

# Research and Application of Artificial Intelligence in the Field of Vision System and Network

Li Kaiyue<sup>1,\*</sup>, Liu Lei<sup>2</sup>

<sup>1</sup>Science and Technology on Complex System Control and Intelligent Agent Cooperation Laboratory, Beijing, China

<sup>2</sup>Beijing Electro-Mechanical Engineering Institute, Beijing, China

\*Corresponding author Email: 782169985@qq.com

**Keywords:** Artificial intelligence technology; Three-dimensional vision system; Genetic neural algorithm; Simulation calculation

**Abstract:** Artificial intelligence technology is more widely used in human production and life, and is also frequently used in computer vision systems. The so-called computer vision technology is to make the machine have the sensory ability similar to human vision. Based on this, for the defects of 2D multimedia recognition is not clear enough, and the recognition time is long, the paper explores a set of artificial intelligence recognition methods with 3D multimedia vision, proposes the image features of 3D multimedia visual targets, and uses genetic neural network technology to make 3D. Artificial intelligence recognition technology for multimedia visual images is realized. Through simulation verification, the proposed 3D multimedia visual image technology can make the recognition image more accurate and the recognition accuracy more stable.

## 1. Introduction

Everything in the real world is three-dimensional, and people see things in two dimensions. Because human vision can quickly sense two-dimensional images from three-dimensional images based on previous experience, and thus see the three-dimensional world<sup>[1]</sup>. When the era of multimedia communication came, an artificial intelligence three-dimensional image recognition technology appeared. Image recognition technology was widely used in people's daily life, such as satellite positioning, biomedicine, machine vision and other fields<sup>[2]</sup>. When people use image recognition, there are problems such as insufficient image recognition accuracy, large recognition image error, long recognition time, and loss of edge information. Therefore, it is necessary to explore an effective method to improve the artificial intelligence recognition of 3D multimedia visual images, which can improve the speed of image recognition, reduce the error of recognition images, improve the accuracy of image recognition, and make the recognized images clearer. There is a certain practicality in life.

According to the problem embodied in the image recognition method described above, a three-dimensional multimedia visual image artificial intelligence recognition method based on neural network algorithm is proposed. This method can reduce the error of identifying images, reduce the length of image recognition, and improve image recognition. The accuracy rate makes the images recognized by multimedia intelligence clearer.

## 2. Artificial intelligence to identify the principle of 3D visual image

The paper combines genetic algorithm and artificial neural network, and uses the artificial neural network's nonlinear mapping ability to approximate the characteristics of nonlinear systems with arbitrary precision, and studies the complex nonlinear relationship between image feature changes and robot joint velocity. This method makes the system omits the inverse of the calculated image Jacobian matrix and the robot Jacobian matrix, which greatly reduces the calculation amount of the system and avoids the error caused by the parameter setting on the system. The principle of

artificial intelligence recognition of 3D multimedia visual image is to test the collected visual image within the contour range of the target image, and denoise the image within the new contour range to obtain the image within the final contour range. A three-dimensional target image recognition key point database, the key identity threshold of the visual three-dimensional image, to determine whether the smart recognition is successful, and finally realize the artificial intelligence recognition of the three-dimensional multimedia visual image<sup>[3]</sup>.

## 2.1 Image Jacques matrix

The relationship between image motion trajectory tracking and image features can generally be described by an image Jacobian matrix. The definition is as follows:

$$\dot{f}=J_i(\gamma)\bullet\dot{\gamma}(1)$$

Where  $f$  is the image feature parameter vector,  $\gamma$  is the coordinate parameter of the robot in the task space, and  $J_i(\gamma)$  is the image Jacobian matrix, which can be expressed as:

$$J_i(\gamma)=\left(\frac{\partial f}{\partial \gamma}\right)=\begin{pmatrix} \left(\frac{\partial f_1(\gamma)}{\partial \gamma_1}\right) & \dots & \left(\frac{\partial f_1(\gamma)}{\partial \gamma_n}\right) \\ \dots & \dots & \dots \\ \left(\frac{\partial f_m(\gamma)}{\partial \gamma_1}\right) & \dots & \left(\frac{\partial f_m(\gamma)}{\partial \gamma_n}\right) \end{pmatrix} \quad (2)$$

In the robot system, the joint motion velocity vector  $\dot{q}$  of the robot at the time of input is differentially related to the change rate of the end effector posture as follows:

$$\dot{\gamma}=J_\gamma\bullet\dot{q}(3)$$

Through the change of formula (3), the relationship between the joint motion of the robot and the image features can be obtained as follows. Let  $J=J_iJ_\gamma$ , get  $\dot{f}=J\bullet\dot{q}$ ,  $J$  is the total image Jacobian matrix of the system.

$$\dot{f}=J_i\bullet\dot{\gamma}=J_iJ_\gamma\bullet\dot{q}(4)$$

## 2.2 Genetic Neural Network

In this paper, the network mapping structure is determined by combining genetic algorithm and artificial neural network, and then the genetic algorithm is used to optimize the approximation to determine the weight and threshold of the network connection. The artificial neural network algorithm is easy to fall into the local extremum of the error function. For larger search spaces and non-differentiable functions, the global extremum points cannot be effectively searched. The genetic algorithm is a global optimization search algorithm. If the two are combined, the local minimum point can be effectively avoided, and the complex nonlinear relationship between the robot joint angular motion vector and the image feature vector is better approximated<sup>[4]</sup>. The genetic neural network algorithm used in this paper is mainly divided into three parts: 1) the determination of BP neural network structure; 2) the addition of genetic algorithm for optimization; 3) the training and prediction of genetic neural network. The weight and threshold information contained in the found optimal individual is assigned to the neural network, and the network is trained to predict the function output. The algorithm flow is shown in Figure 1.

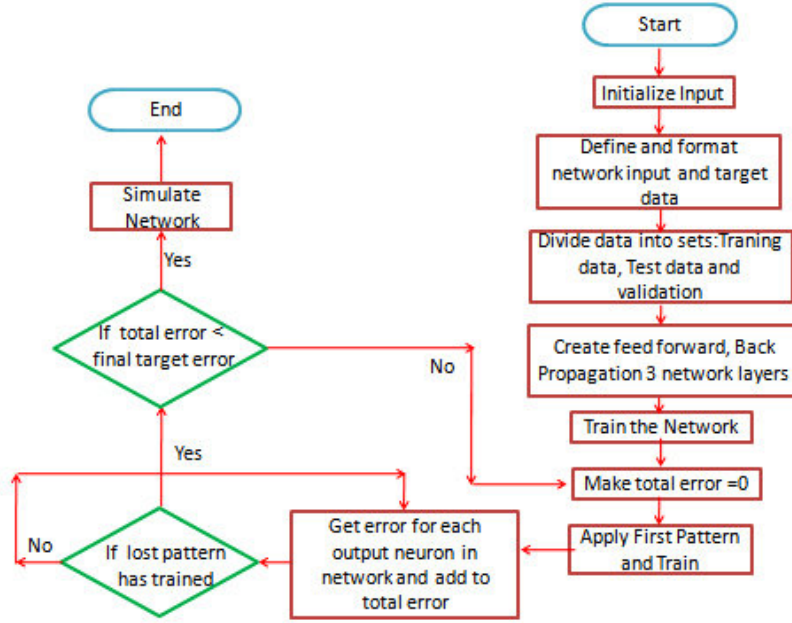


Fig.1 Neural network algorithm steps

### 2.3 3D visual image artificial intelligence recognition

In the 3D visual image recognition, it is necessary to extract the 3D visual image feature parameter vector, and normalize and reduce the dimensionality processing, and establish a 3D visual image recognition model to classify and identify the 3D visual image. The specific implementation process is as follows:

#### 1) Inspection within the target image contour range

In the process of multimedia inductive visual 3D image artificial intelligence recognition, the collected visual images are verified by visual 3D multimedia image recognition. Let  $i$  denote the shape base of the visual image, extract the test structure in the  $i$  shape base, set  $\{T_s\}$  to represent the data set in the test structure, and  $s=1,2,\dots,i$ ; after processing the structural expansion phenomenon in the shape base, the multi-dimensionality test structure can be obtained. Indicates  $wT_s$ ; sets the initial image  $a(x,d)$  sensed by the multimedia visual image;  $F_{wi}(x,d)$  denotes the contour structure parameters obtained by describing the different test structures, then the multi-scale contour gradient parameter calculation formula is as follows.

$$F_w(x,d) = \sum_s^{\varphi} p_s F_{wi}(x,d) \quad (5)$$

$$F_w(x,d) = \sum_s^{\varphi} p_s F_{wi}(x,d) \quad (6)$$

When the  $p_s$  values are equal, after the gradation processing, the contour area inspection image  $G_w(x,d)$  can be obtained; the value range of  $w$  is  $[i,n]$ ,  $\phi_w$  represents the weight coefficient in the different inspection structure, and the new multi-scale method can be used to obtain new The image formula within the contour range is as follows.

$$h(x,d) = \sum_{w=i}^l \phi_w F_w(x,d) \quad (7)$$

The obtained image within the new contour range is denoised to obtain the image  $F(x,d)$  in the final contour range, and the visual 3D image artificial intelligence in the multimedia visual sensing is

used to identify the most accurate data.

## 2) Key point extraction of 3D visual images

By constructing a coordinate system and constructing a component of a key point of a 3D visual image, in the process of identification, a 3D target image recognition key point database is also needed, in which three-dimensional key data of the storage target is set, and the three-dimensional key vector is compared in the database. And the target key vector to complete the 3D visual image recognition, but the number of visual 3D image recognition samples in the database is small, the extracted position has errors, and the calculation formula for extracting the same degree of visual 3D image recognition is as follows.

$$H = \sqrt{(j_1 - k_1)^2 t_1 + (j_2 - k_2)^2 t_2 + \dots + (j_n - k_n)^2 t_n} \quad (8)$$

Where:  $n$  represents the dimension of the key vector of the visual 3D image;  $J_i$  represents the component of the  $i$  shape base of the key vector of the visual 3D image;  $k_i$  represents the component of the  $i$  shape base of the element in the multimedia database;  $t$  represents the  $i$  component weight coefficient in the shape base. Let  $\bar{w}$  denote the key similarity threshold of the visual 3D image. When  $H < \bar{w}$ , observe the sample in the database and the unidentified sample. If the degree of similarity is high, the recognition is successful; the same degree is low, indicating that the recognition fails, and then with the multimedia database. An image sample is compared [5]. In summary, by examining the contour range of the target image and extracting the accurate image of the visual three-dimensional image, the principle of artificial intelligence recognition of the three-dimensional multimedia visual image is completed, and the three-dimensional multimedia visual image recognition of the human body is realized.

## 3. Human body three-dimensional image feature value intelligent recognition and extraction

Firstly, a human body interaction reconstruction model is established to extract the characteristics of human motion. In the process of human body movement, each action posture has its characteristic parameters. When different motion recognition systems make the same technical action, there will be differences. In this paper, the visual characteristics acquisition method is used to distinguish the human motion.

Based on the above-mentioned visual reconstruction of human motion. According to the above assumption, the width of the human motion image is  $W$ , the height is  $H$ , and the background image  $B$  of the motion scene and the current human motion visual image  $I$  are divided into sub-blocks having the following expressions representing the edge contour of the single-frame human motion. Block pheromone.

$$\begin{aligned} \min F(x) &= (f_1(x), f_2(x), \dots, f_m(x))^T \\ \text{s.t.} \quad g_i &\leq 0 \quad i=1, 2, \dots, q \\ h_j &= 0 \quad j=1, 2, \dots, p \end{aligned} \quad (9)$$

Among them, the moment of inertia of the  $g_i$  human body around the vertical axis (Y axis),  $h_j$  is the moment of inertia in the horizontal direction,  $q$  is the length of the mesh formed by the motion track of the human body,  $p$  is the width, and  $f_m(x)$  is the pixel intensity of the edge corner point.

The block pheromone of the edge contour of the single-frame human motion obtained above is clustered, the feature information of the human visual image is acquired, and the feature information is filtered, and the visual feature of the human motion is segmented to obtain the human body motion. Feature analysis of 3D viewpoint switching equations of motion [6].

$$\begin{aligned} \dot{x} &= V \cos \theta \cos \psi v, & \dot{y} &= V \sin \theta \\ \dot{z} &= -V \cos \theta \sin \psi v, & \dot{v} &= \omega_y \sin \gamma + \omega_z \cos \gamma \end{aligned} \quad (10)$$

Among them,  $x, y, z$  is the centroid position of the visual image of the human body motion;  $\psi v$  is the switching angle of the human motion action area, and the analysis of the movement visual characteristics of the human body movement is realized, and the 3D vision is performed for the next step. The comparison provides a data base.

## 4. Genetic neural network visual case simulation

### 4.1 Neural Network Vision Control Design

In this paper, a model of a six-joint robot is built in the MATLAB software using the Robotics toolbox. The Jacobian matrix is used to calculate the relationship between the joint velocity and the image feature point velocity. Then it was simulated and the experimental data between 1 681 joint motion vector and image feature vector was recorded for the training and prediction of genetic neural network. The initial parameters of the neural network are set as: network training times Net. TrainParam. Epochs=100; network learning rate Net<sup>[7]</sup>.TrainParam. Lr=0.1; network performance function error Net. TrainParam. Goal=0.00002; genetic algorithm initial parameter setting: evolution parameter Maxent =10; population size pop = 30; crossover probability Picross = 0.3; mutation probability Mutation = 0.1 and rookie method to select new individuals.

### 4.2 Modeling and Simulation

In the visual servo control system, a trained genetic neural network is employed as the controller of the designed system. The genetic neural network is used to relate the relationship between the school robot joint motion vector and the image feature change vector, so that it can conveniently calculate the joint angle motion vector at any time to achieve the target from any initial position to the desired position and attitude.

In this paper, using the Robotic toolbox in MATLAB, a visual neural network-based visual servo control system is built and simulated in the Simulink environment. The genetic neural network trained above replaces the calculated image Jacobian matrix and Jacobian matrix in the system. In the simulation, the target is a square, the coordinates of the four vertices in world coordinates shaped as:  $[A, B, C, D] = [-1.25, -1.25, -0.75, -0.75; -1.25, -0.75, -0.75, -1.25; 5,5,5,5]$ ; 4 target vertex coordinates in the camera image coordinate system is  $[244.94, 344.69, 351.76, 252.02; 743.68, 750.76, 651.01, 643.93]$ ; desired image feature points position  $[412,412,612,612; 412,612,612,412]$ ; for the specific parameters of the camera; focal length  $f = 0.008m$ , target imaging depth  $h = 0.5m$ . The image is obtained as follows.

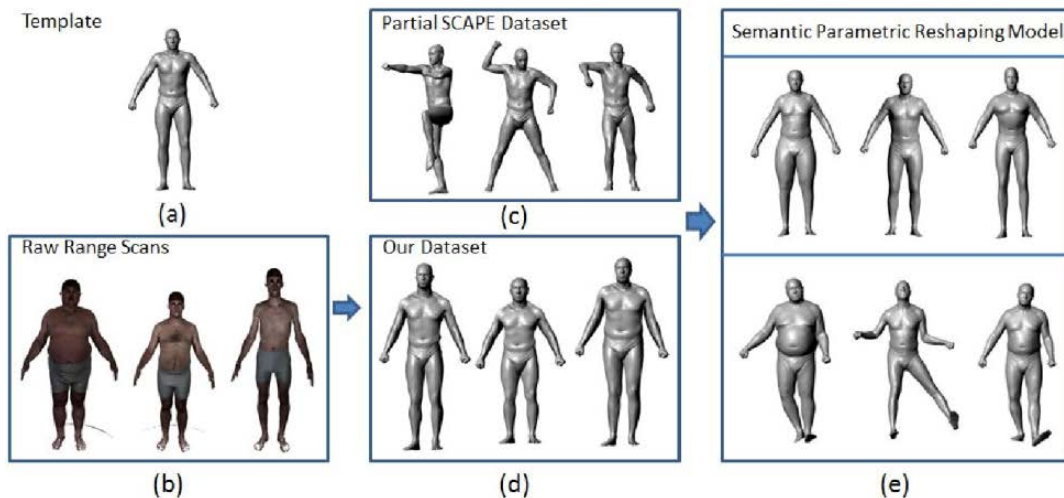


Fig.2 Simulation model of structural characteristics of human motion

## 5. Conclusion

With the rapid development of multimedia intelligent recognition technology, the emergence of 3D visual image artificial intelligence recognition technology is convenient for people's daily life, but due to the low precision of intelligent recognition technology, a 3D multimedia visual image based on genetic neural network is proposed. The artificial energy recognition method shows that the recognition accuracy is improved, the error of the recognition image is reduced, and the recognition performance of the artificial intelligence recognition technology of the three-dimensional multimedia visual image is enhanced, which has certain practical value.

## References

- [1] Qi Jingxiang. Human motion tracking and 3D pose recovery based on visual information features and machine learning. Xidian University, Vol.2 (2012) No.12, p.77-79.
- [2] Huang Luwei. Design and research of intelligent monitoring system for human body motion based on machine vision. Beijing Jiaotong University, Vol.15(2008) No.23, p.33-37.
- [3] Yang Dongfang, Sun Fuchun, Wang Shicheng. Rolling Time Domain Estimation Algorithm for Self-motion of Monocular Vision System. Pattern Recognition & Artificial Intelligence, Vol.1(2013) No.26, p.76-82.
- [4] Wang Yu, Wang Yongtian, Liu Yue. A Moving Object Detection Algorithm Based on Omnidirectional Vision. Pattern Recognition & Artificial Intelligence, Vol.4(2008) No.21, p.488-493.
- [5] Research on real-time tracking algorithm of moving targets based on machine vision. North University of China, Vol.1(2015) No.17, p.15-19.
- [6] Zhao Yanli. Research on Several Technologies of Computer Vision 3D Reconstruction. Nanjing University of Science and Technology, Vol.4(2007) No.14, p.33-39.
- [7] Wan Yitai. BP artificial intelligence neural network topology and algorithm. Heilongjiang Science and Technology Information, Vol.3(2009) No.28, p.84-85.