

Research Article

Yu Shao* and Deden Witarsyah

Fast recognition method of moving video images based on BP neural networks

<https://doi.org/10.1515/phys-2018-0123>

Received Oct 07, 2018; accepted Nov 14, 2018

Abstract: At present, the accuracy of real-time moving video image recognition methods are poor. Also energy consumption is high and fault tolerance is not ideal. Consequently this paper proposes a method of moving video image recognition based on BP neural networks. The moving video image is divided into two parts: the key content and the background by binary gray image. By collecting training cubes. The D-SFA algorithm is used to extract moving video image features and to construct feature representation. The image features are extracted by collecting training cubes. The BP neural network is constructed to get the error function. The error signal is returned continuously along the original path. By modifying the weights of neurons in each layer, the weights propagate to the input layer step by step, and then propagates forward. The two processes are repeated to minimize the error signal. The result of image feature extraction is regarded as the input of BP neural network, and the result of moving video image recognition is output. And fault tolerance in real-time is better than the current method. Also the recognition energy consumption is low, and our method is more practical.

Keywords: BP neural network, moving video image, recognition, binaryzation

PACS: 07.05.Mh, 07.05.Kf, 07.05.Pj

1 Introduction

Previously, moving video image recognition technology was a very strange phrase, but now it has become more and more apart of people's lives. Digital moving video im-

age is the representation of two-dimensional image with finite digital pixels. Its recognition and detection is of great significance to the breakthrough in this field [1, 2]. Image recognition technology is becoming more and more mature. Every year, new technologies and achievements occur at a great pace. In the 21st century, one of the hottest technologies is artificial intelligence. However, image recognition technology is the core of artificial intelligence. It is the eye of future intelligent AI. Its application will inevitably lead to the rapid development of artificial intelligence [3, 4]. To sum up, the research of moving video image recognition methods and technology is ongoing.

In order to automatically recognize and track the moving droplets in the welding video images, Zhang *et al.* proposed the image recognition method based on frame difference. Mean-shift algorithm, aiming at the characteristics of gray images and single backgrounds. In order to solve the problem that the Mean-shift algorithm needs to fetch the target manually in the starting frame, the frame difference method was used to process the first two frames of the video image to get the target window and the center position in order to calibrate them. Combining the Mean-shift algorithm, based on gray histograms, the target template position of the next frame was determined so as to realize the automatic recognition and tracking of moving droplets. The results showed that the method had good real-time performance, but the recognition accuracy was low [5]. Static-video face recognition proposed by Fan Zheyi *et al.* was an identity recognition technology where the training set was a high-quality static image and the test set was a low-quality video sequence. Aiming at the difficulty of image alignment and motion blur, an improved sparsely represented static-video face recognition algorithm was proposed. According to the gradient variance information, the geometric features of facial images under video conditions were achieved. The problem of motion blur was solved by constructing a dictionary by multi-scale filtering of images. The key frames in video sequences were extracted by cross-correlating coefficients between images. Experimental results showed that the algorithm ran stably, but the real-time recognition performance was poor [6]. Xu H.N. *et al.* proposed a motion recognition method based on three-dimensional depth im-

*Corresponding Author: Yu Shao: School of Electronic and Information Engineering, SIAS International University, Xinzheng, 451150, China; Email: michelle_shao@163.com

Deden Witarsyah: School of Industrial Engineering, Telkom University, 40257, Bandung, West Java, Indonesia; Email: dedenw@telkomuniversity.ac.id

age sequence to solve the problem of the high cost of traditional motion recognition algorithms in color video and inadequate two-dimensional information. The algorithm proposed a time depth model (TDM) to describe actions in time dimension. In the three orthogonal Cartesian planes, the depth image sequence was divided into several sub-actions. Also the inter-frame difference and energy accumulation were made for all the sub-actions to form a depth motion map to describe the dynamic characteristics of the action. In the spatial dimension, the spatial pyramid directional gradient histogram (SPHOG) was used to encode the time depth model to get the final descriptor. Finally, support vector machine (SVM) was used to classify actions. Experiments on two authoritative databases of MSR Action 3D and MSR Gesture 3D showed that the method had high recognition accuracy, but the recognition energy consumption was not ideal [7]. Liu Mingzhu *et al.* proposed an image recognition algorithm based on deep learning. A gabor filter was used to extract textural features of video images in four directions: horizontal, vertical, skimming and scratching. Then the depth confidence network was constructed by a RBM incremental depth learning algorithm layer by layer to locate the text region in the extracted texture feature image. In this paper, the feasibility of using morphological processing methods and an OCR character library were also studied to realize text recognition of video image, and the recognition effect were analyzed. The test results showed that the algorithm had good real-time performance [8].

Given the problems existing in the current research results, a fast recognition method of moving video image based on BP neural network is proposed. The detailed process is as follows:

1. A binary method is used to process the moving video image to improve the accuracy of image recognition, enhance the real-time image recognition, and to a certain extent reduce the energy consumption of recognition.
2. The D-SFA algorithm is used to extract the features of moving video images, lay a foundation for further improving the accuracy of image recognition.
3. The result of image feature extraction is input into the BP neural network algorithm [in order] to recognize the moving video image.

The effectiveness of the proposed method is verified experimentally.

2 Method

2.1 Moving video image processing

Image threshold segmentation is a widely used imaging technology. It takes advantage of the differences in gray characteristics between the object to be extracted from the image and its background. It also regards the image as a combination of two types of regions (targets and background) with different gray levels [9]. Among them, the most important is the selection of image threshold, inappropriate threshold selection will affect the quality of the binary image and recognition accuracy. This is because of the influence of uneven illumination, camera distortion, insufficient exposure and narrow dynamic range, results in serious artifacts appearing in the moving video images. Because of the uneven gray distribution and insufficient contrast, the edge of the moving video image is blurred and the details are not clearly distinguished. Also the binarization effect of moving video image is seriously affected.

For this reason, a global threshold algorithm based on the spatial distribution of moving video images and the classification criterion of maximum inter-class variance is used to binarize the recognition of moving video images, which can not only eliminate artifacts, but also maintain the edge integrity of moving video images [10].

Given ideal condition of uniform illumination, no noise and interference, the total gray level of the moving video image changes gently. Supposing the key content of the image is g_1 , the background gray is g_2 , and $0 \leq g_1, g_2 \leq 255$. Supposing that the proportion of key content pixels in a moving video image is r_1 , the proportion of background pixels is r_2 , and $0 < r_1, r_2 < 1, r_1 + r_2 = 1$. The gray mean of moving video images is expressed as Eq. (1):

$$M = r_1 g_1 + r_2 g_2 \quad (1)$$

The variance calculation is shown in Eq. (2):

$$C^2 = r_1 (g_1 - M)^2 + r_2 (g_2 - M)^2 \quad (2)$$

According to Eq. (1) and Eq. (2), it can get:

$$r_1 (g_1 - M) + r_2 (g_2 - M) = 0 \quad (3)$$

According to Eq. (3), there are:

$$(g_2 - M) = -\frac{r_1}{r_2} (g_1 - M) \quad (4)$$

The Eq. (4) is substituted for Eq. (2) and it can get:

$$C^2 = \frac{r_1}{r_2} (g_1 - M)^2 \quad (5)$$

In summary, the grayscale of the key content in the image is:

$$g_1 = M \pm \sqrt{\frac{r_2}{r_1}} C \quad (6)$$

In this way, the grayscale of image's background is:

$$g_2 = M \pm \sqrt{\frac{r_1}{r_2}} C \quad (7)$$

The rough threshold value can be expressed as:

$$T = M - \sqrt{\frac{r_1}{r_2}} C \quad (8)$$

According to the calculation of rough threshold, the fine threshold value of an image with binaryzation is determined. The binarization of moving video images can be reduced to the classification of the two models (targets and backgrounds). Finally, the images are divided into two categories: key content and background [11].

Assuming that a given moving video image has a gray level of $123 \cdots L'$, a total of L' , and a threshold of t , the pixels with gray levels greater than t and less than t are divided into two categories: class 1 and class 2. The total number of pixels in class 1 is $\omega_1(t)$, the average gray value is $\mu_1(t)$, and the variance is $\sigma_1(t)$. The total number of pixels in class 2 is $\omega_2(t)$, the average gray value is $\mu_2(t)$, the variance is $\sigma_2(t)$, and the average gray value of image pixels is $\mu(t)$. The inter-class variance $\sigma_A^2(t)$ and intra-class variance $\sigma_B^2(t)$ can be defined as:

$$\sigma_B^2(t) = \omega_1(t) [\mu_1(t) - \mu(t)]^2 + \omega_2(t) [\mu_2(t) - \mu(t)]^2 \quad (9)$$

where

$$\mu(t) = \omega_1(t) \mu_1(t) + \omega_2(t) \mu_2(t) \quad (10)$$

In pattern classification theory, there are three criteria for separability measurement among different classes: scattering matrix, divergence and Battacharyya distance. The ratio of inter-class variance to intra-class variance corresponds to the scattering matrix, which reflects the distribution of patterns in pattern space. Also the greater the similarity of the pixels of each class are the classification results will be better [12, 13]. Therefore, the maximum inter-class variance criterion function $S(t)$ is used to fine tune the rough threshold.

$$S(t) = \frac{\sigma_B^2(t)}{\sigma_A^2(t)} \cdot T \quad (11)$$

According to Eq. (11), the binarization method based on spatial distribution is combined with the maximum inter-class variance classification criterion to realize the binarization of moving video images. In this way, the contrast between the background and the target is enhanced,

the accuracy of the image recognition is improved, the real-time performance of the image recognition is enhanced, and the energy consumption of the recognition is reduced to a certain extent.

2.2 Feature extraction of moving video image

Based on the results of moving video image processing, the D-SFA algorithm is used to extract the features in the image. In this paper, image feature extraction is divided into three parts: collecting training cubes; extracting features of moving video image by using the algorithm; and constructing feature representations.

Collecting training cube is a method of constructing original input signal $x(t')$ from video sequences. Firstly, the original video is processed and the frame difference image sequence is obtained. A selected frame is used as the initial frame to detect the feature points, and then the optical flow method is used to track the feature points to get the corresponding set of trajectories of all the feature points in the video. For each trajectory in the trajectory set, the pixel values in the neighborhood of each trajectory point $w \times w$ are extracted to form a series of pixel blocks. Considering the time information, the sequence of pixel blocks of each point is integrated by $\Delta t'$ successive frames, and $\Delta t' = 3$ is taken here. After further integrating all the feature points, the training cube is obtained, that is, the input vector $x(t')$ is constructed. Figure 1 shows the process of training the cube.

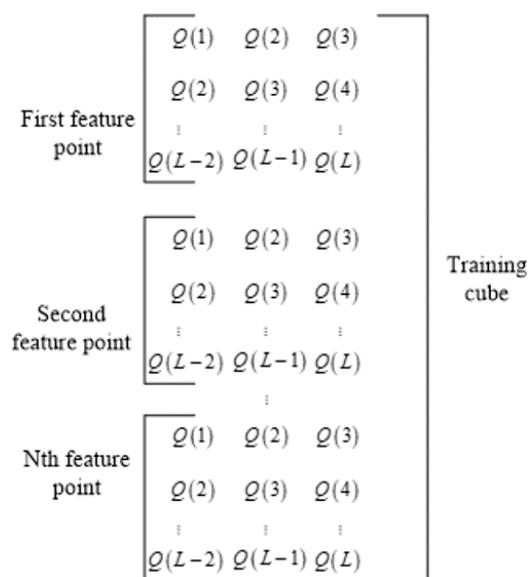


Figure 1: Process of training cube

According to the cube training results, the D-SFA algorithm is used to extract the features of the moving video images. The algorithm is a kind of unsupervised learning algorithm. The idea of extracting image features from video is that the training cubes collected from different kinds of behavioral video are mixed together for machine learning of feature functions, and then the features are extracted from the trained feature functions. Because the supervised information cannot be encoded, the extracted features do not have good discrimination between behaviors. The algorithm introduces supervised information in the learning process. The idea of extracting image features from human behavior video is that the training cubes collected for each type of behavior are used for learning feature functions respectively, so the learning feature functions have the ability to distinguish the inter-class behavior, that is, they are selective to intra-class behavior.

Since feature analysis can minimize the mean square derivative, the fitting degree of a cube to the corresponding feature function can be measured by transforming the cube's square derivative [14]. If the value is small, this cube and the feature function are well fitted. For the C'_i th and j th feature functions of the i th cube, the square derivative is defined.

$$v_{i,j} = \frac{1}{L - \Delta t'} \sum_{t'=1}^{L-\Delta t'} [C'_i(t' + 1) \otimes C'_i(t')]^2 \quad (12)$$

Where, L represents the number of tracked frames, and \otimes represents the transformation operation. Here, L is defined as 15 and \otimes is defined as 3.

Based on the calculation of Eq. (12), the square derivatives are accumulated on all cubes to form the feature of moving video image:

$$f = S(t) \sum_i^N V_i \quad (13)$$

Where, N represents the number of cubes collected in a moving video, $V_i = (v_{i,1}, v_{i,2}, \dots, v_{i,K})$, and K represents the number of feature functions of a moving video image. Through image feature extraction, the accuracy of moving video image recognition is further improved.

2.3 Fast recognition of moving video images based on BP neural network

BP neural network is a multi-layer forward network with at least one layer at each level composed of input layers, output layers, and hidden layers shown in Figure 2.

The main idea of a back propagation algorithm in BP neural network is to divide the learning process into two

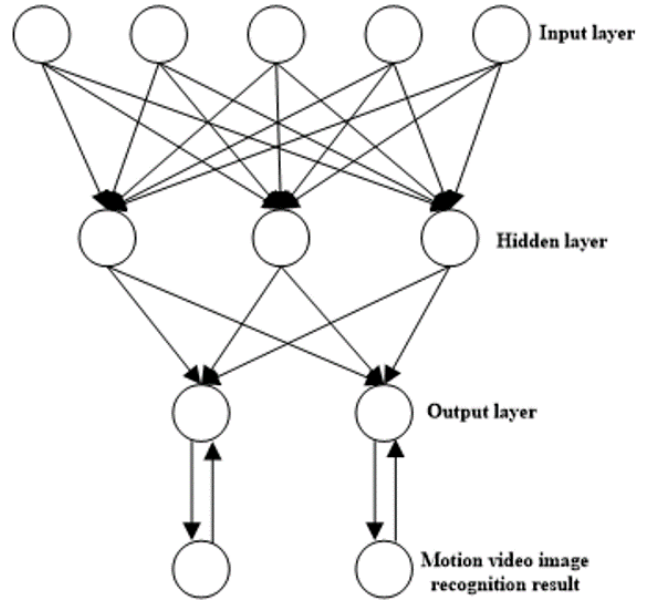
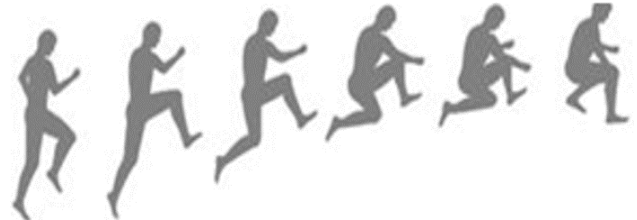


Figure 2: Three-layer BP neural network model



(a) Experimental samples 1



(b) Experimental samples 2

Figure 3: Part of experimental samples

stages: In the first stage (forward propagation process), the input information calculate the actual output value of each unit layer from the input layer, each layer of neuron state only affects the state of the next layer of neurons; In the second stage (back propagation process), assuming that the desired output value is not obtained at the output layer, the difference between the actual output and the desired output is calculated recursively layer by layer, and the error signal tends to be minimized by modifying the weight of the front layer according to the error. It gradually approximates the target by continuously calculating the network weights and deviation changes in the direc-

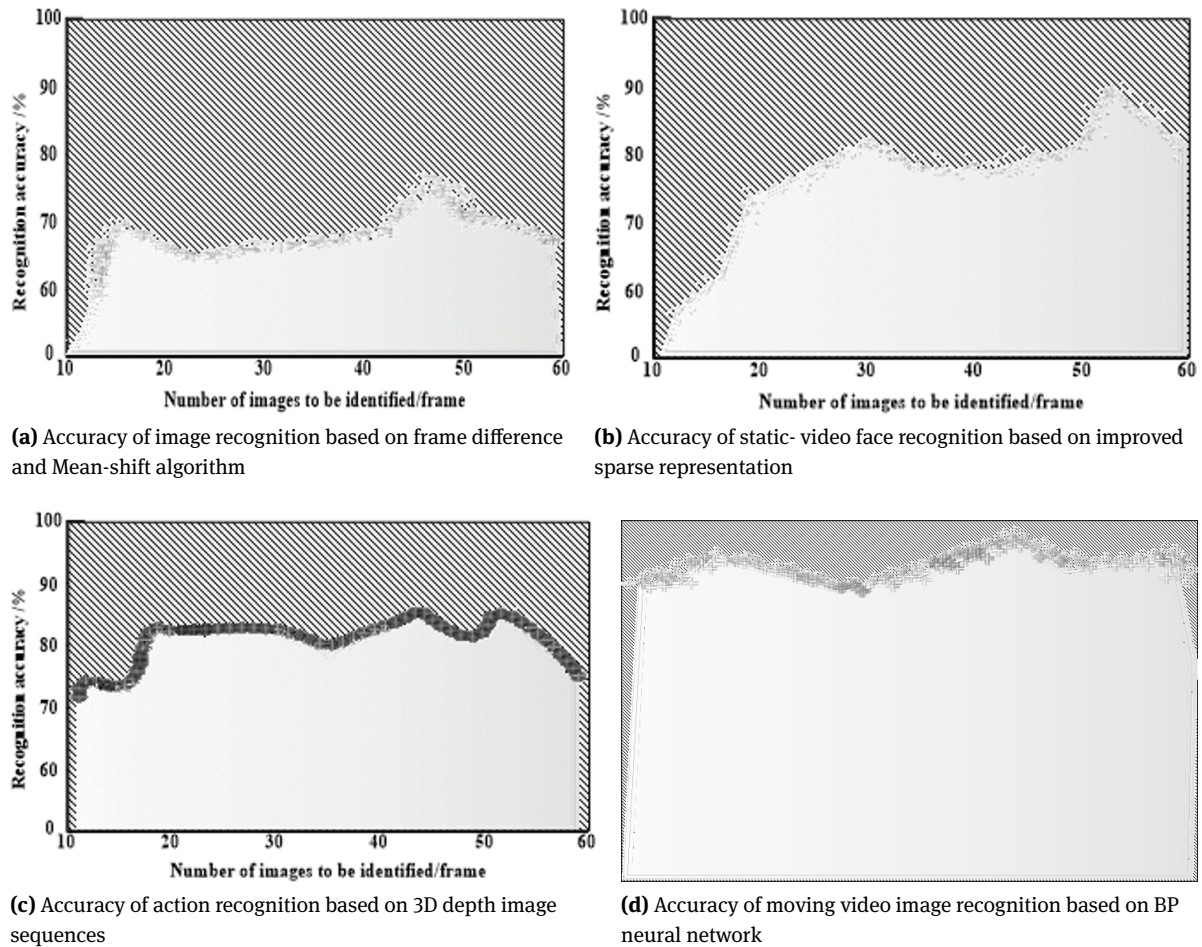


Figure 4: Comparison of accuracy of different image recognition methods

tion relative to the descent of error function slope. Every change of weight and error is directly proportional to the effect of network error [15–21].

Based on the above analysis, it is assumed that the number of cells in the input layer, the middle layer and the output layer is K , H and G respectively. f is added to the input vector in the BP neural network, H is the intermediate output vector, and G is the actual output vector of the network, that is, the recognition results of moving video images are recognized.

Assuming that the weight of the output unit i' to the hidden unit j' is $W_{ij'}$, the weight of the hidden unit j' to the output unit l is W_{jl} . θ_l and $\phi_{j'}$ represent the thresholds of the output unit and the hidden unit. Controlling θ_l in the range of $[1.3, 1.4]$ and $\phi_{j'}$ in the range of $[0.5, 0.6]$ can effectively improve the fault-tolerance of moving image recognition.

A transfer function is a function that reflects the intensity of the stimulus pulse from the lower input to the up-

per node. It is also called a stimulus function. Generally, it is the Sigmoid function that is continuously selected in $(0, 1)$. That is,

$$F(x) = \frac{1}{1 + e^{-x}} \quad (14)$$

The error function $E(x)$ is:

$$E(x) = \frac{1}{2} \sum_{j'=1}^G (G - H)^2 \quad (15)$$

The output $h_{j'}$ of each unit in the middle layer and output layer can be expressed as:

$$H_{j'} = F \left(\sum_{i'=0}^K W_{ij'} + \phi_{j'} \right) \quad (16)$$

$$G_l = F \left(\sum_{j'=0}^H W_{jl} + \theta_l \right) \quad (17)$$

In the BP neural network algorithm, the gradient descent method is used to adjust weights.

$$W_{ij'}(n' + 1) = W_{ij'}(n') + \eta \xi_{j'} x_i'' \quad (18)$$

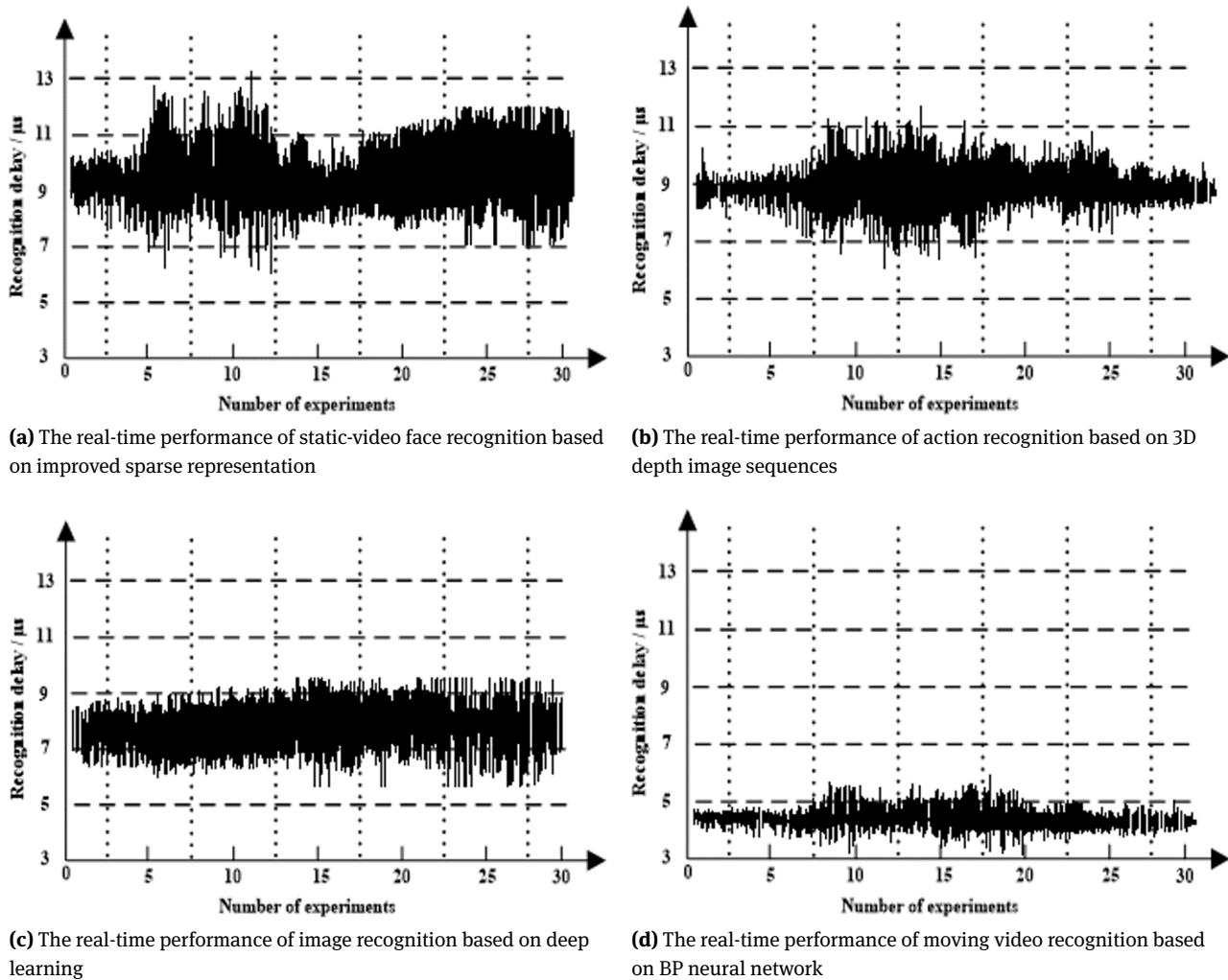


Figure 5: Real-time performance comparison of different image recognition methods

Where, $x_{i'}''$ represents the output or external input of point i' , η represents learning rate, and $\xi_{i'}$ represents error.

Equation (18) is used to modify the weights and thresholds, and the error signals are returned continuously along with the original. By modifying the weights of neurons in each layer, the error signals propagate to the input layer one by one. Through the forward propagation process, the two processes are repeated, making the error signal smallest. When all the errors meet the requirements, the moving video image feature f is input into BP neural network, and the recognition result of the moving video image is obtained.

$$G = \frac{f \odot H \cdot W_{i'j'} (n' + 1)}{E(x)} \cdot F(x) \quad (19)$$

The result of Equation (19) is the result of moving video recognition based on BP neural network.

3 Results

In order to validate the method of motion video image recognition based on BP neural network, the experiment of motion video image recognition is carried out using behavior recognition database. The database contains 50 kinds of single behaviors, including bending, running, single-foot jumping, double-foot jumping, in-situ jumping, waving, sidetracking, walking, single-arm waving and double-arm waving. Each behavior is completed by 9 different individuals. Figure 3 is part of the experimental samples. The experimental platform is built on Matlab.

- Accuracy of image recognition
- Real-time performance of image recognition
- Energy consumption of image recognition
- Fault-tolerance of image recognition

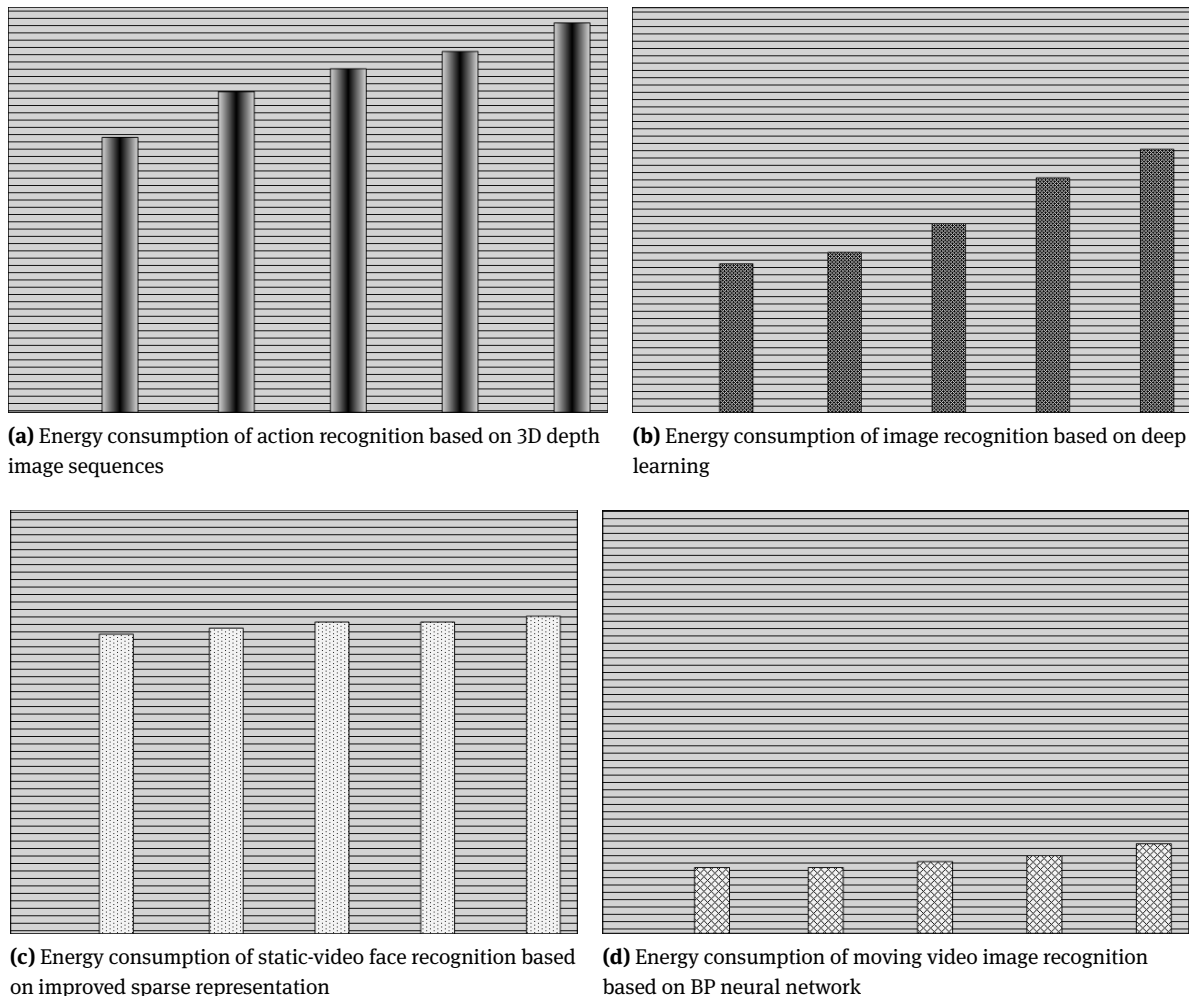


Figure 6: Comparison of different image recognition methods for identifying energy consumption

The results are as follows:

Figure 4 shows that the combination of the frame difference method and a Mean-shift algorithm has the worst recognition accuracy. Other current methods do not have strong recognition accuracy, the reliability is poor. The recognition method of moving video image based on BP neural network does not change with the number of moving video images to be recognized, and the highest recognition accuracy is 99%.

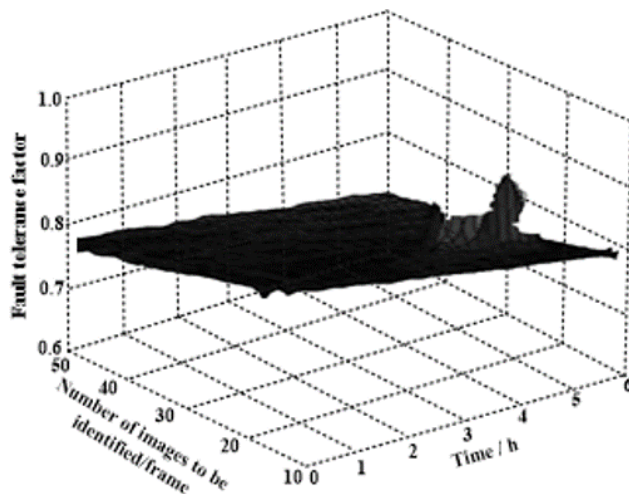
In the real time experiment of image recognition, the number of images to be recognized is defined as 100.

In Figure 5, different methods show different real-time performance in a certain number of images to be recognized. Current image recognition algorithms and methods do not have stability and persistence in real-time, and the recognition process has a high delay. Moving video image recognition method based on BP neural network can eff-

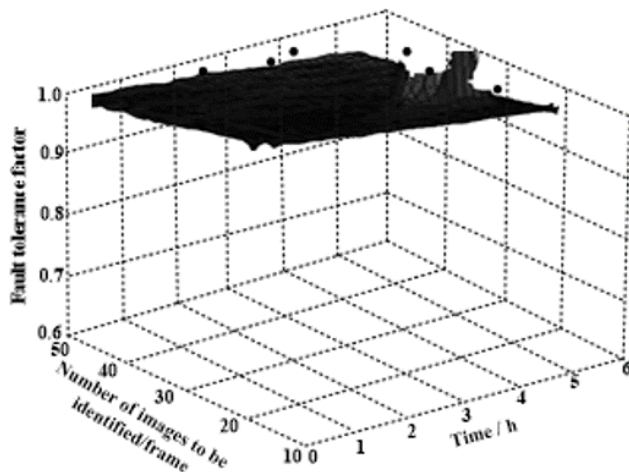
fectively control the recognition delay below 6 μ s, which is feasible.

The average energy consumption of action recognition based on 3D depth image sequences is 121 nJ/bit, that of image recognition based on depth learning is 104 nJ/bit, the improved sparse representation for static-video face recognition is 116 nJ/bit, and the moving video recognition based on BP neural network is 88.6 nJ/bit. From the experimental data, the proposed method has lower energy consumption.

The results of the experiments in Figures 4 to 6 show that the proposed methods show superior performance. This is mainly due to the binarization of the image and the extraction of the image features before the proposed method is used to enhance the contrast between the background and the content of the image. Provides support for reducing the energy consumption, improving the recognition accuracy and enhancing the real-time recognition.



(a) Fault-tolerance of static-video face recognition based on improved sparse representation



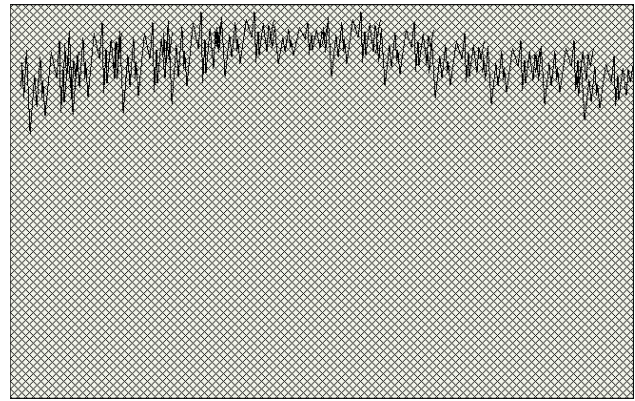
(b) Fault-tolerance of moving video image recognition based on BP neural network

Figure 7: Comparison of fault-tolerance of different image recognition methods

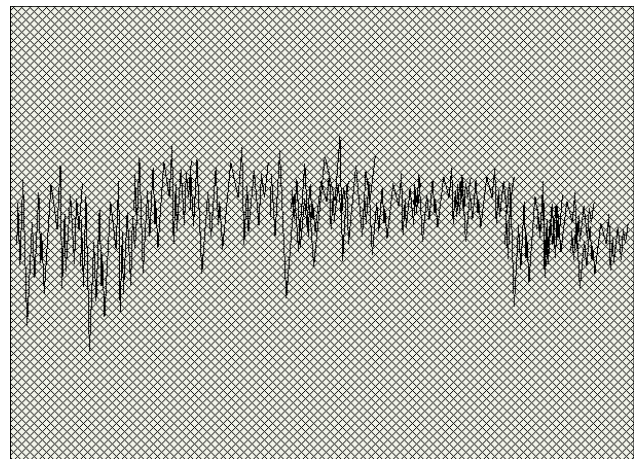
As can be seen from Figure 7, compared with the current research, the fault-tolerance of moving video image recognition based on BP neural networks is better. The proposed method sets the thresholds of the output unit θ_l and the hidden unit 1 and $\phi_{j'}$, which effectively enhances the fault-tolerance in the process of moving video image recognition.

4 Discussion

In this discussion, the fault-tolerance of the moving video image recognition method based on BP neural network is observed with the different range of the hidden unit



(a) Fault-tolerance of moving video image recognition based on BP neural network in $\phi_{j'} \in [0.5, 0.6]$



(b) Fault-tolerance of moving video image recognition based on BP neural network in $\phi_{j'} \in [0.7, 0.8]$

Figure 8: Influence of different values of $\phi_{j'}$ on fault tolerance of image recognition

threshold $\phi_{j'}$. $\phi_{j'}$ is defined in two ranges of $[0.5, 0.6]$ and $[0.7, 0.8]$ respectively, so as to observe the influence of the value of $\phi_{j'}$ on the fault-tolerance of image recognition. The results are as follows:

As shown in Figure 8, the fault-tolerance coefficient of moving video image recognition based on BP neural network is larger in $\phi_{j'} \in [0.5, 0.6]$, and that of the proposed method fluctuates continuously in $\phi_{j'} \in [0.7, 0.8]$, and the overall fault-tolerance coefficient is smaller than 0.9. From the discussion results, it can be seen that the method defines $\phi_{j'}$ in the interval of $[0.5, 0.6]$, which can adjust the image fault-tolerance coefficient to the maximum.

5 Conclusions

As the focus of current research, moving video image recognition has attracted wide attention and many scholars have embarked on research. At present, there are some defects in the related research methods and the performance of the algorithms employed. Thus, a moving video image recognition method based on BP neural network is proposed. Through image processing, image feature extraction and image recognition, the detection and recognition of moving video images are completed. Experimental results show that the proposed method is robust. The following suggestions are put forward for the next research.

BP neural network algorithms have certain advantages in image recognition. They can be combined with the constantly updated new algorithms or methods to further improve the accuracy of image recognition.

Image denoising or enhancement algorithms should be added to further improve the performance of the recognition method.

Acknowledgement: Science and Technology Breakthrough Project of Henan Provincial Science and Technology Department.

Project name: Study on complex image retrieval method based on content diversity (182102210547)

References

- [1] Li F.P., Liang J.G., Du X.F., et al. Research on Intelligent Patrol Robot Based on Image Processing Technology, *Autom. Instrum.*, 2017, (6), 10-12.
- [2] Liu C.Q., Chen B., Pan Z.H., et al. Research of Target Recognition Technique via Simulation SAR and SVM Classifier, *J. China Acad. Electron Inform. Techn.*, 2016, 11(3), 257-262.
- [3] Liu L.L. Research on Image Segmentation Technology for a License Plate Recognition, *Bull. Sci. Techn.*, 2017, 33(4), 125-129.
- [4] Wang M., Ju H.Z., Yao G.Q. Plane Target Recognition in the High Resolution Remote Sensing Image, *Sci. Techn. Eng.*, 2017, 17(18), 265-270.
- [5] Zhang S.Y., Zhu X.L., Wang Y.G., et al. Recognition and Tracking Algorithms of Moving Droplet Based on Inter-Frame Difference Method Combined with Mean-Shift, *J. Shanghai Jiaotong Univ.*, 2016, 50(10), 1605-1608.
- [6] Fan Z.Y., Zeng Y.J., Jiang J., et al. Improved Still-to-Video Face Recognition Algorithm Based on Sparse Representation, *J. Signal Process.*, 2016, 32(5), 567-574.
- [7] Xu H.N., Chen E.Q., Liang C.W. Three-dimensional spatio-temporal feature extraction method for action recognition, *J. Comput. Appl.*, 2016, 36(2), 568-573.
- [8] Liu M.Z., Zheng Y.F., Fan J.F., et al. Area Location and Recognition of Video Text Based on Depth Learning Method, *J. Harbin Univ. Sci. Techn.*, 2016, 21(6), 61-66.
- [9] Cao Y.Q., Cheng W., Huang X.S. Simulation Research on Tracking and Recognition of Moving Objects in Video Images, *Comput. Simulation*, 2017, 34(1), 191-196.
- [10] Tang W., Wang X.T., Wang M.X. Study of FPGA and DSP-based Vehicle License Plate Recognition System, *Comput. Meas. and Control*, 2016, 24(2), 297-299.
- [11] Zhang Q., Xia S.B., Guo P., et al. The Application of Image Recognition Technology in the Measurement Error Detection of Smart Meter, *Electron. Des. Eng.*, 2017, 25(19), 187-189.
- [12] Das R., Thepade S., Ghosh S. Framework for Content-Based Image Identification with Standardized Multiview Features, *Etri J.*, 2016, 38(1), 174-184.
- [13] Zeng H., Kang X. Fast Source Camera Identification Using Content Adaptive Guided Image Filter, *J. Forensic Sci.*, 2016, 61(2), 520-526.
- [14] Zhang F., Hao G., Shao M., et al. An Adipose Tissue Atlas, An Image-Guided Identification of Human-like BAT and Beige Deposits in Rodents, *Cell Metab.*, 2018, 27(1), 252-262.
- [15] Kara I. Investigation of Ballistic Evidence through an Automatic Image Analysis and Identification System, *J. Forensic Sci.*, 2016, 61(3), 775-781.
- [16] Klinshov V., Maslennikov O., Nekorkin V. Jittering Regimes of Two Spiking Oscillators with Delayed Coupling, *Appl. Math. Nonlinear Sci.*, 2016, 1(1), 197-206.
- [17] Oyekale A.S. Cocoa Farmers' Safety Perception and Compliance with Precautions in the Use of Pesticides in Centre and Western Cameroon, *Appl. Ecol. Env. Res.*, 2017, 15(3), 205-219.
- [18] Gao W., Wang Y., Basavanagoud B., Jamil M.K. Characteristics Studies of Molecular Structures in Drugs, *Saudi Pharm. J.*, 2017, 25(4), 580-586.
- [19] Liu Z. What is the Future of Solar Energy? Economic and Policy Barriers, *Energy Sources Part B-Econ. Plan. Pol.*, 2018, 13(3), 169-172.
- [20] Hosamani S.M., Kulkarni B.B., Boli R.G., Gadag V.M. Qspr Analysis of Certain Graph Theoretical Matrices and their Corresponding Energy, *Appl. Math. Nonlinear Sci.*, 2017, 2(1), 131-150.
- [21] Torres-Martinez A., Sanchez A.J., Alvarez-Pliego N., Amalia Hernandez-Franyutti A., Carlos Lopez-Hernandez J., Bautista-Regil J., Gonadal Histopathology of Fish From La Polvora Urban Lagoon in the Grijalva Basin, Mexico. *Rev. Int. De Contaminacion Ambiental*, 2017, 33(4), 713-717.