detection process in the UAV grasping task can learn from the biological mechanism of the eagle. Various eagle-inspired studies on composite image sensors have made some progress. Deng and Duan[7] designed a biological eagle-eye-based visual platform simulating the eagle's vision system for target detection. However, the load capacity and available space of UAVs are not sufficient for the device. Fu et al.[8] presented a lightweight vision system based on an eagle-eye structure. However, the system needs an additional servo control module and image transmission module, and the image processing process needs to be transferred to the ground station. Therefore, a simple and efficient design is required for the vision devices used on the UAV platform while simulating the eagle-eye's structural characteristics.

After image acquisition, detecting targets in unfamiliar environments has long been a challenge in computer vision. In [9], the image was first decomposed into large-scale perceptually homogeneous regions. Then, the target saliency map was computed using feature similarity and spatial difference based on global contrast cues. However, these methods are affected by the background and cannot filter out the disturbance of surface vegetation. Bonnin-Pascual and Ortiz[10] used intensity, color, and orientation features to address defect detection problems. The Camshift algorithm[11] was applied for a moving target to detect and track a motion target in a continuous image sequence. In an application, the UAV needs to determine the position of the target to be captured during flight.

Research Article
Manuscript received on January 17, 2022; accepted on May 30, 2022; published online on March 7, 2023
Recommended by Associate Editor Lu-Ping Shi
Colored figures are available in the online version at https://link.springer.com/journal/11633

During the movement of UAVs, the image background changes constantly, which makes saliency detection difficult. Saliency detection mechanisms exist in biological visual systems, which can achieve the preferential allocation of computational resources[12]. The saliency detection based on the deep learning method has a good effect, but it is challenging to apply it on a microprocessor at a UAV[13−15].

Inspired by eagle's visual system, Duan et al.[16] applied contrast sensitivity, eagle-eye vision adaptation and other mechanisms to many applications, such as small target detection and visual measurement. In [17], by imitating the structure of the eagle eye and the biological mechanism of visual information processing of the eagle's visual system, the problem of salient target detection in the UAV platform grabbing task was solved. Li et al.[18] also presented an algorithm based on properties in the eagle eye to solve the problem of autonomous landing of UAVs on water.

In this paper, we present a biological eagle-eye inspired hardware and software integration system to solve the target detection problem. The main contributions of the paper are as follows:

1) A novel static and dynamic feature-based saliency detection algorithm is designed. The saliency cues are calculated based on the color attention mechanism and the midbrain inhibitory circuit structure of the eagle.

2) An eagle-eye inspired multi-camera sensor is presented to simulate the double fovea structure on the retina of the eagle eye, which is used for image acquisition.

3) According to the specialty of the multi-camera sensor, a cooperative positioning method between two cameras is designed by switching image sources.

The remaining section of the paper is organized as follows. The saliency detection method inspired by the eagle midbrain inhibition mechanism is introduced in Section 2. The hardware and software design of the eagle-eye inspired multi-camera sensor is presented in Section 3. Section 4 shows the experimental results of our saliency detection method and UAV navigation. The conclusion and perspectives are shown in Section 5.

2 Target detection algorithm

Our moving object detection method mainly contains two parts. The first part is cluster-based saliency detection[19], which represents the static feature. The second part is eagle-eye inspired dynamic perception, which denotes motion characteristics. The flowchart of our algorithm is shown in Fig. 1. Our method can consider both the static and dynamic characteristics of moving objects.

2.1 Static image saliency

The single image saliency calculation is inspired by the cluster-based method[19]. In general, the generic dataset images and camera-captured images are usually stored in RGB color space. To make the calculation process closer to the physiological characteristics of raptors, the

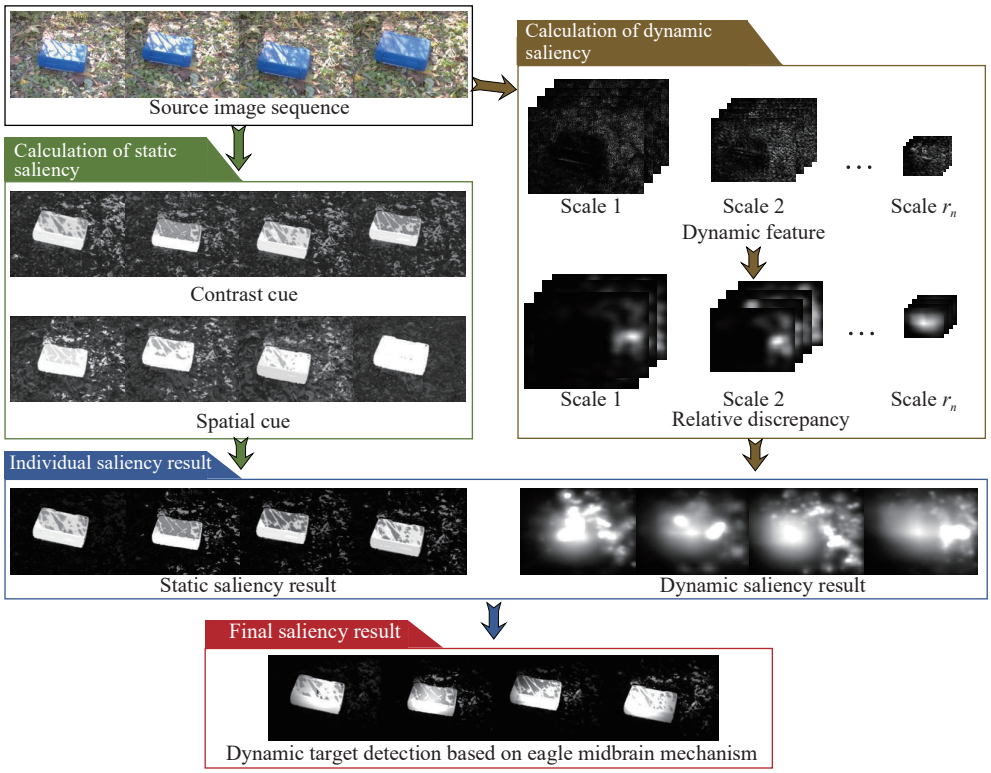


Fig. 1 Framework of our dynamic target detection algorithm

RGB color space was first converted to the LMS color space. The conversion formula from the RGB color space to the LMS color space is as follows:

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.381\,1 & 0.578\,3 & 0.040\,2 \\ 0.196\,7 & 0.724\,4 & 0.078\,2 \\ 0.024\,1 & 0.128\,8 & 0.844\,4 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (1)$$

Raptors without prior training tend to gaze at certain salient areas of the scene[20]. Generally, salient areas have specific image features, such as luminance, color, and edge density. Therefore, we use cluster-based saliency cues in [19] to simulate the biological mechanisms of a raptor's visual system.

First, the contrast cue represents the color saliency of the target. This cue measures the color difference between the current pixel block and the surrounding pixel block. The formula of contrast cue $\lambda^c(k)$ was calculated as follows:

$$\lambda^c(k) = \sum_{i=1, i\neq k}^{N^k} \frac{n^i}{N} \left\| \mu^k - \mu^i \right\|_2 \quad (2)$$

where k denotes cluster C^k, N^k represents the total cluster number, the L_2 norm quantifies the color and luminance features, n^i denotes the pixel number of cluster C^i, N denotes the pixel number of the whole image, and μ denotes the average of the three color channels in the cluster.

Second, we use the spatial cue to simulate the raptor's gaze effect. The visual saliency in the center of the eagle field is higher than that around the field of view (FOV). The following formula can characterize the spatial saliency:

$$\frac{1}{\|x_p - o\|^2} \quad (3)$$

where x_p is the coordinate of pixel p in the image. o represents the center position of the entire image. $\|\cdot\|^2$ is the Euclidean distance calculation symbol.

Raptors tend to bring objects of interest to a retinal fixation area near the central vision field[21]. Each cluster has similar low-level image feature information, and the pixels are spatially continuously distributed. The spatial cue λ^s of cluster C^k is described as follows:

$$\lambda^s(k) = \frac{1}{n^k} \sum_{i=1}^{N^k} \frac{1}{\|x_i - o\|^2} \quad (4)$$

where n^k is the total number of pixels in image block C^k, considering the adjacent frame relationship in the image sequence, targets that repeat in the central area of an image will have higher spatial saliency. Thus, (4) is rewritten as

$$\lambda^s(k) = \frac{1}{n^k} \sum_{j=1}^{M} \sum_{i=1}^{N^k} \left[\frac{1}{\left\| x_i^j - o^j \right\|^2} \cdot \delta\left(c(x_i^j) - C^k \right) \right] \quad (5)$$

where M is the total number of images in the image sequence, j represents the number of images, and $\delta(\cdot)$ denotes the Kronecker delta function. $c(\cdot)$ is a mapping relationship from pixel to cluster, and $c(x_i^j)$ represents the cluster where pixel x_i^j is located. Since $\left\| x_i^j - o^j \right\|$ is calculated as the pixel distance in (5), there is a significant variation and discontinuity in the data. Therefore, normalization is performed after calculating the spatial saliency of all clusters.

The static saliency $S_s(j)$ of image j in the image sequence is derived from the combination of spatial and color cues. Pixels in cluster C^k have the same static saliency value. The static saliency $S_s(j)$ can be calculated as follows:

$$S_s(j) = \lambda^s(k) \cdot \lambda^c(k). \quad (6)$$

2.2 Midbrain network inhibition mechanism-based dynamic perception

For the target detection task of UAVs while moving, target location extraction in the field vegetation cover environment cannot be satisfied by the low-level image features alone. In addition to low-level image features such as color and morphology, the target also has a relative movement of the UAV. The precondition to obtain the dynamic characteristics of the target is to eliminate background information interference. The eagle midbrain plays an essential role in eagle visual information processing. According to the midbrain stimulus selection network[22, 23], a dynamic feature extraction framework based on midbrain inhibition was extracted. The framework is shown in Fig. 2.

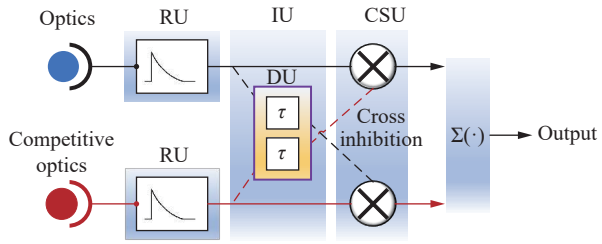


Fig. 2 Dynamic feature extraction framework based on midbrain inhibition

The dynamic perception extraction framework is established according to the intercellular inhibition and the structure of the cross inhibition between the optic tectum and the nucleus isthmus large cell in the midbrain. The dynamic perception extraction framework consists of six parts: input optics, response unit (RU), delay unit (DU),

inhibition unit (IU), competitive selection unit (CSU) and output unit. Each part corresponds to Fig. 2.

First, using a one-dimensional vector mask $s = [-1, 0, 1]$, the gradient is calculated in the horizontal and vertical directions of the image. We define g_i^x as the gradient in the x direction and g_i^y as the gradient in the y direction of pixel x_i. Then, the gradient response G_i in pixel x_i can be obtained by

$$G_i = \sqrt{(g_i^x)^2 + (g_i^y)^2}. \tag{7}$$

Gradient response G_i is then input to the inhibitory section of the midbrain network. The inhibitory effect is achieved mainly through the interaction of delay signals with nondelay signals. In this process, the interference information will be inhibited. We suppose that the delay parameter in inhibition is τ and the output of DU is $G_i^d(t)$ at time t. Then, the output of IU with DU can be expressed as

$$G_i^\tau(t) = \frac{G_i(t) - G_i^d(t)}{\tau}. \tag{8}$$

Similarly, the delay output $G_i^d(t+1)$ for moment $t+1$ is

$$G_i^d(t+1) = G_i^d(t) + G_i^\tau(t) \tag{9}$$

where the initial state is $G_i^d(0) = 0$.

The output of DU and IU in the opposite competitive branch can be expressed as ${G_i}'(t)$ and $G_i^{d'}(t)$, respectively. Then, the output $R_i(t)$ of the CSU is calculated as

$$R_i(t) = G_i(t)G_i^{d'}(t) - {G_i}'(t)G_i^d(t). \tag{10}$$

The connection of the eagle visual system is an extended two-dimensional structure. After synthesizing the x and y directions, the dynamic feature extraction framework of the midbrain inhibition output $O_i(t)$ is

$$O_i(t) = \sqrt{(R_i^x)^2 + (R_i^y)^2}. \tag{11}$$

After processing the dynamic feature extraction framework of midbrain inhibition, there is still some background interference in the results. The distribution difference between the target object response and background interference response can be used to filter out interference further. According to [24], textural contrast is defined as

$$D(i) = \begin{bmatrix} \sum_{l\in P_i} O_x^2(l) & \sum_{l\in P_i} O_x(l)O_y(l) \\ \sum_{l\in P_i} O_x(l)O_y(l) & \sum_{l\in P_i} O_y^2(l) \end{bmatrix} \tag{12}$$

where $O_x(l)$ and $O_y(l)$ denote the gradient in the horizontal and vertical directions of the dynamic feature, respectively. l is the pixel in the neighborhood of pixel i.

Generally, the background area has a circular feature distribution, while the target area has an elliptical feature distribution. Therefore, the difference in distribution can be obtained by calculating the eigenvalues of $D(i)$. The relative discrepancy $S_M(i)$ between two eigenvalues of $D(i)$ is

$$S_M(i) = (\lambda_1 - \lambda_2)^2. \tag{13}$$

The relative distance between the UAV and target varies from far to near during UAV flight. Meanwhile, the total number of pixels the target occupies in the image changes from less to more. The dynamic feature extraction based on midbrain inhibitory synthesis of multiple scales adapts to substantial scale changes. The dynamic feature of image j in scale s_r is denoted by $S_M(j, r)$. The multiscale synthesis result $S_M(j)$ is

$$S_M(j) = \frac{1}{n} \sum_{r\in\{r_1, r_2, \ldots, r_n\}} S_M(j, r) \tag{14}$$

where n is the total number of scales. The final result for dynamic saliency detection based on the eagle midbrain inhibition mechanism is

$$S(j) = S_s(j) \times S_M(j). \tag{15}$$

3 Eagle-eye inspired multi-camera sensor

The search task for ground targets has high requirements for visual sensors. Because the target position is unknown, it is difficult to ensure search efficiency and accuracy when a UAV searches according to the established route by relying on a single fixed focus camera. Ordinary RGB cameras have limitations. Although the short-focus camera has a large FOV, it has a poor resolution ratio for long-distance targets. The long-focus camera can collect the characteristics of long-distance target objects, but the FOV is small, and the target is easy to move out of the FOV. Among living animals, eagles have the highest chromatic visual acuity threshold[25]. The eagle eye can consider a wide field of vision and high spatial resolution. According to the physiological structure of the eagle eye, a multi-camera sensor imitating the eagle eye double fovea is designed to complete the pose measurement of the target.

3.1 Design scheme of the eagle-eye inspired multi-camera sensor

To combine the characteristics of a large field of vision with high resolution, we designed a multi-camera sensor based on a eagle-eye biological structure. The ret-

inal structure of the eagle eye is different from that of other animals. This unique visual structure is helpful for eagle flight positioning and target capture. Such advantages are consistent with the search task requirements of the UAV platform. The relative distance between the target and the UAV keeps changing during searching. In the process of the UAV approaching the target, the wide FOV is first used to search the target and detect the position of the target in the image. Then, the target feature information is extracted by using the high-resolution region. The eagle eye retina has two high-resolution imaging core areas, the deep fovea and the shallow fovea[26]. Two foveae can be used for different functional imaging and image processing. Inspired by the biological structure of the eagle eye retina, we designed and constructed a multi-camera sensor.

The structure of the eagle eye retina brings ideas to hardware design. The deep fovea on the eagle eye retina has maximum visual acuity. To imitate this structure, we use a long-focus camera to capture the details of the target. The shallow fovea on the retina has less visual acuity than the deep fovea. However, the shallow fovea provides a wider FOV. We use a short-focus camera to imitate the shallow fovea. Thus, considering the combination of long-focus and short-focus dual cameras, a multi-camera sensor with two eagle eye-like fovea structures is constructed. The installation components of the multi-camera sensor are 3D printed with nylon. The dimensions of the device are carefully designed to ensure structural compactness. We use a Basler daA1280-54uc camera as a visual sensor, which is small in size and mass and has high imaging quality.

According to the pinhole camera model[27], focal length parameters are designed. In the application scenario, the upper limit of the distance between the camera and target is 10 m. The lower limit is 15–20 cm according to the grasping situation. According to the pinhole camera model and camera parameters, the focal length of the short-focus camera is 2.8 mm and that of the long-focus camera is 12 mm. This set of parameters ensures that the target can occupy over 100 pixels on a long focal length camera image at a distance. In close grasp, the short-focus camera FOV can fully cover the target. Two cameras collaborate through task logic design. The multi-camera sensor can search the target in a wide-angle view and further judge the details of the target. It can continuously monitor and detect a wide range of visual fields and analyse and measure targets with high resolution. The eagle-eye inspired multi-camera sensor is shown in Fig. 3.

3.2 Processing logic design of the eagle-eye inspired multi-camera sensor

The target detection and visual measurement task process are shown in Fig. 4. The whole task is mainly divided into two stages. UAV cruising along preset routes due to a lack of prior information about target locations. During the cruise, the multi-camera sensor device with an angle of 30° between the optical axis and the horizontal plane continuously collects images. Due to the slight field-of-view angle of the long-focus camera, it is difficult to determine whether the target appears in the view at this stage. At this time, detecting the image of the long-focus camera can cause unnecessary waste of computation power. To solve this problem, the image of the short-focus camera is first processed. Dynamic target detection based on midbrain inhibition (DTDMI) is used to calculate the target area. The target saliency detection results are shown in Fig. 4(g). The target is marked with a green rectangular box. Figs. 4(e) and 4(f) show the images collected by the eagle-eye inspired multi-camera sensor. These two images are collected simultaneously by the short-focus camera and the long-focus camera. The two images are aligned by speeded up robust features (SURF)[28]. The corresponding area of the long-focus camera's FOV in the short-focus camera's FOV is represented by a red box. In the image captured by the long focal camera, the DTDMI processing results are combined with the SURF matching results. It can be known that the target is not in the FOV of the long-focus camera at the current moment.

First, the camera coordinate system is defined in the same way as in [29]. Due to the small number of pixels in the long-focus camera image, it is difficult to accurately estimate the relative distance between the UAV and the target. Therefore, the UAV navigation information is

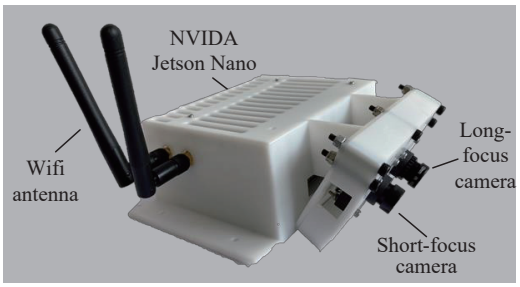


Fig. 3 Eagle-eye inspired multi-camera sensor

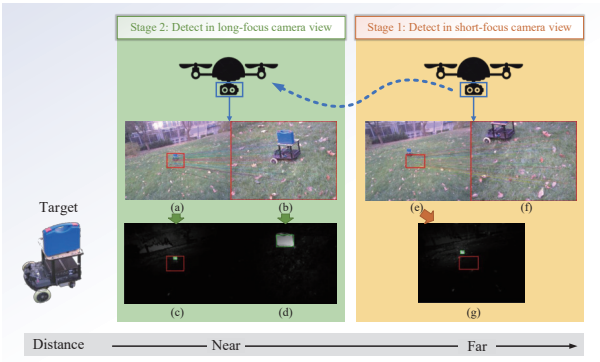


Fig. 4 Capture task flow diagram

roughly estimated using the following methods:

$$E_i = \Delta p_i \beta_i \frac{FOV}{M_i} \tag{16}$$

where $i \in \{x, y\}$, E_i is the estimated estimative distances of the target in the x and y directions, Δp_i denotes the difference between the pixel value of the target center and the center of the image in direction i, β_i represents the pixel size in direction i, and M_i indicates the resolution in direction i. The offset of the target in the camera coordinate system from directions x and y is calculated. Based on the calculated relative position, the UAV can correct the preset navigation points and keep approaching the target position.

In the process of UAV visual navigation, two-camera image matching and saliency detection of short-focus camera images are continuously performed. When the target is detected to be within the image matching area (as shown in Fig. 4(c)), navigation proceeds to the second stage. This stage mainly completes the accurate position measurement. The image captured by the long-focus camera is shown in Fig. 4(b), and the result of saliency detection using DTDMI is shown in Fig. 4(d). Through adaptive threshold segmentation, the target specific region is extracted from the image. Then, the minimum circumscribed rectangle of the target area is selected. The four vertices of the rectangle are used as marker points for pose estimation by the robust solution to the perspective-n-point problem (RPnP) algorithm[30]. The calculated relative position information is fed back to the UAV platform controller to provide the UAV platform navigation information in real time. The pose estimation diagram is shown in Fig. 5.

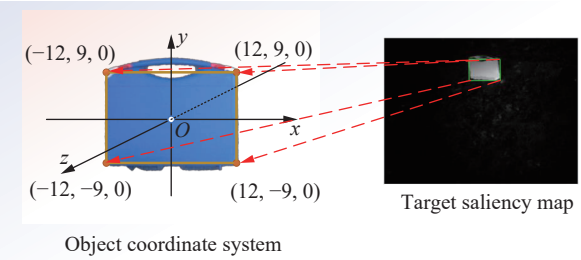


Fig. 5 Feature point matching

The translation matrix and rotation matrix of the object's coordinate system relative to the camera′s coordinate system can be obtained by the RPnP algorithm. Considering the navigation task requirements, visual information mainly provides the relative location relationship between the UAV and the target. Therefore, the rotation matrix is ignored, and the translation matrix is mainly used as the judgment basis. According to the x, y and z direction relative positions in the translation matrix, the position control command is fed back to the flight controller to complete the navigation process.

4 Experimental results

4.1 Image saliency

In this section, we have designed several experiments for dynamic target detection based on the midbrain inhibition (DTDMI) method proposed in this paper. The overall experiment was carried out in the Matlab R2020a environment, and the computer was configured as follows: Win10 system, processor for AMD Ryzen 5800H, 16 GB memory configuration. In the experiment, a variety of mature image saliency detection methods are used for comparison, including co-saliency (CS)[19], maximum symmetric surround (MSS)[31], covariance (COV)[32], visual attention model (VAM)[33], spectral residual (SR)[34], saliency aware (SA)[35], saliency optimization (SO)[36], saliency filters (SF)[37], geodesic saliency (GS)[38], manifold ranking (MR)[39], and co-saliency using implicit rank-sparsity (CoIRS)[40]. The algorithm is first tested on the densely annotated video segmentation (DAVIS)[41] and Freiburg-Berkeley motion segmentation (FBMS)[42] datasets. The DTDMI will then be tested in images acquired by an eagle-eye inspired multi-camera sensor.

Comparisons on DAVIS and FBMS datasets. The DAVIS dataset contains a total of 50 video sequences. The videos are densely annotated with high quality and high resolution. The dataset has two resolution images of 480p and 1080p. The FBMS dataset goes further on the basis of FBMS dataset. Compared with the FBMS dataset, the FBMS dataset adds 33 additional video sequences. Every video sequence is carefully annotated at the pixel level. These two datasets cover all kinds of situations of moving targets. It mainly includes target movement, target occlusion, etc.

We evaluate the DTDMI method in the above two datasets. As mentioned before, we compare our method with eight single image saliency methods: MSS, COV, VAM, SR, SO, SF, GS and MR. We also compare our method with three multiple image saliency methods: CS, CoIRS and SA. The SA method uses spatiotemporal information of image sequences to obtain a saliency map. Figs. 6–9 show the visual saliency detection result on two datasets. Among the comparison algorithms, SR and COV have no clear salient object boundary. This will bring difficulties to the subsequent target positioning process. Although SF and MSS can detect the target area, the target information is incomplete. For example, the above two saliency detection methods in swing sequences do not entirely detect all human body parts. The COV method cannot determine the shape of the target but can only roughly locate the target position in the image. GS are seriously disturbed by the image background. From the processing results of these methods, it is difficult to extract an accurate target area. In addition, SF and MR misidentified saliency targets in the cats01 sequence.

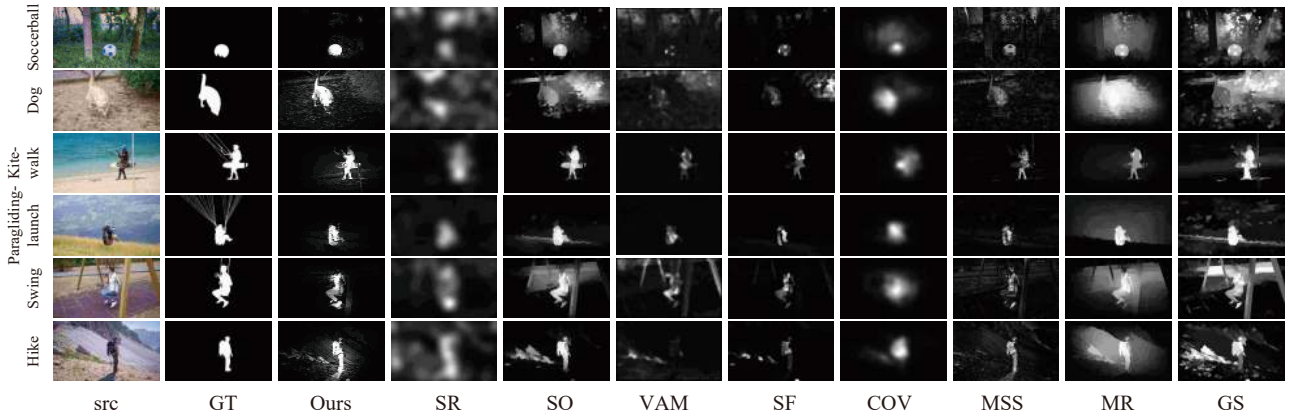


Fig. 6 Single image saliency detection results of the proposed method and eight other comparison methods on the DAVIS dataset

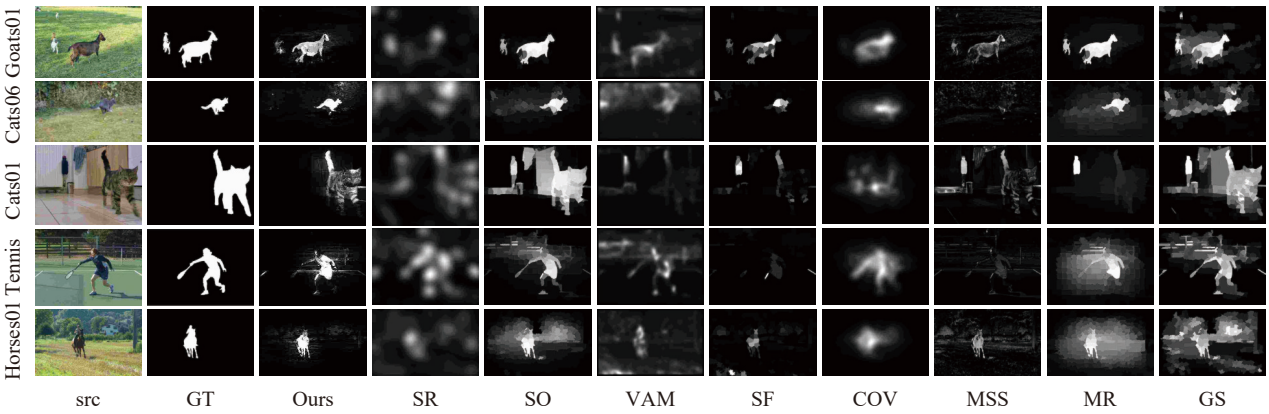


Fig. 7 Single image saliency detection results of the proposed method and eight other comparison methods on the FBMS dataset

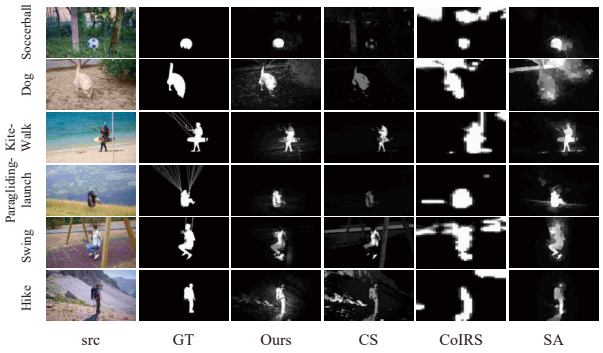


Fig. 8 Multiple image saliency detection results of the proposed method and three other comparison methods on the DAVIS dataset

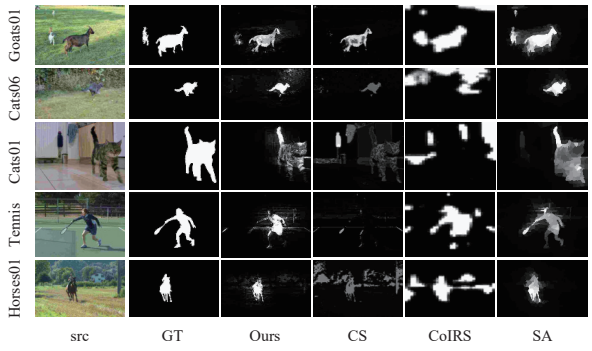


Fig. 9 Multiple image saliency detection results of the proposed method and three other comparison methods on the FBMS dataset

Moreover, COV is weakened because the salient target of this image sequence is near the edge. The proposed method has the best performance among the ten results. Compared with other methods, the DTDMI has a more apparent significant target boundary. For the cosaliency detection algorithm, the CS method uses the connection between multiple images to strengthen the saliency information. However, the interference information may be enhanced in sequential images. The saliency areas calculated by the CoIRS are relatively rough. Furthermore, the DTDMI method has a better inhibitory effect on background interference. The DTDMI method also works well when the target location is next to the edge of the image. It can be found that our method's result is closest to the ground truth.

Figs. 10 and 11 show the quantitative results. For both datasets, we use five methods for quantitative analysis. Precision versus recall curve analysis and receiver operating characteristic (ROC) curve analysis are quantitative indicators. To comprehensively measure the performance of the algorithm, the F-measure[39], mean absolute error (MAE), and area under the curve (AUC)[43] are also com-

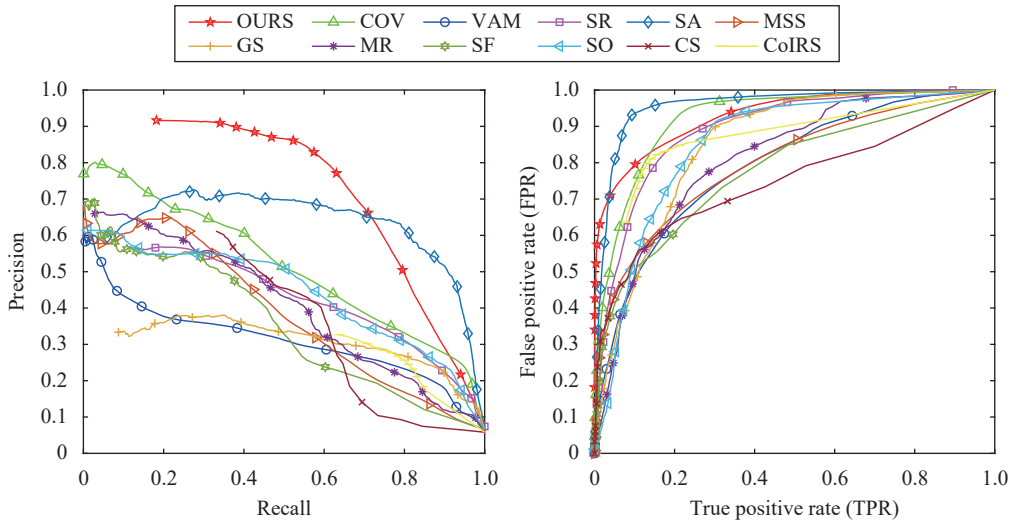


Fig. 10 Comparison of precision-recall (PR) curves and ROC curves on the DAVIS dataset

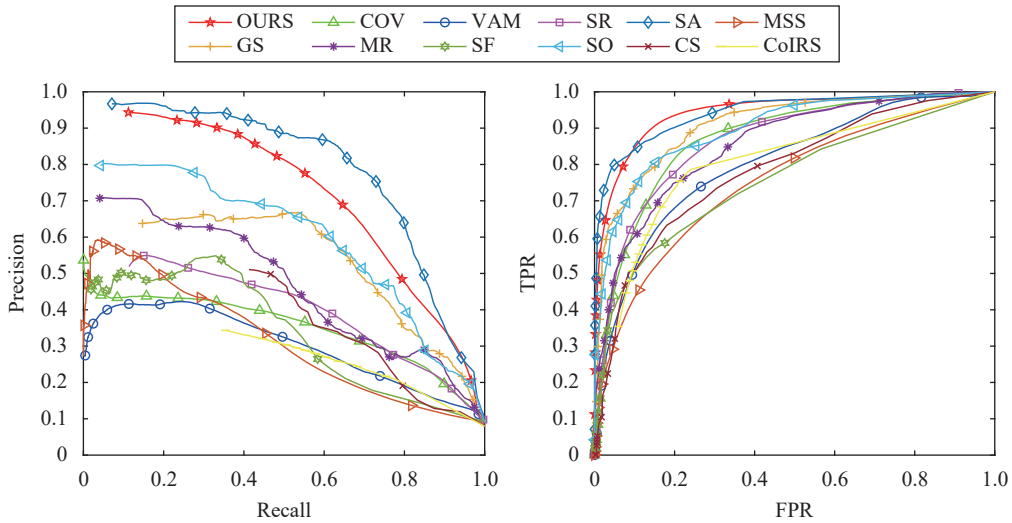


Fig. 11 Comparison of PR curves and ROC curves on the FBMS dataset

puted. The F-score in this experiment is calculated as follows:

$$F = \frac{\left(1+\beta^2\right) \times precision \times recall}{\beta^2 \times precision + recall} \tag{17}$$

where β^2 is the weight coefficient, which is used to adjust the importance of the precision relative to the recall. According to the recommendation in [44], β^2 is set to 0.3.

By comparing the data, our proposed method is found to be superior to other methods in most cases. The maximum F-score, AUC and MAE scores are shown in Table 1. Compared with other methods, the DTDMI method has better performance under these three evaluation criteria. The SA method has a higher AUC on the DAVIS dataset. SA also has a better F-score on the FBMS dataset.

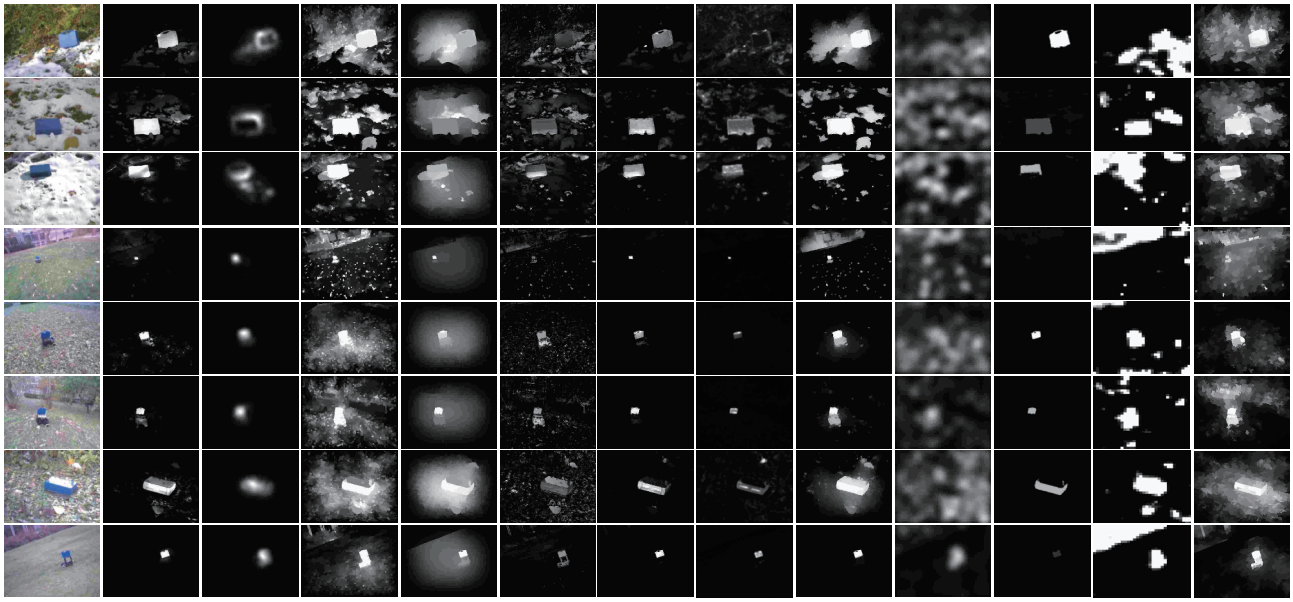
Comparisons on camera capture images. In this section, we use the eagle-eye inspired multi-camera sensor to take a set of measured images. The background in the image is mainly snow, land, and vegetation. Saliency objects to be detected are placed between greenery. In this scene, the target to be detected has specific color and shape characteristics. The saliency detection result is shown in Fig. 12. In the task setting, saliency detection results will be applied to the relative position measurement between the target and the camera. Therefore, methods such as SR, which cannot detect clear boundaries of targets, are not applicable. In addition, we can also see that SO, MR and GS do not distinguish background from target very well. The area around the target is also detected to be salient. In contrast, our proposed method can detect the target′s location accurately. Because of the combination of dynamic information in the algorithm, background interference is well suppressed.

4.2 Relative position estimation experiment

In the experiment, we used the eagle-eye inspired multi-camera sensor to collect images in real time and

Table 1 F-score and AUC of our method and other algorithms

Criteria	Dataset	COV	SR	MSS	GS	MR	SF	SO	VAM	CS	CoIRS	SA	Ours
F-score	DAVIS	0.528 5	0.449 6	0.443 0	0.349 6	0.419 0	0.438 7	0.493 3	0.317 7	0.441 7	0.362 5	0.655 5	**0.714 7**
AUC		0.923 1	0.894 6	0.802 6	0.851 1	0.812 0	0.779 9	0.854 0	0.793 7	0.749 7	0.834 1	**0.958 2**	0.930 3
MAE		0.084 0	0.184 4	0.096 1	0.205 2	0.202 6	0.074 3	0.159 3	0.117 5	0.077 0	0.140 0	0.087 6	**0.061 5**
F-score	FBMS	0.390 2	0.419 3	0.350 9	0.609 4	0.519 7	0.419 5	0.598 8	0.344 2	0.407 8	0.327 3	**0.763 5**	0.698 6
AUC		0.871 2	0.867 1	0.754 0	0.910 4	0.855 7	0.758 3	0.899 1	0.803 5	0.784 9	0.791 3	0.942 8	**0.943 1**
MAE		0.111 5	0.177 1	0.112 4	0.150 1	0.173 8	0.090 7	0.132 3	0.148 3	0.097 0	0.149 1	0.081 3	**0.066 9**



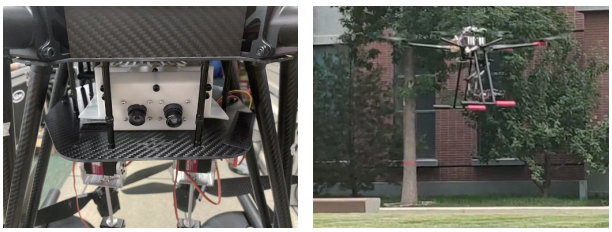
src Ours COV GS MR MSSS SF VAM SO SR CS CoIRS SA

Fig. 12 Saliency detection results of the proposed method and eleven other comparison methods on the camera capture image

process them online. The image processor adopts an NVIDIA Jetson nano with four core ARM A57 CPU and 64 GB storage. The camera installation and outdoor navigation tests are shown in Fig. 13.

The framework of the experimental process is shown in Fig. 14. In Period 1 of the experiment, the multi-camera sensor is approximately 6 m away from the target. In this period, the long-focus camera is mainly used as the image source. Target saliency detection in short-focus camera view and image matching based on SURF were performed simultaneously.

The detailed flow of image processing is as follows. After the image is acquired, the visual information is processed using the DTDMI method. The three-channel RGB image collected by the camera is processed as a single-channel grayscale saliency map. The grayscale image is then binarized with $T^{th} = 120$. Then, we use the minimum bounding rectangle (MBR) to select a connected region with a nonzero pixel value in the binary image. The center of the MBR serves as the basis for roughly measuring the offset of the target in the x and y directions under the camera coordinate system. The target to the camera is calculated by (16).



(a) (b)

Fig. 13 Hardware of the UAV navigation platform: (a) Device installation mode on the UAV; (b) UAV navigation experiment.

As the UAV approaches, the target enters the FOV of the long-focus camera. The target's relative position of the switch of the data source is mainly completed. The image sequence with target information is captured by a long-focus camera. The shape characteristics of the target as a priori information are known. We filter the final target area based on the aspect ratio and size of the MBR. The four vertex coordinates of the MBR can be acquired. As Fig. 5 shows, pixel points are matched to known three-dimensional coordinates to obtain the target's relative position to the camera.

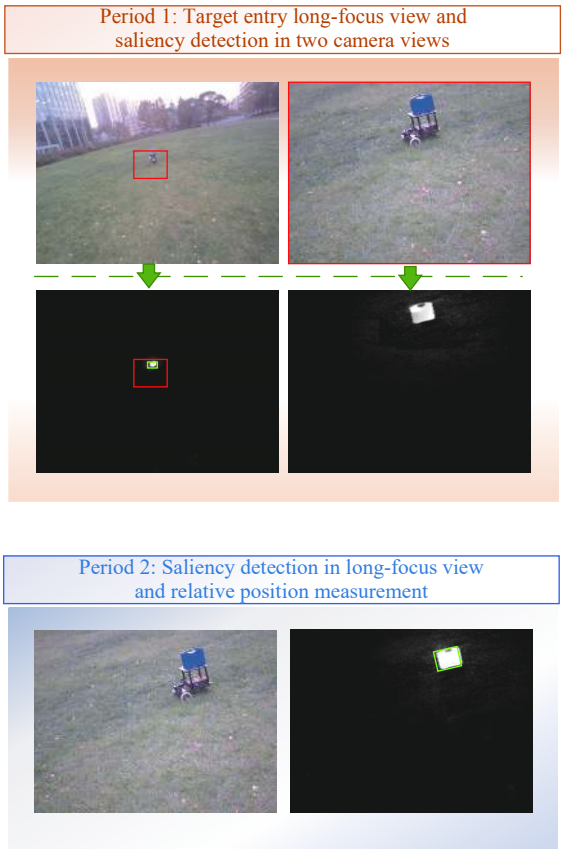


Fig. 14 Outdoor UAV navigation experiment flow

The target is within the FOV of the long-focus camera in Period 2. The navigation information of the UAV mainly comes from processing the image from a long-focus camera. Images are still processed using DTDMI methods. The DTDMI method can accurately detect saliency targets under the interference of the environment. After extracting the MBR where the target is located, the RPnP method is used for relative position measurement. The relative position measurement results using the RPnP method are shown in Fig. 15.

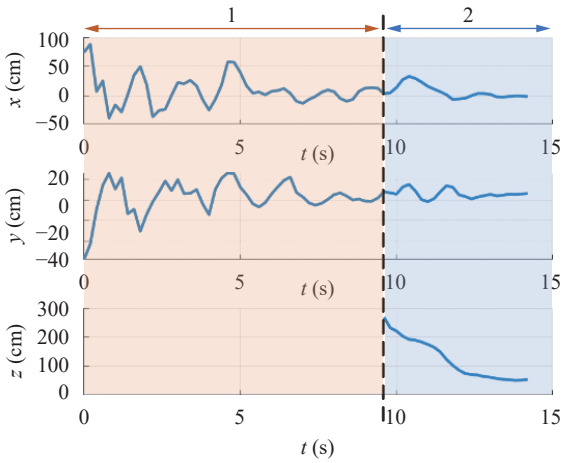


Fig. 15 Relative position measurement result

In the overall process of UAV navigation, the measured relative positions of the different cameras can be smoothly connected. During the experiment, the UAV keeps moving toward the target. Because of (16), only the relative position information parallel to the imaging plane direction can be obtained in Period 1. This period focuses on directional navigation to ensure that the UAV moves in the target direction. Period 1 also plays a connecting role. Navigation information transitions from estimation to accurate calculation. In Period 2, the exact relative position is calculated by the RPnP algorithm. UAV can navigate accurately with visual measurement information.

5 Conclusions

This paper proposes a novel eagle-eye inspired vision system based on a combination of a long-focus camera and a short-focus camera. An image processing method called DTDMI is proposed for salient target detection. Dynamic saliency features are extracted using the cross-inhibition mechanism of the eagle visual system. Comparative experiments are conducted, and the results verify that the eagle-eye inspired visual system can handle target detection challenges in outdoor UAV capture tasks. The DTDMI method can eliminate complex background interference and accurately detect salient targets. The relative position can be measured based on the multi-camera sensor. Experimental results demonstrate the feasibility and effectiveness of the designed biological visual system and the proposed target detection method.

Future works will focus on detection-tracking software design. In addition, due to its small size and light weight, the eagle-eye inspired multi-camera sensor can be used as a perception device on other unmanned vehicles, including the remote operated vehicle and unmanned surface vehicle.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Nos. T2121003, U1913602 and U19B2033), Science and Technology Innovation 2030–Key Project of “New Generation Artificial Intelligence”, China (No. 2018AAA0100803).

Declarations of conflict of interest

The authors declared that they have no conflicts of interest to this work.

References

[1] W. R. T. Roderick, M. R. Cutkosky, D. Lentink. Bird-inspired dynamic grasping and perching in arboreal environments. *Science Robotics*, vol. 6, no. 61, Article number eabj7562, 2021. DOI: 10.1126/scirobotics.abj7562.

[2] D. Alshamaa, A. Cherubini, R. Passama, S. Pla, L. Damm,

S. Ramdani. RobCap: A mobile motion capture system mounted on a robotic arm. *IEEE Sensors Journal*, vol. 22, no. 1, pp. 917–925, 2022. DOI: 10.1109/JSEN.2021.3128386.

[3] D. Morrison, P. Corke, J. Leitner. Learning robust, real-time, reactive robotic grasping. *The International Journal of Robotics Research*, vol. 39, no. 2–3, pp. 183–201, 2020. DOI: 10.1177/0278364919859066.

[4] J. Thomas, G. Loianno, K. Sreenath, V. Kumar. Toward image based visual servoing for aerial grasping and perching. In *Proceedings of IEEE International Conference on Robotics and Automation*, IEEE, Hong Kong, China, pp. 2113–2118, 2014. DOI: 10.1109/ICRA.2014.6907149.

[5] J. González-Martín-Moro, J. L. Hernández-Verdejo, A. Clement-Corral. The visual system of diurnal raptors: Updated review. *Archivos de la Sociedad Española de Oftalmología*, vol. 92, no. 5, pp. 225–232, 2017. DOI: 10.1016/j.oftal.2016.11.019.

[6] L. Reymond. Spatial visual acuity of the eagle Aquila audax: A behavioural, optical and anatomical investigation. *Vision Research*, vol. 25, no. 10, pp. 1477–1491, 1985. DOI: 10.1016/0042-6989(85)90226-3.

[7] Y. M. Deng, H. B. Duan. Biological eagle-eye-based visual platform for target detection. *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 6, pp. 3125–3136, 2018. DOI: 10.1109/TAES.2018.2845178.

[8] Q. Fu, S. T. Wang, J. Wang, S. N. Liu, Y. B. Sun. A lightweight eagle-eye-based vision system for target detection and recognition. *IEEE Sensors Journal*, vol. 21, no. 22, pp. 26140–26148, 2021. DOI: 10.1109/JSEN.2021.3120922.

[9] M. M. Cheng, J. Warrell, W. Y. Lin, S. Zheng, V. Vineet, N. Crook. Efficient salient region detection with soft image abstraction. In *Proceedings of IEEE International Conference on Computer Vision*, IEEE, Sydney, Australia, pp. 1529–1536, 2013. DOI: 10.1109/ICCV.2013.193.

[10] F. Bonnin-Pascual, A. Ortiz. A flying tool for sensing vessel structure defects using image contrast-based saliency. *IEEE Sensors Journal*, vol. 16, no. 15, pp. 6114–6121, 2016. DOI: 10.1109/JSEN.2016.2578360.

[11] Y. Zhang. Detection and tracking of human motion targets in video images based on Camshift algorithms. *IEEE Sensors Journal*, vol. 20, no. 20, pp. 11887–11893, 2020. DOI: 10.1109/JSEN.2019.2956051.

[12] X. C. Cao, Z. Q. Tao, B. Zhang, H. Z. Fu, W. Feng. Self-adaptively weighted co-saliency detection via rank constraint. *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 4175–4186, 2014. DOI: 10.1109/TIP.2014.2332399.

[13] M. Chancán, L. Hernandez-Nunez, A. Narendra, A. B. Barron, M. Milford. A hybrid compact neural architecture for visual place recognition. *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 993–1000, 2020. DOI: 10.1109/LRA.2020.2967324.

[14] Y. H. Luo, M. Xu, C. H. Yuan, X. Cao, L. Q. Zhang, Y. Xu, T. J. Wang, Q. Feng. SiamSNN: Siamese spiking neural networks for energy-efficient object tracking. In *Proceedings of the 30th International Conference on Artificial Neural Networks and Machine Learning*, Springer, Bratislava, Slovakia, pp. 182–194, 2021. DOI: 10.1007/978-3-030-86383-8_15.

[15] Z. Y. Yang, Y. J. Wu, G. R. Wang, Y. K. Yang, G. Q. Li, L. Deng, J. Zhu, L. P. Shi. DashNet: A hybrid artificial and spiking neural network for high-speed object tracking, [Online], Available: https://arxiv.org/abs/1909.12942, 2019.

[16] H. B. Duan, L. Xin, Y. Xu, G. Z. Zhao, S. J. Chen. Eagle-vision-inspired visual measurement algorithm for UAV's autonomous landing. *International Journal of Robotics and Automation*, vol. 35, no. 6, pp. 94–100, 2020. DOI: 10.2316/J.2020.206-0221.

[17] H. B. Duan, X. B. Xu, Y. M. Deng, Z. G. Zeng. Unmanned aerial vehicle recognition of maritime small-target based on biological eagle-eye vision adaptation mechanism. *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 5, pp. 3368–3382, 2021. DOI: 10.1109/TAES.2021.3075524.

[18] X. Li, H. B. Duan, J. C. Li, Y. M. Deng, F. Y. Wang. Biological eagle eye-based method for change detection in water scenes. *Pattern Recognition*, vol. 122, Article number 108203, 2022. DOI: 10.1016/J.PATCOG.2021.108203.

[19] H. Z. Fu, X. C. Cao, Z. W. Tu. Cluster-based co-saliency detection. *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 3766–3778, 2013. DOI: 10.1109/TIP.2013.2260166.

[20] S. Ohayon, W. Harmening, H. Wagner, E. Rivlin. Through a barn owl's eyes: Interactions between scene content and visual attention. *Biological Cybernetics*, vol. 98, no. 2, pp. 115–132, 2008. DOI: 10.1007/s00422-007-0199-4.

[21] J. Orlowski, O. Ben-Shahar, H. Wagner. Visual search in barn owls: Task difficulty and saccadic behavior. *Journal of Vision*, vol. 18, no. 1, Article number 4, 2018. DOI: 10.1167/18.1.4.

[22] C. A. Goddard, S. P. Mysore, A. S. Bryant, J. R. Huguenard, E. I. Knudsen. Spatially reciprocal inhibition of inhibition within a stimulus selection network in the avian midbrain. *PLoS One*, vol. 9, no. 1, Article number e85865, 2014. DOI: 10.1371/journal.pone.0085865.

[23] S. P. Mysore, E. I. Knudsen. Reciprocal inhibition of inhibition: A circuit motif for flexible categorization in stimulus selection. *Neuron*, vol. 73, no. 1, pp. 193–205, 2012. DOI: 10.1016/j.neuron.2011.10.037.

[24] W. Kim, C. Kim. Spatiotemporal saliency detection using textural contrast and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 4, pp. 646–659, 2014. DOI: 10.1109/TCSVT.2013.2290579.

[25] S. Potier, M. Mitkus, A. Kelber. High resolution of colour vision, but low contrast sensitivity in a diurnal raptor. *Proceedings of the Royal Society B-Biological Sciences*, vol. 285, no. 1885, Article number 20181036, 2018. DOI: 10.1098/rspb.2018.1036.

[26] V. A. Tucker. The deep fovea, sideways vision and spiral flight paths in raptors. *Journal of Experimental Biology*, vol. 203, no. 24, pp. 3745–3754, 2000. DOI: 10.1242/jeb.203.24.3745.

[27] P. Cornic, C. Illoul, A. Cheminet, G. Le Besnerais, F. Champagnat, Y. Le Sant, B. Leclaire. Another look at volume self-calibration: Calibration and self-calibration within a pinhole model of Scheimpflug cameras. *Measurement Science and Technology*, vol. 27, no. 9, Article number 094004, 2016. DOI: 10.1088/0957-0233/27/9/094004.

[28] K. Gao, H. Aliakbarpour, J. Fraser, K. Nouduri, F. Bunyak, R. Massaro, G. Seetharaman, K. Palaniappan. Local feature performance evaluation for structure-from-motion and multi-view stereo using simulated city-scale aerial imagery. *IEEE Sensors Journal*, vol. 21, no. 10,

pp. 11615–11627, 2021. DOI: 10.1109/JSEN.2020.3042810.

[29] C. Liu, Y. Huang, Y. M. Rong, G. Li, J. Meng, Y. L. Xie, X. L. Zhang. A novel extrinsic calibration method of mobile manipulator camera and 2D-LiDAR via arbitrary trihedron-based reconstruction. *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24672–24682, 2021. DOI: 10.1109/JSEN.2021.3111196.

[30] S. Q. Li, C. Xu, M. Xie. A robust $O(n)$ solution to the perspective-n-point problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1444–1450, 2012. DOI: 10.1109/TPAMI.2012.41.

[31] R. Achanta, S. Süsstrunk. Saliency detection using maximum symmetric surround. In *Proceedings of IEEE International Conference on Image Processing*, IEEE, Hong Kong, China, pp. 2653–2656, 2010. DOI: 10.1109/ICIP.2010.5652636.

[32] E. Erdem, A. Erdem. Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of Vision*, vol. 13, no. 4, Article number 11, 2013. DOI: 10.1167/13.4.11.

[33] H. B. Duan, Y. M. Deng, X. H. Wang, F. Liu. Biological eagle-eye-based visual imaging guidance simulation platform for unmanned flying vehicles. *IEEE Aerospace and Electronic Systems Magazine*, vol. 28, no. 12, pp. 36–45, 2013. DOI: 10.1109/MAES.2013.6693667.

[34] X. D. Hou, L. Q. Zhang. Saliency detection: A spectral residual approach. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Minneapolis, USA, 2007. DOI: 10.1109/CVPR.2007.383267.

[35] W. G. Wang, J. B. Shen, R. G. Yang, F. Porikli. Saliency-aware video object segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 20–33, 2018. DOI: 10.1109/TPAMI.2017.2662005.

[36] W. J. Zhu, S. Liang, Y. C. Wei, J. Sun. Saliency optimization from robust background detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Columbus, USA, pp. 2814–2821, 2014. DOI: 10.1109/CVPR.2014.360.

[37] F. Perazzi, P. Krähenbühl, Y. Pritch, A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Providence, USA, pp. 733–740, 2012. DOI: 10.1109/CVPR.2012.6247743.

[38] Y. C. Wei, F. Wen, W. J. Zhu, J. Sun. Geodesic saliency using background priors. In *Proceedings of the 12th European Conference on Computer Vision*, Springer, Florence, Italy, pp. 29–42, 2012. DOI: 10.1007/978-3-642-33712-3_3.

[39] C. Yang, L. H. Zhang, H. C. Lu, X. Ruan, M. H. Yang. Saliency detection via graph-based manifold ranking. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Portland, USA, pp. 3166–3173, 2013. DOI: 10.1109/CVPR.2013.407.

[40] Y. L. Chen, C. T. Hsu. Implicit rank-sparsity decomposition: Applications to saliency/co-saliency detection. In *Proceedings of the 22nd International Conference on Pattern Recognition*, IEEE, Stockholm, Sweden, pp. 2305–2310, 2014. DOI: 10.1109/ICPR.2014.400.

[41] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, A. Sorkine-Hornung. A benchmark dataset and evaluation methodology for video object segmentation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Las Vegas, USA, pp. 724–732, 2016. DOI: 10.1109/CVPR.2016.85.

[42] P. Ochs, J. Malik, T. Brox. Segmentation of moving objects by long term video analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 6, pp. 1187–1200, 2014. DOI: 10.1109/TPAMI.2013.242.

[43] A. Borji, L. Itti. State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, 2013. DOI: 10.1109/TPAMI.2012.89.

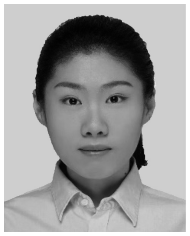
[44] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk. Frequency-tuned Salient Region Detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Miami, USA, pp. 1597–1604, 2009. DOI: 10.1109/CVPR.2009.5206596.

Yi-Min Deng received the B. Sc. and Ph. D. degrees in control science and engineering from School of Automation Science and Electrical Engineering, Beihang University, China in 2011 and 2017, respectively. He is currently an associate professor at School of Automation Science and Electrical Engineering, Beihang University, China. He is enrolled in the Young Elite Scientists Sponsorship Program by CAST (Chinese Association for Science and Technology) and Young Top Talent Support Program by Beihang University.

His research interests include biological computer vision and autonomous flight control.

E-mail: ymdeng@buaa.edu.cn (Corresponding author)

ORCID iD: 0000-0003-1533-3839



Si-Yuan Wang received the B. Sc. degree in automation from School of Automation and Electrical Engineering, University of Science and Technology Beijing, China in 2019. She is currently a master student in guidance, navigation and control at School of Automation Science and Electrical Engineering, Beihang University, China.

Her research interests include bioinspired computation and artificial intelligence.

E-mail: wang_siyuan@buaa.edu.cn

Machine Intelligence Research

Citation: Y. M. Deng, S. Y. Wang. Biological eagle–eye inspired target detection for unmanned aerial vehicles equipped with a manipulator. *Machine Inteligence Research*, vol.20, no.5, pp.741–752, 2023. https://doi.org/10.1007/s11633–022–1342–3