

Received February 24, 2019, accepted March 23, 2019, date of publication April 3, 2019, date of current version April 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2909098

# Unmanned Aerial Vehicle Perception System Following Visual Cognition Invariance Mechanism

QIRUI ZHANG<sup>1</sup> AND RUIXUAN WEI

Graduate School, Air Force Engineering University, Xi'an 710038, China

Corresponding author: Qirui Zhang (xianyangrui@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61573373.

**ABSTRACT** Humans have a fundamental ability that is fulfilling visual perception with complex brain functions, while intelligent sensors onboard UAVs do not have. The difficulties mainly lie in the inexhaustive study of brain science and the homologous mathematical description. In this paper, a visual cognition invariance mechanism has been proposed, which first review studies in brain research, moving from theoretical to practical. These terms are then reconsidered from brain pathways of visual perception. Next, a conceptual model of visual cognition invariance mechanism is proposed and the mathematical deduction is fulfilled. Furthermore, an experimental unmanned aerial vehicle is built to practically implement the proposed algorithm. The simulation results and experimental practices have validated the effectiveness and celerity of our method. Finally, a general discussion and proposals for addressing future issues are given.

**INDEX TERMS** Unmanned aerial vehicles, hybrid intelligent systems, object recognition, visual cognition.

## I. INTRODUCTION

To meet changing environmental demands, humans make rapid, strategic adjustments to how they deploy their intentional and cognitive resources [1], such that when we encounter similar but different difficulties, we tend to refocus our attention and recall memory-relevant aspects so as to work out recognition problems [2].

Although tremendous advances have been achieved in brain research, understanding own brain visual and cognitive system still remains one of the important research challenges to address, because knowing how we perceive the environment may allow us to replicate our biologic abilities into artificial systems, enabling us to build humanoid algorithms and machines for a large number of applications.

In theoretical research, Poggio [3], Marr and Hildreth [4], and Marr and Poggio [5] have tried to find out the human visual and cognitive systems as a homogeneous substrate, looking forward to explaining them with a few general wiring and plasticity rules. However, this view fades as time goes by, replaced by a highly pluralistic view that human vision system is inhomogeneous, with many specialized parts to be explored separately [6]. Of particular interests are the ventral stream [7] and dorsal stream [8].

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenbao Liu.

In practical experiments, intelligent machines draw lessons from human vision system are already happening. From smart cameras applying movement correction [9], to facial identification used in police and security operations [10], to the completely self-driving cars employing human vision models [11]. And for the near future, the perception system of autonomous unmanned aerial vehicles (UAVs) which uses human vision mechanism for reference is on the verge of happening [12], [13].

Despite there are extensive theoretical research and practical experiments, one of the major hurdles of these humanoid vision studies remains that they do not truly understand a scene like humans would [14], but rely on amounts of training and longtime computing. Replicating the process that humans perceive outside world remains an enormous challenge, the resolution of which would constitute a great contribution on the way to true humanlike perception system.

Biomimetic technology research has a great potential application. In 1944, Griffin [15] proposed that the blind people can obtain a skill, which is to navigate without eyes but using echolocation. Similar approaches using bats for reference can be found in radar and sonar systems [16]. Dolphin sonar has a better advantage compared with traditional sonar systems in environment perception [17], [18]. In complex environment, humanlike methods often provide excellent performance in perceiving objects and processing

information. Imai [19] proposed a humanlike three-layer model to better perceive outside world. In 2018, a system for autonomous vehicles is built to give autonomous vehicles the ability to process information like a human [20]. Regarding the cognition of intelligent creatures, Fawaz *et al.* [21] and Ding *et al.* [22] proposed cognitive architectures that use the concept of Internet of Things for the surveillance of amateur drones. More studies imitating human beings can be found in [23]–[25], indicating a huge prospect.

Here, a human vision inspired approach of UAV's onboard perception system is presented. We first take an explanation to humans' visual neural pathways in perception process, especially the visual information flow in dorsal stream and ventral stream. Then a humanlike model called parallel visual stream network (PVSN) which incorporates these neuroscience insights in a structured perception system of UAVs. In addition to the development of PVSN, we applied the model to a variety of visual cognition tasks that required rapidity and veracity. Finally the proposed model is tested against other cognizing models in a hypothetical scenario where multiple objects are situate in the environment, and the objects must be recognized as fast and accurate as models can. Furthermore, PVSN is experimented in a real scenario (in which there are two distinct objects, true and false) on the basis of a visual UAV platform, and the UAV must recognize the objects, localize the true objects and reach to them sequentially.

## II. BRAIN PATHWAYS OF VISUAL PERCEPTION

The brain pathways of visual perception have been illustrated by the neurophysiologist Fuster [26]. He pointed out that humans' visual perception is a complex system with multi channels. Every channel has its own task, they work independently as well as parallel to fulfill the general task of visual perception together. These parallel pathways including space and time channel, color information channel, left eye and right eye information channel, spatial azimuth information channel, etc. An exhausting description of all these brain pathways involved is impossible due to the lack of brain-scientific and neuro-scientific evidence. In this section we will briefly describe the human brain pathways and relating areas we think are critical to the proper functioning of visual perception architecture.

Han *et al.* [27] found that the pattern formed by movement can also cause the activation of the dorsal and ventral regions. This also supports the view that there is a synergy between the two pathways (“what” pathway and “where” pathway) separated in visual perception. Figure 1 shows the general description of visual perception. The visual information processing starts with retina, projecting through the lateral geniculate to the primary visual cortex, then to the senior visual cortex with further processing function [28]. Retinal cells are divided into large cells(M) and smalls cells(P) according to their functions. The large cells are responsible for processing information related to movement and brightness and then project to the V1 area. The small cells are related to colors and shapes, and then project to the V2 area. The visual

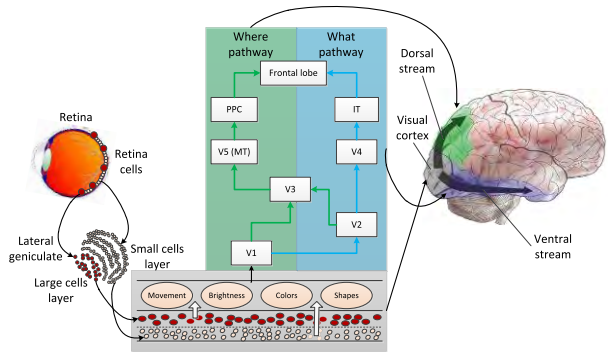


FIGURE 1. The two brain pathways in visual perception.

information is projected further into the outer striate cortex after V1 and V2 processing. In the outer striate cortex stage, the visual information processing is accomplished through two anatomical and functional pathways separately [29]. One is the dorsal visual pathway, which begins from V1, and goes through V2 and V3, then to the parietal lobe through MT area (visual area V5) and from there to the posterior parietal cortex (PPC). The dorsal stream, what also known as the “where pathway”, is related with visual stimuli for spatial location and motion information processing. The other is the ventral visual pathway, which starts from V1, and goes through V2 and V4, then to the inferior temporal cortex (IT). The ventral stream, what also known as the “what pathway”, is responsible for object characteristics including shape, color, size and texture, etc. The two pathways both start at V1 and V2, and terminate at the frontal lobe [30], what also called the visual cognition invariance in perceiving outside world [31], [32].

Regardless of whether the two pathways are encoded by a single neuron or multiple units, the knowledge of visual perception provide functional insights into mechanism of the brain that a modeler can seek to build a humanoid model by referring to human's visual cognition invariance mechanism.

## III. VISUAL COGNITION INVARIANCE MECHANISM

### A. OUTLINE OF PARALLEL VISUAL STREAM NETWORK

A graphical representation of our proposed architecture is given in Fig.2, which proposes that acting and perceiving upon objects of UAVs' perception system in environment is like in the case of two brain pathways in visual perception of humans, and each pathway is responsible for specific function.

In PVSN, objects are modeled as a combination of features and movements, which can be detected by cameras, providing input of environment information. Then the information is preprocessed to the formation of two visual maps, the object identity feature map (IFM) and the object spatial movement map (SMM), which stand for the map information of ventral pathway and dorsal pathway respectively. Bidirectional crosstalk between IFM and SMM ensures the object corresponds to the appropriate spatial location in the environment.

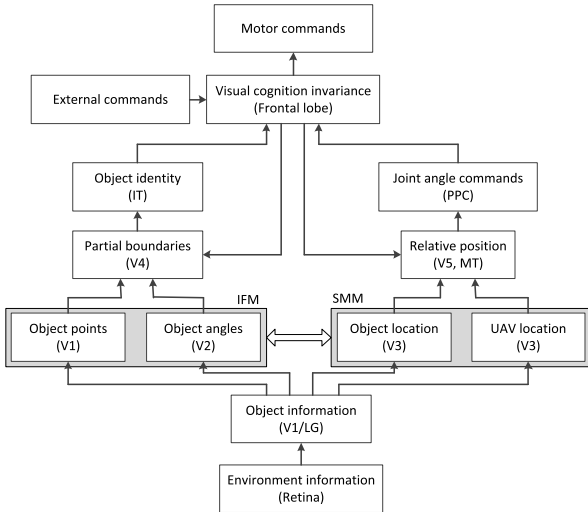


FIGURE 2. Outline of parallel visual stream network.

The IFM is then separated to individuals depends on partial boundaries module. A winner-take-all algorithm in object identity module is applied to ensure the object and spatial representation that reached resonance first will continue for the further processing. In the dorsal pathway, SMM is divided to tow maps, object location and UAV location. Once an object is chosen, the relative position between object and UAV is selected, and a library of joint angle commands is then also selected based on the former module. The visual cognition invariance module will select the opt object and the flying plan most relevant to the current context and suppress the irrelevant ones on the basis of real-time external commands and input information. Feedbacks of visual cognition invariance module will update the PVSVM configuration towards real human visual mechanism.

For the PVSN architecture, to achieve such complicate processes, a number of components are required to explicate, which are described in the following sections together with mathematical algorithms derivation.

**B. THE COMPONENTS AND ALGORITHMS OF PVSN**

Supposing the environment captured by camera is image  $I(x, y)$  and the environment information is in convolution with Gass kernel function  $G(x, y, \sigma)$  at different scales in order to obtain stable feature points [33]. Figure 3 shows the schematic of different scales constitute Gauss Pyramid. The environment information is converted to  $\Omega$  scales. Each scale is determined by scale layers  $S$  and scale factor  $k$ . The difference of Gaussians (DoG)  $D(x, y, \sigma)$  is applied showing the difference between the two adjacent scales of  $G$ :

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (1)$$

To object information module, it functions on converting chaotic data  $D(x, y, \sigma)$  to sorted data map (IFM and SMM) as shown in Fig.4, where  $OP$ ,  $OA$ ,  $OL$  and  $UL$  represents the feature of object points, angles, locations and UAV location respectively.

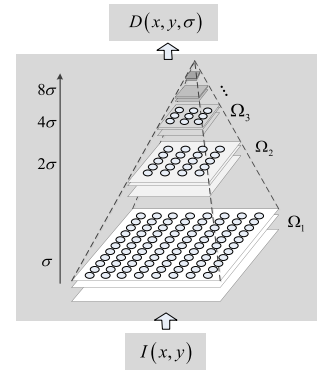


FIGURE 3. Schematic of the environment information module.

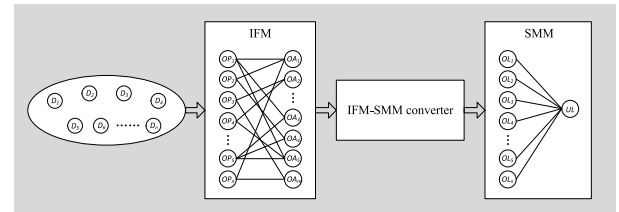


FIGURE 4. Schematic of the object information module.

For the  $k$ -th DoG  $D_k$ , its Taylor expansion at local extreme point  $(x_0, y_0, \sigma_0)$  is:

$$D_k(x, y, \sigma) = D_k(x_0, y_0, \sigma_0) + \frac{\partial D_k^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D_k^T}{\partial X^2} X \quad (2)$$

where the first and second order derivatives can be derived by neighborhood difference approximation [34].

Supposing the higher order derivative of  $D_k(x, y, \sigma)$  is zero, the precise extreme position  $X_{max}$  can be obtained as:

$$X_{max} = - \left( \frac{\partial^2 D_k}{\partial X^2} \right)^{-1} \frac{\partial D_k}{\partial X} \quad (3)$$

Once  $X_{max}$  is determined, to enhance the stability of matching and to improve noise immunity, the feature points with low contrast and unstable edge are removed by using Eq.(4) and Eq.(5) respectively.

$$D_k(X_{max}) = D_k + \frac{1}{2} \frac{\partial D_k^T}{\partial X} X_{max} \quad (4)$$

$$H_k = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (5)$$

where  $H_k$  is the partial derivative of Hessian matrix at the optimal feature point.

Assuming  $\alpha$  and  $\beta$  is the maximum and minimum eigenvalue, and  $\alpha = \gamma\beta$ :

$$\frac{Tr(H)^2}{Det(H)} = \frac{(D_{xx} + D_{yy})^2}{D_{xx}D_{yy} - (D_{xy})^2} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(\gamma + 1)^2}{\gamma} \quad (6)$$

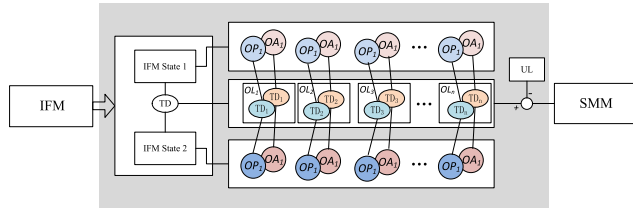


FIGURE 5. IFM-SMM converter.

We can infer from Eq.(6) that the threshold of feature points is only defined by the ratio of  $\alpha$  and  $\beta$ , it is relevant to specific eigenvalues. So if we assign a constant threshold  $\tau$ :

$$D_k = \Xi \cdot D_k \quad (7)$$

where  $\Xi$  is the threshold function  $\Xi = \begin{cases} 1, & \text{if } \gamma < \tau \\ 0, & \text{if } \gamma \geq \tau \end{cases}$ .

Furthermore, the gradient value  $m_k$  and direction angle  $\theta_k$  of feature points can be obtained by its neighborhood:

$$\theta_k(x, y) = \arctan \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \quad (9)$$

where  $L_k(x, y) = G_k(x, y, \sigma) * I(x, y)$ .

So IFM can be defined by:

$$\begin{cases} OP_k = \Xi \cdot X_{\max} \\ OA_k = [m_k, \theta_k] \end{cases} \quad (10)$$

Then IFM-SMM converter transforms IFM to SMM via applying temporal difference (TD). Figure 5 shows the schematic of IFM-SMM converter. The movement of feature point and corresponding angle can be defined by difference  $f_{TD}^k$  of the former state  $S_1$  (IFM state 1) and current state  $S_2$  (IFM state 2):

$$OL_k = f_{TD}^k(S_1, S_2) = \left\{ f_{TD}^k(OP_1^1, OP_1^2) \oplus f_{TD}^k(OA_1^1, OA_1^2) \right\} \quad (11)$$

where  $\oplus$  represents the sum mapping relation, indicating that SMM not only includes TD of the two components of IFM, but also involves the mapping relation between both.

For ventral stream, the processing of IFM is depicted in Fig.6. The ventral visual hierarchy has two modules, partial boundaries module (PBM) and object identity module (OIM). They work together to decompose IFM into distinct sets of features using partial boundary algorithm. And then every set is identified by color, shape and texture, finally the identities are integrated as output of object identity.

Assuming two feature points in IFM are  $D_q = \langle OP_q, OA_q \rangle$  and  $D_p = \langle OP_p, OA_p \rangle$ .  $\Gamma(D_q, D_p)$  is the function of PBM,

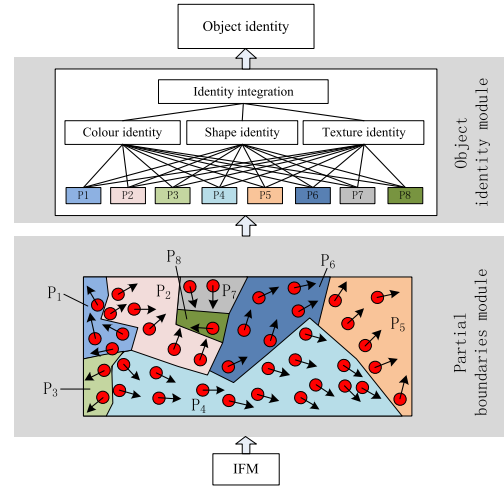


FIGURE 6. Schematic of information flow in ventral visual hierarchy.

to judge whether there is a coordination between  $D_q$  and  $D_p$ :

$$\Gamma(D_q, D_p) = \frac{\langle OP_q, OA_q \rangle \odot \langle OP_p, OA_p \rangle}{\langle OP_q, OA_q \rangle \otimes \langle OP_p, OA_p \rangle}, \quad \Gamma \in [-\omega, \omega] \quad (12)$$

where  $\odot$  represents the Euclidean distance similarity computation symbol,  $\otimes$  is the feature orientation similarity computation symbol.  $\omega$  is the threshold value, and  $+\omega$  (or  $-\omega$ ) indicates the max Euclidean distance uniformly (or not). When  $\Gamma(D_q, D_p) = 0$ , they are orthogonal between each other.

The region in same partial boundary  $P_k$  can be defined as:

$$\Gamma_k(D_q, D_p) \in [0, 1], \quad \forall D_q, D_p \in P_k \quad (13)$$

Finally each  $P$  is testified and classified by its color, shape and texture, then all the characteristics are integrated as a whole object identity value, as the object identity module in Fig.6 shows.

For dorsal stream, the input SMM has a map of object location and UAV location, but the relation between the two is independent, vague and uncertain. The network between objects is definite and precise, but the UAV is not in the net. As a result, the UAV can only know its location from remote ground stations, causing the time difference in controlling and inaccuracy in recognizing. But by combining the net and the UAV, the precise location can be calculated as the relative position module in Fig.7 shows.

Let  $E = (E_x, E_y, E_z)$  be the relative positions between the UAV and SMM network, for the  $k$ -th SMM:

$$E_k = \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}_k^T = \begin{bmatrix} OL_x - UL_x \\ OL_y - UL_y \\ OL_z - UL_z \end{bmatrix}_k^T \quad (14)$$

$$m_k(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (8)$$

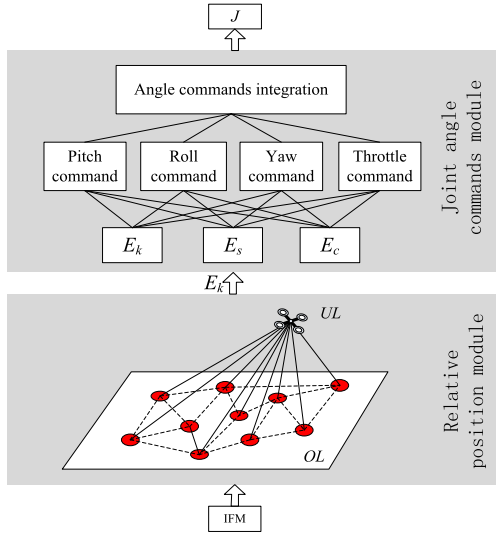


FIGURE 7. Schematic of information flow in dorsal visual hierarchy.

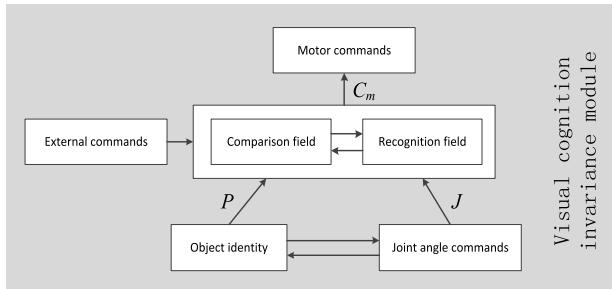


FIGURE 8. Schematic of the visual cognition invariance module.

The goal of joint angle commands module is to find the optimal joint angle commands  $J$ , within the permissible range of performance  $\mathfrak{R}$ , to minimize the  $\mathbf{E}_k$  between a calculation  $\mathbf{E}_c$  based on SMM and the information sensors onboard  $\mathbf{E}_s$ , which can be described as a Tikhonov minimization problem [35], [36]:

$$\min_{J \in \mathfrak{R}} \left( \sum_{i=1}^n \zeta_i \|E_c - E_s\| + \lambda (\mathbf{E}_k + J) \right) \quad (15)$$

where  $\zeta_i$  is the weight of  $i$ -th object.  $\mathbf{E}_k + J$  is to use control commands to minimize  $\mathbf{E}_k$ , which is controlled by the regularization parameter  $\lambda$ . The norm  $\|E_c - E_s\|$  penalizes large error to prevent over-fitting.

As shown in figure 8, the decision to which object to reach next is determined by the coordinated actions of the external commands, object identity and joint angle commands in the visual cognition invariance module. In its most basic form, an adaptive resonance theory (ART) [35] of two interconnected fields (the comparison field and recognition field) is applied.

In the VCIM model, to the  $k$ -th object identity  $\mathbf{P}_k$  and joint angle commands  $\mathbf{J}_k$ , the recognition field  $RF_k$  can be defined as:

$$RF_k(\mathbf{P}_k, \mathbf{J}_k) = (\mathbf{P}_k - \mathbf{J}_k)^T \sum_k^{-1} (\mathbf{P}_k - \mathbf{J}_k) \quad (16)$$

TABLE 1. The algorithm implementation of PVSN.

---

Input:  $I(x, y)$ , Initialize:  $\sigma, \tau, \omega, \lambda, f_{mr}, C^*$

for  $D(x_0, y_0, \sigma_0)$  do

$$X_{\max} = -\left(\frac{\partial^2 D_k}{\partial X^2}\right)^{-1} \frac{\partial D_k}{\partial X}$$

$$D_k(X_{\max}) = D_k + \frac{1}{2} \frac{\partial D_k^T}{\partial X} X_{\max}$$

$$D_k = \Xi \cdot D_k$$

for  $i = \{1: N_{objects}\}$  do

$$L_k(x, y) = G_k(x, y, \sigma) * I(x, y)$$

$$m_k(x, y) = \frac{\sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}}{L(x+1, y) - L(x-1, y)}$$

$$\theta_k(x, y) = \arctan \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}$$

$$IFM: \begin{cases} OP_k = \Xi \cdot X_{\max} \\ OA_k = [m_k, \theta_k] \end{cases}$$

end for

for  $j = \{1: N_{objects}\}$  do

$$SMM: OL_k = f_{TD}^k(S_1^j, S_2^j) = \{f_{TD}^k(OP_1^j, OP_1^j) \oplus f_{TD}^k(OA_1^j, OA_1^j)\}$$

end for

for  $p = \{1: N_{objects}\}, q = \{1: N_{objects}\}, p \neq q$

$$\Gamma_k(D_q, D_p) = \frac{\langle OP_q, OA_q \rangle \square \langle OP_p, OA_p \rangle}{\langle OP_q, OA_q \rangle \otimes \langle OP_p, OA_p \rangle}, \Gamma \in [-\omega, \omega]$$

if  $\Gamma_k(D_q, D_p) \in [0, 1]$

$$\{D_q, D_p\} \in P_k$$

else  $\Gamma_k(D_q, D_p) \in [0, 1]$

$$D_q \in P_k, D_p \notin P_k$$

end if

end for

if  $\min_{J_i \in \mathfrak{R}} \left( \sum_{i=1}^n \zeta_i \|E_c - E_s\| + \lambda (\mathbf{E}_k + J) \right)$

$$J_{optimal} = J_k$$

$$RF_k(\mathbf{P}_k, \mathbf{J}_k) = (\mathbf{P}_k - \mathbf{J}_k)^T \sum_k^{-1} (\mathbf{P}_k - \mathbf{J}_k)$$

$$f(RF_k) = \frac{\sum_{i=1}^n cf_i \cdot RF_k}{\sum_{i=1}^n cf_i}$$

if  $f(RF_k) \cdot C^* \leq f_{thr}$

$$C_m = J \mapsto C_m$$

else

$$C_m = J_{new}$$

end if

end if

end for

---

where  $\sum_k$  is the correlation covariance matrix between  $\mathbf{P}_k$  and  $\mathbf{J}_k$ .

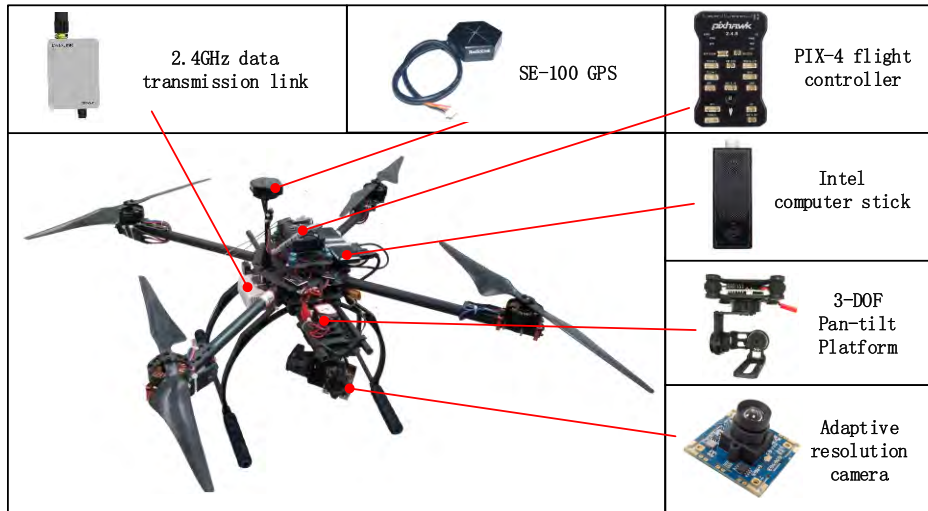


FIGURE 9. Perceiving and cognizing UAV platform.

To the comparison field  $CF = \{cf_1, cf_2, \dots, cf_n\}$ , the comparison function can be defined as :

$$f(RF_k) = \frac{\sum_{i=1}^n cf_i \cdot RF_k}{\sum_{i=1}^n cf_i} \quad (17)$$

Finally the motor commands  $C_m$  is defined as:

$$C_m = \begin{cases} J \mapsto C_m, & \text{iff } (RF_k) \cdot C^* \leq f_{thr} \\ J_{new}, & \text{iff } (RF_k) \cdot C^* > f_{thr} \end{cases} \quad (18)$$

where  $\mapsto$  is the mapping symbol, which turns joint angle commands to corresponding motor commands.  $C^*$  and  $f_{thr}$  represents motor commands data base and its threshold value respectively.  $J_{new}$  is the set of new joint angle commands preparing to add to  $C^*$ .

### C. THE ALGORITHM IMPLEMENTATION OF PVSN

To fulfill the proposed method, the detailed algorithm implementation of PVSN should be given. Here we use pseudocode (see in Table 1) to describe how the algorithm runs.

The environment is perceived by the camera onboard. To one of the captured frames, it is regarded as an input picture  $I(x, y)$ . Before applying the proposed algorithm, the parameters  $\sigma, \tau, \omega, \lambda, f_{thr}, C^*$  should be initialized. For every DoG of  $I(x, y)$ , we first calculate  $X_{max}$  and  $D_k$  to remove the unsuitable feature points. Then to every objects, the object features and its corresponding angles are calculated, indicating the features of different objects. So IFM of can be obtained by the combination of  $OP_k$  and  $OA_k$ . Next, the IFM-SMM converter can transfer IFM to SMM by using TD method so as to get  $OL_k$ . On the basis of IFM and SMM, the partial boundary algorithm is applied to calculate  $\Gamma_k(D_q, D_p)$ , so as to define whether  $D_q$  and  $D_p$  belongs to the same region or not. Finally, the optimal joint angle commands  $J_{optimal}$  can be obtained by finding the optimal solution of

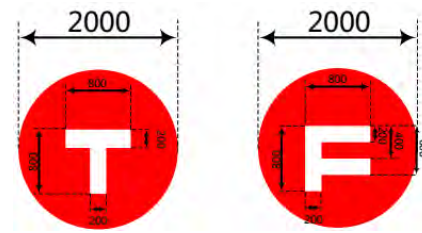


FIGURE 10. The true and false objects.

Tikhonov minimization problem, and the motor commands  $C_m$  is defined by the relation between  $f(RF_k) \cdot C^*$  and  $f_{thr}$ .

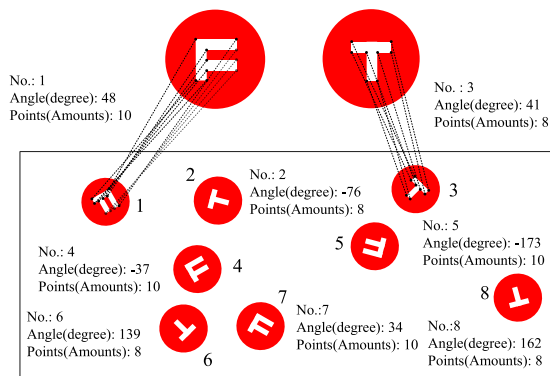
### IV. VISUAL COGNITION INVARIANCE MECHANISM

In Sections 2 and 3 we presented a visual cognition invariance mechanism of humans and a multiple hypotheses framework to follow VCIM associated with UAV flight. This section presents an experimental validation of the presented approach in the UAV perception and cognition context.

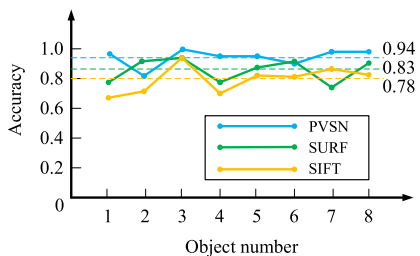
The experiments have been conducted on the perceiving and cognizing UAV (P&C-UAV) as Figure 9 shows. The P&C-UAV perceives environment from its embedded adaptive resolution camera on the basis of 3-DOF pan-tilt platform, and current flight status (altitude, attitude, flight speed and GPS position, etc.) from highly integrated sensors in PIX-4 flight controller. Then the environment images and status information are transmitted to airborne Intel computer stick, where the images and information are processed to generated flight offset instructions. Finally, the instructions are transported to PIX-4, which converts the instructions to motor commands. Furthermore, to monitor whether the algorithms and modules are functioning well or not, a wireless data transmission link (2.4GHz) between UAV and ground station is built to observe the real-time operation of processes.

### V. EXPERIMENTS

The process described in Sec. 3 generated successful VCIM and UAV's maneuvering policies. The mechanisms and poli-



(a)



(b)

FIGURE 11. IFM simulation result and comparisons. (a) IFM simulation result. (b) Comparisons of PVSN, SIFT and SURF.

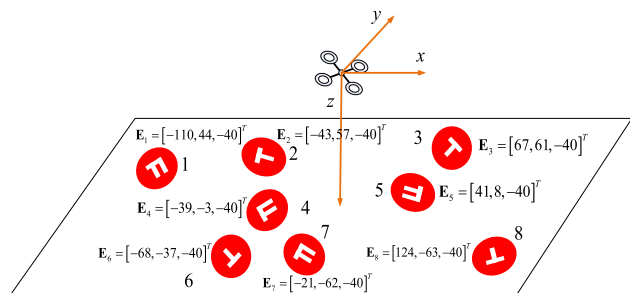
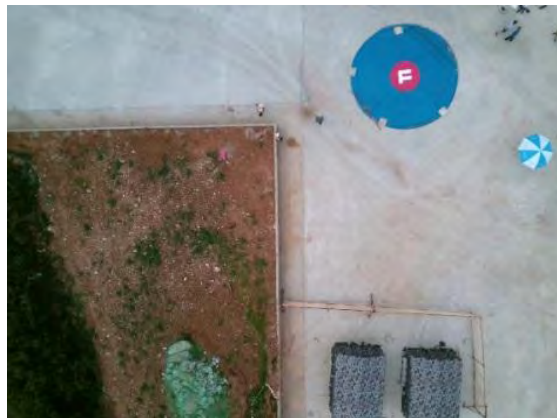


FIGURE 12. SMM simulation result.

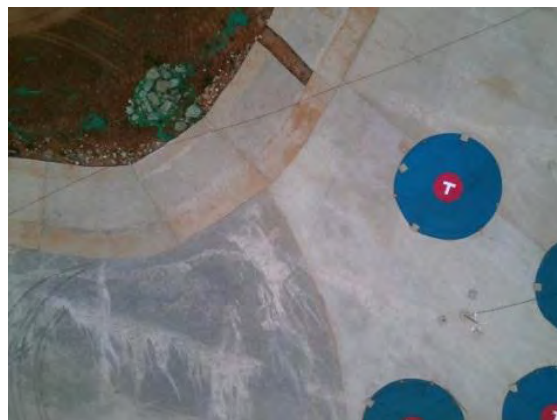


FIGURE 13. Experimental environment with multiple objects.

cies were tested using a computer simulation as well as P&C-UAV flight test. Sections 5.1 and 5.2 describe the simulated training and test results. Section 5.3 describes the experimental results applying real-time PVSN on a P&C-UAV.



(a)



(b)



(c)

FIGURE 14. Three experimental images from P&C-UAV. (a) Experimental image 1. (b) Experimental image 2. (c) Experimental image 3.

A. PVSN SIMULATION TRAINING

The visual cognition invariance mechanism naming PVSN was trained in Matlab 16a, Inter Core i5. Suppose the objects are in two classes, the true object with letter “T” and the false object with letter “F”, the standard size of objects is shown in Fig. 10.

Take the true and false objects in Fig.10 as an example, the shape feature is extracted as a benchmark, define the

TABLE 2. Comparisons of three methods of experimental image 1.


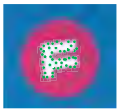




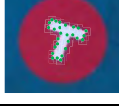
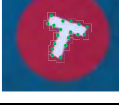
Algorithm	Experimental image	Feature points	True similarity	False similarity	Calculation time/(s)	Relative position/(m)
SIFT			0.21	0.79	0.4	/
SURF			0.17	0.83	0.31	/
PVSN			0.06	0.94	0.08	[28,35,-40]

TABLE 3. Comparisons of three methods of experimental image 2.

Algorithm	Experimental image	Feature points	True similarity	False similarity	Calculation time/(s)	Relative position/(m)
SIFT			0.77	0.23	0.35	/
SURF			0.81	0.19	0.29	/
VCIM			0.92	0.08	0.06	[37,12,-39]

midpoint of shape features as the center of rotation and the features are rotated, clockwise and anticlockwise, 90 times with a step of 2 degrees, therefore there are 180 samples of objects.

A test image and its simulation results (IFM and SMM) is shown in Fig.11 and Fig.12, respectively. In IFM, the objects points and angles are compared with samples to get object identity (whether it is true or false object). The simulation result shows that after PVSN training, an object can be mapped with samples with an average accuracy rate 0.94 as shown in Fig.11(b), which is much higher than scale-invariant feature transform (SIFT, with accuracy result 0.78) [38] and speeded up robust features (SURF, with accuracy result 0.83) [39].

Assuming the flight height of UAV is 40 meters, in SMM simulation result (as is shown in Fig.12), the relative positions between UAV and SMM network are depicted with three-dimensional space coordinates (take the UAV as the coordinate system), which shows the UAV was able to localize itself and objects.

**B. EXPERIMENTAL RESULTS ON A P&C-UAV**

In Section 3 we presented a multiple hypotheses framework and algorithms to fulfill VCIM via two pathways and we showed how it allows to improve the intelligence and accuracy with regard to the constraints associated with UAV flight. This part presents an experimental validation of the suitability of the presented approach in the P&C-UAV context.

Fig.13 shows the experimental environment, in which the multiple objects are put on the ground and a P&C-UAV is flying with equipment listed in Section 4 onboard. To confirm the VCIM's effectiveness, comparison experiments on P&C-UAV are done using two typical algorithms (SIFT and SURF) and the proposed one.

The UAV sends the images from its embedded camera to the ground station (personal computer, PC) via a wireless analogical link of 2.4GHz. Images are processed on the ground station with SIFT, SURF and VCIM. Take three images as an example (as is shown in Fig14(a), (b) and (c)), the task considered was to autonomously distinguish multiple true and false objects regarding the real-time flight.



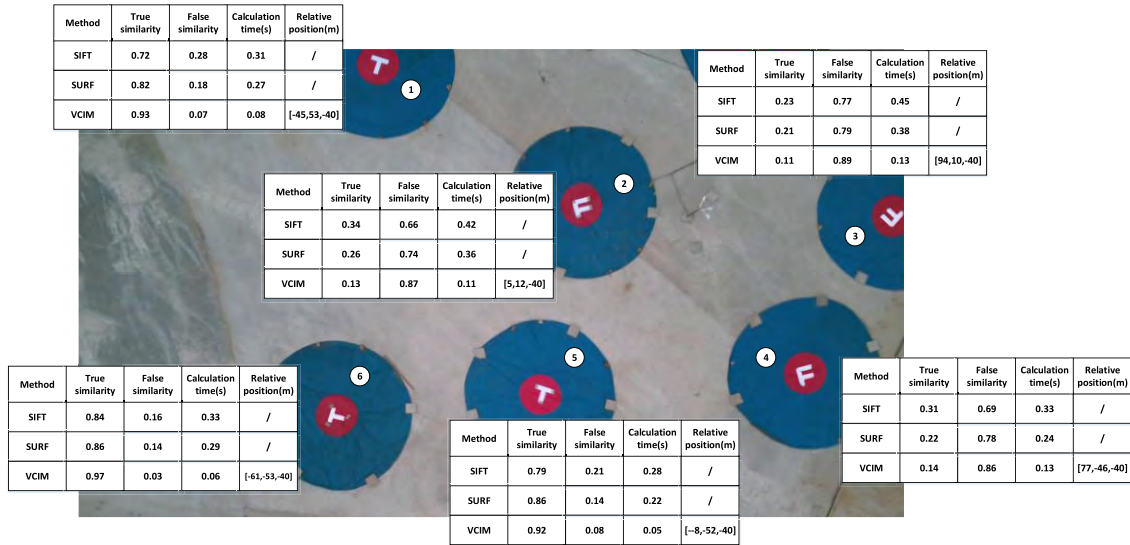


FIGURE 15. Experimental results of multiple objects.

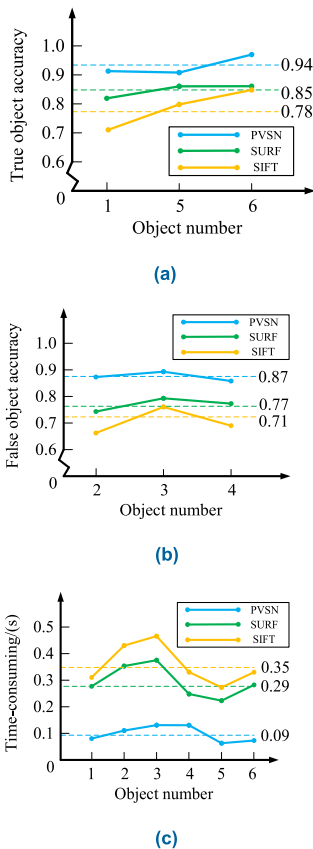


FIGURE 16. Comparison results of SIFT, SURF and PVSN. (a) Comparisons of true objects. (b) Comparisons of false objects. (c) Comparisons of time-consuming.

Table 2 and Table 3 show the experimental results of three methods, which indicate the proposed method has a higher accuracy in calculating similarity of true and false objects for it has fewer feature points, and the calculation time of PVSN performs much faster processing ability (which cost less than

1/3 time compared with the other two). Furthermore, PVSN can get the relative position between object and UAV via two pathways, giving out the motion commands to UAV to adjust flight, while SIFT and SURF can only recognize objects.

For Fig. 15(c), there are multiple objects needing recognize. So the objects are sorted and the results are labeled on the image.

To make a clear comparison of three methods, the results are classified and recognized in Fig.16. In Fig.16 (a) and (b), the similarity of true objects and false objects are drew with broken lines, indicating the processing accuracy of each object. Result shows that PVSN has higher accuracy in classifying and recognizing objects. Furthermore, the proposed method costs much less time in image processing and data calculating than the other two methods (see in Fig.16(c)).

## VI. CONCLUSIONS

In terms of humans, we have argued how VCIM can follow brain-like way in perceiving and cognizing outside world, which is similar to natural “dorsal stream” and “ventral stream” in visual processing. After reviewing terminology in the context of brain and visual perception, a conceptual constructive model for objects cognizing has been proposed. Following are some concluding remarks.

(a). A humanoid mechanism and algorithm is built referring to two brain pathways in visual perception. The proposed VCIM not only performs well in simulations but also in experiments.

(b). Compared with the other two methods, PVSN has a better performance both in simulated and experimental environment. It has higher accuracy in recognizing objects and has less time in calculating. Besides, PVSN can obtain the relative position between UAV and objects while SIFT and SURF unable to.

(c). The proposed constructive mechanism is expected to shed new insight on our understanding of visual pathways in

brain, which can directly reflected in the design of humanlike algorithms. Furthermore, a P&C-UAV platform is built to confirm the proposed method is practically feasible.

Still, there are several issues in need of attention and further investigations, including dynamic objects and multi-UAV cooperative recognition.

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their careful review and helpful suggestions that greatly improved the manuscript. They would also like to thank Hongxuan Lu and Zhuofan Xu for suggesting improvements after reading early versions of this manuscript.

## REFERENCES

- [1] T. Moore and M. Zirnsak, "Neural mechanisms of selective visual attention," *Annu. Rev. Neurosci.*, vol. 18, no. 1, pp. 193–222, 2017.
- [2] T. Egnér and J. Hirsch, "Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information," *Nature Neurosci.*, vol. 8, no. 12, pp. 1784–1790, Dec. 2005.
- [3] T. Poggio, "A theory of how the brain might work," *Cold Spring Harbor Symposia Quant. Biol.*, vol. 55, no. 55, p. 899, 1990.
- [4] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London. B, Biol. Sci.*, vol. 207, pp. 187–217, Feb. 1980.
- [5] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proc. Roy. Soc. London B, Biol. Sci.*, vol. 204, no. 1156, pp. 301–328, 1979.
- [6] G. A. Rousselet and S. J. J., "Thorpe and M. Fabrethorpe, How parallel is visual processing in the ventral pathway?: Trends in Cognitive Sciences," *Trends Cognit. Sci.*, vol. 8, no. 8, pp. 363–370, 2004.
- [7] M. Cauchoux, "Fast ventral stream neural activity enables rapid visual categorization," *NeuroImage*, vol. 125, no. 280, pp. 280–290, 2016.
- [8] K. Sakreida, "Affordance processing in segregated parieto-frontal dorsal stream sub-pathways," *Neurosci. Biobehavioral Rev.*, vol. 69, no. 89, pp. 89–112, 2016.
- [9] Y. Wang, "Mobile image based color correction using deblurring," *Comput. Imag. XIII*, vol. 9401, Mar. 2015, Art. no. 940107.
- [10] I. S. Hemdan, A. Karungaru and K. Terada, "Facial features-based method for human tracking," in *Proc. Workshop Frontiers Comput. Vis.*, Feb. 2011, pp. 1–10.
- [11] H. Surden and M. A. Williams, "Technological opacity, predictability, and self-driving cars," *Social Sci. Electron.*, vol. 38, p. 121, Jan. 2016.
- [12] A. R. Cummings, "UAV-derived data for mapping change on a swidden agriculture plot: Preliminary results from a pilot study," *Int. J. Remote Sens.*, vol. 38, nos. 8–10, pp. 2066–2082, 2017.
- [13] S. Y. Guznov, "Visual search training techniques in a UAV simulator environment: Pilots, performance, workload, and stress," Ph.D. dissertation, Univ. Cincinnati, Cincinnati, OH, USA, 2008.
- [14] G. Donald, "Visual Turing test for computer vision systems," *Proc. Nat. Acad. Sci.*, vol. 112, no. 12, pp. 23–3618, Mar. 2015.
- [15] D. R. Griffin, "Echolocation by blind men, bats and radar," *Science*, vol. 100, no. 2609, pp. 589–590, 1944.
- [16] A. Balleri *et al.*, "Special section on biologically-inspired radar and sonar systems - Analysis of acoustic echoes from a bat-pollinated plant species: insight into strategies for radar and sonar target classification," *IET Radar Sonar Navigat.*, vol. 6, no. 6, pp. 536–544, Jul. 2012.
- [17] W. W. Au and D. W. Martin, "Insights into dolphin sonar discrimination capabilities from human listening experiments," *J. Acoust. Soc. Amer.*, vol. 86, no. 5, pp. 1662–1670, Nov. 1989.
- [18] P. Dobbins, "Dolphin sonar—modelling a new receiver concept," *Bioinspiration Biomimetics*, vol. 2, no. 1, pp. 19–29, 2007.
- [19] M. Imai, "Humanlike conversation with gestures and verbal cues based on a three-layer attention-drawing model," *Connection Sci.*, vol. 18, no. 4, pp. 379–402, Dec. 2006.
- [20] L. Li, K. Ota, and M. Dong, "Humanlike driving: Empirical decision-making system for autonomous vehicles," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 6814–6823, Aug. 2018.
- [21] W. Fawaz, C. Abou-Rjeily, and C. Assi, "UAV-aided cooperation for FSO communication systems," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 70–75, Aug. 2018.
- [22] G. Ding, Q. Wu, L. Zhang, Y. Lin, T. A. Tsiftsis, and Y.-D. Yao, "An amateur drone surveillance system based on the cognitive Internet of Things," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 29–35, Jan. 2018.
- [23] M. Bearzi and C. A. S. Bigger, "Better brain: Observations of chimpanzees and dolphins strengthen the notion that humanlike intelligence may not be uniquely human," *Amer. Scientist*, vol. 98, no. 5, pp. 402–409, 2010.
- [24] L. Robertsson, B. Iliev, R. Palm, and P. Wide, "Perception modeling for human-like artificial sensor systems," *Int. J. Hum. Comput. Stud.*, vol. 65, no. 5, pp. 446–459, May 2007.
- [25] K. Madani *et al.*, "A human-like visual-attention-based artificial vision system for wildland firefighting assistance," *Appl. Intell.*, vol. 7, pp. 1–23, Sep. 2018.
- [26] J. M. Fuster, "Frontal lobe and cognitive development," *J. Neurocytol.*, vol. 31, nos. 3–5, pp. 373–385, 2002.
- [27] S. Han, "Neural substrates for visual perceptual grouping in humans," *Psychophysiology*, vol. 38, no. 6, pp. 926–935, 2010.
- [28] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, 1962.
- [29] L. G. Ungerleider and J. V. Haxby, "What' and 'where' in the human brain," *Current Opinion Neurobiol.*, vol. 4, no. 2, pp. 157–165, 1994.
- [30] N. K. Logothetis and J. T. P. Poggio, "Shape representation in the inferior temporal cortex of monkeys," *Current Biol.*, vol. 5, no. 5, pp. 552–563, Aug. 1995.
- [31] X. Miao and R. P. Rao, "Learning the Lie groups of visual invariance," *Neural Comput.*, vol. 19, no. 10, pp. 2665–2693, 2007.
- [32] Z. C. Wen, "Feature extraction based on visual invariance and species identification of weed seeds," *Trans. Chin. Soc. Agricult. Eng.*, vol. 27, no. 3, pp. 631–635, May 2011.
- [33] S. Ekvall, D. Kragic, and P. Jensfelt, "Object detection and mapping for service robot tasks," *Robotica*, vol. 25, no. 2, pp. 175–187, 2007.
- [34] S. Se and D. J. Lowe, "Little Vision-based mobile robot localization mapping using scale-invariant features," in *Proc. IEEE Int. Conf. Robot. Automat.*, Aug. 2001, pp. 1–10.
- [35] J. Taylor and G. Saliency, "Attention, active visual search, and picture scanning," *Cognit. Comput.*, vol. 3, no. 1, pp. 1–3, 2011.
- [36] G. Deco and E. T. Rolls, "A Neuro dynamical cortical model of visual attention and invariant object recognition," *Vis. Res.*, vol. 44, no. 6, pp. 621–642, 2004.
- [37] E. O. Postma and P. T. W. Hudson, "Adaptive resonance theory," in *Handbook of Brain Theory Neural Networks*. vol. 931. Cambridge, MA, USA: MIT Press, 1995, pp. 87–90.
- [38] C. J. Xi and S. X. Guo, "Image target identification of UAV based on SIFT," *Procedia Eng.*, vol. 15, pp. 3209–3215, May 2011.
- [39] W. Jia, "Fast algorithm based on SURF for UAV target identification," *Comput. Eng. Appl.*, to be published.



**QIRUI ZHANG** received the B.S. and M.S. degrees in control science and engineering from Air Force Engineering University, China, in 2014 and 2017, respectively, where he is currently pursuing the Ph.D. degree in control science and engineering.



**RUIQUAN WEI** received the Ph.D. degree from Xi'an Jiaotong University, Xi'an, China. He is currently a Professor with Air Force Engineering University. His research interest includes control science and engineering.

• • •