A Bionic Method of Moving Object Detection with Multi-feature Fusion Based On Frog Vision Characteristics

Xiaogang Tang¹,², Sun’an Wang¹, Hongyu Di¹, Litian Liu²
¹School of mechanical Engineering, Xi’an Jiaotong University, Xi’an, Shaanxi, P.R. China
²Department of Information Equipment, Academy of Equipment, Beijing, P.R. China

Abstract - In the complex natural background, the image features of moving objects usually change severely. And the kinematics and morphological features of dynamic target are unconspicuous due to the fast movement, unpredictable kinetic law and the accompanied scale transformation. The methods of motion detection based on one single morphological, statistics or kinetic features would not meet the requirements. Inspired by the visual characteristics of frog eye and the physiology characteristics of dynamic response of the retinal neural circuit, a spatial-temporal moving target recognition method based on the frog’s vision is provided. This method introduces a biomimetic recognition algorithm of multi-feature fusion for dynamic target detection based on BP neural network. The experimental results show that the method achieves to inhibit the background information effectively and enhance the multidimensional moving target information through the mechanism of spatial and temporal characteristics and multi-feature fusion, which is better than the method based on single feature. The algorithm principle provides a biomimetic approach for motion detection.

Keywords: Frog’s vision characteristics, Complex natural background, Moving object detection, Multi-feature fusion, BP neural networks

1 Introduction

Moving object detection is one key technology of infrared tracking, precision-guided, air defense alerting, large field target detection and satellite remote sensing [1]. The effectiveness and real-time performance of the moving object detection algorithm will directly influences the precision and efficiency of the subsequent video servo and tracking. Moving object detection in complex natural background is much closer to practical application and thus attracts much attention. But due to the complicated background and unpredictable motion, the feature of dynamic target images is unstable. In addition, the target is usually of small size and accompanied with scale transformations like rotation, scaling, etc. Therefore, the moving object detection in complex background has always been a hot and difficult problem in the field of machine vision and image processing.

Moving object detection method based on morphological features is practical merely for the specific target and background [2-3]. It only considers certain natural attributes of the target. In complex backgrounds, the algorithm’s robustness is insufficient. And when the target is micro compared to the background, the configuration of the target such as texture, color or shape is unconspicuous, making the effectiveness of that method limited. Another technique for moving object detection is based on the target and background statistical model. It has good performance in the simple background [4-6]. However, the statistical features of the target and the background are not stable in a dramatic changing environment, leading to a significant decline in the performance of the algorithm. The method of adaptive filtering combined with infrared technology and motion estimation has good real-time and recognition performance [7-9]. But it also has limitation that the threshold value and the motion feature of the target require a prior knowledge.

Therefore, it is still a difficulty to detect moving objects in complex background on computer vision, restricting the development of automation and information processing technology. But for the vertebrates, the visual behavior such as specific object detection, spatial location and motion tracking is an easy task. In the process of biological evolution, a variety of creatures developed a perfect tool for capturing the target, their vision systems, and provides an inspiration on the solution of the problem [10-11]. Frog has a relatively simple brain, and mainly relies on eye retinal neural for the target recognition. Besides, frog is sensitive to the dynamic targets of certain characters and is insensitive to static ones. At present, there is no unified model or fixed mechanism of frog vision on target identification. Most of the researches concentrate on the computer simulation of its specific features. According to frog eye sight is limited and its Predation depends on the target size, Zhi-ling Wang have proposed A Fuzzy Region Understanding Tactic for Object Tracking Based on Frog’s Vision Characteristic [11], through filtering, maintain, merge
method, the background interference was reduced and target texture, image contour features were highlighted gradually. Zhi-yong Li have presented a motion recognition method based on the bionic intelligence[12], moving target was determined through edge extraction and thermodynamic entropy threshold segmentation, and the algorithm had better effect. But none of the papers discussed the optic nerve mechanism of proposed method, and relatively narrowly focused the computer simulation of frog certain visual features.

This paper mainly focuses on the moving object detection in complex backgrounds. On the basis of summarizing the frog visual imaging features, the physiology characteristics of the frog retinal neural circuit and its response to dynamic target are induced. The impact factors of the frog visual neural circuit response to the target size, gray variance, shape characteristics and target motion feature are clarified. Then spatial-temporal recognition mechanism based on the frog’s vision is come up with and a biomimetic recognition algorithm of multi-feature fusion for dynamic target is designed. This method combines the morphological and kinematic features of dynamic objects effectively and achieves the recognition of the moving objects in complex background through effectively inhibiting background information and enhancing dynamic target features in multi-dimension. The algorithm principle is simple and easy to implement, providing a biomimetic approach for motion detection.

2 Frog eyes features and bionic ideas

2.1 Features of frog eyes and physiological structure

Frog eye has a unique visual characteristics and physiological structure[13,14], which ensures the effectiveness of its predatory behavior in complex natural environment. The field of view of frog eye is limited, and it can only handle grayscale information, while it is sensitive to the rate of change of the gray-scale shading. Frog eye only has the ability to identify the specific size of the target, and it’s more sensitive to the curved edge features of the target. Besides, the frog eye can only identify the moving target with the particular law of motion, but it fails to recognize static target. Frog eye is selective to the direction of motion of the dynamic target, which is more sensitive to the transverse crossing goal than the vertical crossing goal.

The unique visual characteristics of frog eyes rely on its unique physiological structure. The analysis of frog eye retinal nerve structure shows that the optic nerve fibers are divided into four major categories: darken detector, sustained contrast detector, convex edge detector and motion edge detector. The four types of nerve fibers achieve to recognize the target in grayscale images which has a specific size and a specific law of motion. R3, which is a basic physiological unit to compose the four kinds of nerve fibers, is one of the most important neurons of the frog retina, whose physiological characteristics determine the frog visual imaging characteristics and dynamic target recognition capabilities. Figure 1 is a simplified model of frog retinal ganglion cells R3[15]. Photoreceptors (Rec) convert light to membrane potential. Horizontal cells (H) reflect the environmental background brightness. Hyperpolarizing bipolar cells (HBC) and depolarizing bipolar cells (DBC) detect the instant changes of light on and light off, which means they can produce depolarization and hyperpolarization response to light stimulus. For motion stimulation, bipolar cells have a continuous output. ATH and ATD, two types of amacrine cells, are high-pass filter with a threshold, which respond to negative and positive changes of corresponding bipolar cells. The R3 cells can accept input of the ATD and ATH cells in all access roads within the receptive field. That means the increase in brightness stimulates R3 to response through ON channel, and the decrease in brightness stimulates R3 to response through OFF channel. Thus the R3 cells integrate the input of all channels in the space within the receptive field. The result shows that R3 cell is selective to size and speed of moving stimulation, because of its special structure and size. The receptive field size of the R3 cell determines the ability of neurons to detect stimulation. The difference of excitatory and inhibitory input synaptic delay and the time constant is an important basis to determine the moving type of the simulation.

Figure 1. Overall structure of simplified R3 model

2.2 Bionic ideas of moving target detection

The above analysis shows that the physiological structure of the frog eye’s R3 cell determines the spatial (the response characteristics to the stimulus of specific grayscale, edge and shape) and temporal (the response characteristics to the specific moving feature) characteristics of the moving target recognition. The moving target recognition process of frog eye is the result
of multi-feature fusion. Based on the frog visual properties and biological identification mechanism, this paper constructs the bionic structure and identification mechanism of moving target recognition, which is shown in figure 2. The four bionic recognizers respectively identify the gray, gray change rate, shape and moving feature, and the single feature recognition results are weighted to input to the multi-feature fusion recognition neural network. The multi-feature fusion recognition result is input to frog brain to compare with the experience threshold of prey and predator and determine the target property and decide the follow action.

![Figure 2. Moving target detection principle based on frog’s vision characteristics](image)

### 3 The Proposed Method

#### 3.1 Model expression of dynamic target

From the bionic moving target recognition structure shown in figure 2, we can see that frog brain acts as a guide for the identification mechanism. On the one hand, frog brain should adjust the neural network weights to improve the accuracy of target recognition when the environment changes. On the other hand, the threshold to determine predators and prey is the result of long term learning and can adjust its values depending on the scene and the target feature. Therefore, the target recognition mechanism of biological vision is an experienced judgment process. We build the following dynamic target model to simulate the dynamic target recognition process based on the comprehensive decision-making experimental threshold.

**Dynamic target:**

\[
\text{Object} = [\text{color}, \text{edge}, \text{shape}, \text{velocity}] \quad (1)
\]

In the expression, color stands for the color information of the object, and we choose the space chromaticity information using HIS model, namely \( \text{color} = \{H : 0 \sim 255\} \). Edge stands for the edge information of the object, and we will take the gray space of edge feature expression, namely \( \text{edge} = E_{m \times n} = (e_{ij})_{m \times n}, e_{ij} = \{0 \sim 1\} \). Shape is the shape information of the object, it obeys the particular shape feature that its expression is \( f(x) \), \( \text{shape} = S_{m \times n} = (s_{ij})_{m \times n}, s_{ij} = f(s_{ij}) \). Velocity stands for the speed information of the object, which obeys the law of a particular motion that expression is \( g(y) \), and is used for the recognition of movement style. which is \( \text{velocity} = \{V : 0 \sim 1\} \). When \( V \in g(y), V = 1 \), otherwise \( V = 0 \).

#### 3.2 Bionic algorithm implementation

##### 3.2.1 The bionic algorithm design

In order to achieve the bionic structure of moving target recognition and identification mechanism proposed in this paper, the bionic algorithm is to solve three major problems. (1) Rapid extraction method of dynamic spatial and temporal characteristics. (2) The mechanism of multi-feature fusion. (3) Dynamic threshold adjustment method. This paper uses chroma bandpass filter, edge detection based on Canny operator and shape detection based on Hough transform to realize the rapid extraction of the spatial feature of moving target. We use motion detection method based on optical flow to realize the rapid extraction of the temporal feature of moving target. Meanwhile, we also construct multi-input single-output three-layer neural network to realize the recognition mechanism of multi-feature fusion of moving target. In this paper, the objective determination threshold is set to a fixed value, and the dynamic adjust method of the threshold is to be studied in the future. The dynamic target recognition bionic algorithm is shown in figure 3.

![Figure 3. Dynamic target recognition bionic algorithm](image)

##### 3.2.2 Rapidly extraction method of moving target features

Dynamic sampling information model for target identification bionic platform is sequence images represented by RGB image model. If the image resolution is \( M \times N \), to simulate the frog eye monochromatic visual features, the RGB image is transferred into a HIS image. Based on the prior knowledge, the band pass filter algorithm of H (chroma) information is designed and the target recognition of monochrome vision based on frog eye is realized. Chromaticity band pass filter is shown as follows:
\[ h_{ij} = \begin{cases} 1, & 0 \leq H_{ij} \leq H_{max} \\ 0, & \frac{H_{ij}}{H_{max}} \leq H_{min} \end{cases} \]

(2)

\( h_{ij} \) is the filtering result of the point \((i,j)\) in image matrix space through chromaticity band pass filter. \( H_{ij} \) is the chromaticity H component of the point \((i,j)\) in image matrix space. \( H_{min} \) and \( H_{max} \) are the dynamic threshold ranges according to the prior knowledge, \( i = 0,1 \cdots M - 1, \quad j = 0,1 \cdots N - 1 \).

To simulate the recognition feature of the frog eye for the particular shape of a target, a circular target is adopted for feature extraction in this paper. The basic thought to detect circle based on Hough transforms is to transfer the image spatial domain to the parameter space and use certain parameters which will satisfy most of the boundary points to describe the curves in the image. By setting the accumulator to accumulate, the point corresponding to the peak is the information needed. The principle of Hough transforms detecting the image spatial resolution of the curve is as follows. The general form of analytical curve parametric representation is:

\[ f(x,a) = 0 \]

(3)

In the expression, \( x \) is a point on the curve (two-dimensional vector), and \( a \) is a point in the parameter space. For a circle, the radius is \( r \), and the circle whose center coordinates is \((a,b)\) can be represented in the parameter space as:

\[ (x_i - a)^2 + (y_i - b)^2 = r^2 \]

(4)

At this time, the point \( x = [x_i, y_i]^T \), \( a = [a, b, r]^T \) is three dimensional in parameter space. For a circle in the image space, its radius is fixed and every point on the circumference of a circle forms a set of cone which has an equal \( r \) and different \( a, b \). The points of the circle in image space which is an intersection of a bunch of cone mapping to the parameter space correspond to the center coordinates and the radius of the circle.

To simulate the recognition feature to the specific edge characteristics, this paper uses the edge detection method based on Canny operator, which has a high speed of extraction and is a mature method. The edge detection progress based on Canny operator is shown in Figure 4.

Figure 4. Edge detection based on Canny operator

Analysis of dynamic response of the frog visual characteristics and retinal nerve loop shows that moving target identification of the frog eye takes into account both the spatial characteristics (response characteristics to specific gray scale, shape, edge) and temporal characteristics (response characteristics to specific moving feature) of the moving target. The spatial-temporal identification mechanism makes full use of the morphological and kinematic characteristics of the target, which improves the effectiveness of target recognition.

Optical flow is the instantaneous speed of pixel movement which is projected onto the imaging plane by moving target. Optical flow field is the apparent change of grayscale mode. Since the optical flow field contains the motion information of the image sequence, it is feasible to use the motion information of different elements to divide moving targets in image sequences.

In this paper, we adopt the gradient-based LK algorithm proposed by Lucas and Kanade\(^{[16]}\). Let \( I(x,t) \) represents the gray value of the image at point \( x \) at the time \( t \), and \( v = (u, v) \) is the optical flow at this point. Obviously the grayscale values and the initial grayscale values of the image have the following relationship:

\[ I(x,t) = I(x - vt, 0) \]

(5)

The image consistency assumptions:

\[ \frac{dI(x,t)}{dt} = 0 \]

(6)

Coupled with the above Taylor expansion of formula (1), we get the following gradient constraint equation:

\[ \nabla I(x,t) \cdot v + \frac{dI(x,t)}{dt} = 0 \]

(7)
Thus, we get: \[ v = \frac{I_z}{I_x}, \] where the \( I_z \) and \( I_x \) are defined as follows:

\[ I_z = \frac{\partial I}{\partial x} \bigg|_{x(t)}, \quad I_x = \frac{\partial I}{\partial t} \bigg|_{x}. \]

Besides, by approximately calculating of the gradient constraint equation using the weighted least squares method, we can get the optimized optical flow field in each small interval \( \Omega \):

\[ \sum_{x \in \Omega} [W(x)]^2 \cdot [\nabla I(x,t) \cdot v + \frac{dI(x,t)}{dt}]^2 \]  \hspace{1cm} (8)

where \( W(x) \) is the window function.

The recognition process of the motion feature based on the optical flow shows below:

Assuming the feature point in image A is \( u = [u_x, u_y]^T \), motion estimation is to find the point:

\[ v = u + d = [u_x + d_x, u_y + d_y]^T \]  \hspace{1cm} (9)

In image B, where \( d = [d_x, d_y]^T \), which is the displacement vector of feature point \( u \). And \( v \), which is approximately calculated by solving the gradient constraint equation using the weighted least squares method, is the optical flow field value of feature point \( u \).

### 3.2.3 Strategy of multi-feature fusion

Because the external conditions and moving target changing have a bad influence, the moving target recognition method based on a single feature is limited. On the basis of the analysis of frog eye’s recognition mechanism, this paper has designed a multi-feature fusion mechanism based on neural network. The fusion strategy takes full advantage of the spatial and temporal characteristics, and enhances multi-dimensional information of moving target. The strategy of multi-feature fusion based on neural network is shown in figure 5.

Where, \( O_i(x_i, y_i), (i = 1, 2, 3, 4) \) is the recognition result of the four rapid recognition method, which is to identify the central coordinate of the target. \( O_R(x_i, y_i) \) is the real central coordinate of moving target, and \( O_{FBMF}(x_f, y_f) \) is the objective central coordinate of multi-feature fusion recognition. And we have:

\[ O_{FBMF} = \frac{\sum_{i=1}^{4} O_i \cdot W_i}{\sum_{i=1}^{4} W_i} \]  \hspace{1cm} (10)

Where \( W_i(i = 1, 2, 3, 4) \) is the nonlinear mapping matrix of the BP neural network input \( O_i, (i = 1, 2, 3, 4) \) and the output \( O_{FBMF} \), whose value is determined by BP neural network weights matrix \( W'_l(l = 1, 2, 3; i = 0, 1 \cdots k - 1; j = 0, 1 \cdots n - 1) \), and \( l \) is the layer ordinal number of the neural network, \( k \) is the number of neurons in each layer, \( n \) is the input number of a layer in the neural network. The BP model of 3 layers was used in the paper, and the function Sigmoid was chosen to be the activation function of neurons, which is:

\[ f(x) = \frac{1}{1 + e^{-x}} \]  \hspace{1cm} (11)

The structure of BP neural network based on multi-feature fusion is shown in figure 6.

The neural network was trained by inputting image sequences in which the central coordinate of the target had been calibrated in advance, so that the nonlinear mapping matrix \( W_i(i = 1, 2, 3, 4) \) was determined.
4 Moving object detection experimental results

In order to show the effectiveness of the algorithm, we conducted comparative experiments between a simple feature recognition and multi-feature fusion recognition in both simple and complex environment. In order to show the robustness of the algorithm, we conducted recognition experiments combining spatial and temporal characteristics together in strong similarity background interference and strong similarity target interference. The result is shown in Figure 10 to Figure 13.

Figure 7. Recognition comparison experiment under simple background

Figure 8. Recognition comparison experiment under complex background

Figure 9. Experiment under strongly similar target interference

Figure 10. Experiment under strongly similar background interference

The comparative experiments in Figure 7 and 8 shows that the recognition method based on the single morphological characteristics (for example only edge
detection and shape feature detection) of the target is ineffective. The detection method of frog bionic multi-features fusion (FBMF), which enhances the target feature and suppresses the background in multiple dimensions, is effective. Figure 9 shows the experiment that was conducted under strongly similar target interference. Targets with completely same characteristics of chrominance, shape and edge can’t be recognized effectively by spatial characteristics. But our algorithm combines the spatial and temporal characteristics of target so that the target feature has been enhanced. Thus, in the condition of strong similarity interference, the algorithm can detect moving target effectively. Experiment under strongly similar background interference in figure 10 shows that the wide range of background with the same color as the target cannot be effectively suppressed by filtering method. Our algorithm uses chrominance, edge, shape, and other multi-feature fusion method, which can effectively suppress the background information. Meanwhile, the moving feature recognition method is added to the algorithm, which made the detection more effective. The experimental results show that the algorithm based on bionic multi-feature fusion is robust and significant.

5 Conclusions

Based on frog's eyes visual characteristics, physiological structure and the dynamic responses of the retinal neural circuits, a spatial-temporal recognition mechanism based on the frog’s vision is provided, and a biomimetic recognition algorithm based on neural network with multi-feature fusion for dynamic target detection is designed. The algorithm effectively combines the kinematic characteristics with the morphological characteristics of the moving target, and recognizes the moving objects in complex background through effectively inhibiting background and enhancing dynamic target features in multi-dimension, which has stronger robustness and effectiveness. Besides, this method make integrated use of filtering, edge extraction, image transformation and other image processing methods, which is easy to implement. Meanwhile, one point should be proposed that even the real frogs would be tricked by baits, so the method of multi-features fusion proposed in this paper could continue to improve. But the moving target recognition method that simulates the biological visual processing mechanisms provides a new approach for this kind of problem under the complex natural condition.

6 References


