Application of Human Body Recognition Technology in Scientific Research Sites

Yugang Liu\(^1\) and Xiaofeng Zhang\(^2\)

\(^1\)Engineering College of Information, Engineering University of PAP, Shaanxi, China
\(^2\)Engineering College of Information, Engineering University of PAP, Shannxi, China

Email: \(^1\)43927400@qq.com; \(^2\)zxf801031@126.com

Abstract. In schools and various scientific research institutes, there are a lot of confidential places, such places usually have a person in charge of the management of the scientific research platform. It is an effective method to detect people entering the secret place by video surveillance and give early warning and warning to outsiders. In the video monitoring in normal use, the recognition method of human body is a kind of dynamic pattern recognition, in order to be able to quickly form to the variety of the human body for effective identification, this paper uses the fuzzy clustering method to filter the static image elements in the video image, in view of the dynamic pixel area, by using convolutional neural network to identify the human form, improve the system recognition robustness for optimized recognition effect.

1. Introduction

The monitoring images in scientific research sites are generally dynamic images, and the recognition of dynamic video images is different from that of static images. The recognized objects in the dynamic images will change in shape with time series\([1]\), and the "empty picture" without the recognized body is often found to be entered in frequent use, resulting in a relatively low recognition rate. Therefore, filtering out static elements in "empty screen" can improve the recognition efficiency of dynamic image \([2]\). When someone enters the monitoring range, it will block the pixels in the static "empty frame" image. Therefore, for continuous frames in the video recording, this paper adopts the method of fuzzy clustering (FCM) to do static identification of different areas in a single frame \([3]\) and filter out the static pixels.

In the subsequent human body recognition, the convolutional neural network (CNN) is very robust to two-dimensional image transformation and fault-tolerant to morphing images \([4]\). Therefore, this paper takes the filtered images of continuous frames as the input data of CNN to identify the human body in the dynamic video.

2. Fuzzy Clustering (FCM)

Fuzzy C-mean clustering (FCM) is a clustering algorithm that USES membership degree to determine the degree of each data point to a certain clustering. This algorithm is an improvement on the earlier hard C-mean clustering (HCM) method \([5]\).

When batch processing is running, FCM determines the clustering center \(c_i\) and membership matrix \(U[j]\) by following steps:

Step 1: initialize the membership matrix \(U\) with random Numbers between (0,1) to satisfy the constraints in equation (1)
\[ \sum_{i=1}^{c} u_{ij} = 1 \quad \forall j = 1, \ldots, n \quad (1) \]

Step 2: Use Equation (2) to calculate the \( c_i \) of \( c \) cluster centers, \( i = 1, \ldots, c \).

\[ c_i = \frac{\sum_{j=1}^{n} u_{ij} x_j}{\sum_{j=1}^{n} u_{ij}} \quad (2) \]

Step 3: Calculate the value function according to Equation (3). If it is less than the threshold, or if its change from the value of the last value function is less than a certain threshold, the algorithm stops.

\[ J(U, c_1, \ldots, c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} d_{ij}^2 \quad (3) \]

Step 4: Calculate the new \( U \) matrix with (4). Return to Step 2.

\[ u_{ij} = \frac{1}{\sum_{k=1}^{c} (d_{ij})^{2/(m-1)}} \quad (4) \]

3. Convolutional Neural Network

As a deep neural network, the convolutional neural network adopts local perception technology. Instead of a fully connected structure, the network layer contains the convolutional layer and the lower sampling layer, which greatly reduces the number of weights and improves the efficiency of network operation [6]. It is beneficial to the recognition of two-dimensional shape, and it is also applicable to two-dimensional images with translation, scale, tilt or other forms of deformation. Convolutional neural network mainly consists of two stages [7]:

Forward transmission stage:

Forward propagation is the calculation from input parameters to output results. The output of the previous layer is the input of the current layer, and then through the activation function, the output of the current layer is calculated and passed from layer to layer. Forward propagation is divided into convolution and down-sampling operation, and the calculation of convolutional layer is expressed as formula (5):

\[ x_j^n = f \left( \sum_{i \in M_j} x_i^{n-1} \cdot K_{ij}^n + b_j^n \right) \quad (5) \]

Where \( M_j \) is the set of input characteristic graphs, \( x_j^n \) denotes the \( j \) th characteristic graph of the \( n \)th layer, \( K_{ij}^n \) is the convolution kernel function, \( f() \) is the activation function, generally uses the Sigmoid function as the activation function, \( b_j^n \) is the offset parameter.

The convolution layer and the down sampling layer of convolution neural network are often carried out alternately. The next layer of convolution layer is the lower sampling layer. The calculation of the lower sampling layer is expressed as formula (6):

\[ x_j^{n+1} = f(\beta_{j}^{n+1} \text{down}(x_j^n) + b_j^{n+1}) \quad (6) \]

Where \( \beta \) is the weight constant of the characteristic graph of the down sampling layer, and \( \text{down}() \) is the down sampling function.

In the forward propagation stage, convolution and sampling of convolution neural network are alternately carried out until the output layer gets the output result. In the \( n \)-layer convolution neural network, \( X \) is taken as the input sample of the network, and \( f_n \) is used as the input sample. \( N \) is the activation function adopted by each layer. \( W^{(n)} \) is the connection weight of each layer. The calculation performed by this process network can be expressed by formula (7):
\[ y = f_n \left( \ldots \left( f_2 \left( f_1 \left( X \cdot W^{(1)} \right) \right) \right) \ldots \right) W^{(n)} \]  

The stage of backward propagation:

Backward propagation refers to the error calculation between the result calculated by forward propagation and the label value of a given sample. The error function is expressed as formula (8):

\[ E = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{C} (y^{d}_{ji} - y_{ji})^2 \]  

N is the number of training samples in the input image of convolutional neural network, C is the number of neurons in the output layer of the network, \( y^{d}_{ji} \) is the expected output value of the jth output node of the ith sample, \( y_{ji} \) is the actual output value of the jth output node of the ith sample. After calculating the error between the actual output and the expected output, the network is back propagated according to the principle of minimizing the error, and the network connection weight matrix is adjusted.

4. Human Body Recognition Method

4.1. Experimental Database

The algorithm test database is the monitoring video in a scientific research place, and we select a group of indoor and indoor data as samples. The database includes walking, running, jumping, waving and other actions. Each action is completed by 10 people. The background is fixed and the view angle is fixed. It conforms to the characteristics of the monitoring video image, as shown in Figure 1.

![Figure 1. Monitor video fragments](image_url)

4.2. Experimental Process

The video format is 20 frames per second, and 11 s video is intercepted to ensure sufficient recognition of moving features in continuous time sequence segments. If the video is extracted every 11 frames, 20 frames can be extracted, which can avoid filtering out the phenomenon that the recognized object does not change in consecutive frames. The pixels at the same position in 20 frames form a vector X.

Set the fuzzy weighted index \( m \) of FCM [8] to different values. When \( m \) is large, the background filtering intensity is too high to make the recognized subject almost unable to recognize. When \( m \) is small, there is too much residual in the background, the recognition rate will also be reduced.

Therefore, taking the recognition rate of FCM in CNN [9] with different \( m \) values as the output standard, this paper only discusses the recognition results when \( m \) is rounded. The optimization of \( m \) value will be improved in the follow-up research. When \( m \) is taken as 4, Figure 2 can be obtained.
Then, in order to avoid the effect of fuzzy clustering [10] on innocent people, fuzzy clustering is carried out for vector groups in 20 frames. After calculation, the largest number of classes is still pixels, which can filter out the relatively static pixel positions in the video content.

After filtering out the static pixels of the video, the convolutional neural network is used to recognize the human body. In this experiment, the lenet-5 convolutional neural network model (as shown in Figure 3 is used for reverse deduction, and two fully connected networks are used to reduce a large number of training parameters in the deep network and ensure the robustness of all aspects of the network. The fuzzy network weights are trained by the known human body image data, and then the extracted 60 video images are input into the network in turn, and the square sum of the output errors between the two frames is calculated. If the average error is within the range of (0,1), it can be determined that there is a moving human body in the video image in the time series.

4.3. Experimental Results
The FCM neural network is used to extract 11 samples of human video with a moving interval of 100 seconds.

| number | Average recognition time (s) | Average recognition rate |
|--------|-----------------------------|-------------------------|
|        | FCM-CNN CNN FCM-CNN CNN |
| 1      | 0.52s 0.62s 0.93 0.85       |
| 2      | 0.32s 0.38s 0.95 0.83       |
| 3      | 0.66s 0.51s 0.95 0.87       |
| 4      | 0.83s 0.92s 0.93 0.75       |
| 5      | 0.37s 0.67s 0.94 0.76       |
| 6      | 0.58s 0.26s 0.91 0.83       |
| 7      | 0.43s 0.56s 0.89 0.73       |
| 8      | 0.88s 1.23s 0.98 0.89       |
| 9      | 0.73s 0.85s 0.97 0.88       |
| 10     | 0.71s 0.90s 0.96 0.86       |

Through the test and analysis, fcm-cnn method has a high recognition accuracy, and the recognition speed is improved compared with CNN method. It can effectively and accurately identify whether there is a moving person in the monitoring area by calculating the video image within 11 seconds in the application environment.

5. Conclusion
In this paper, the human recognition in the monitoring image of scientific research sites is studied. For the conventional image recognition mode, firstly, the environmental background that interferes with human recognition is separated by fuzzy clustering method, and then the filtered image is recognized.
by using convolutional neural network, which not only reduces the computational complexity of convolutional neural network, but also improves the robustness of image recognition Experiments show that the method is effective and efficient.

6. Reference

[1] Yu Zhu, Jiangkun Zhao, Yining Wang, et al. Overview of human behavior recognition algorithm based on deep learning [J]. Acta automatica Sinica, 2016, 42 (6): 848-857

[2] Guangfeng Lin, Yindi Fan. Human detection and recognition in intelligent video surveillance [J]. Computer applications, 2007, 27 (S2): 91-92

[3] Zhen Ding, Zhongshan Hu, Jingyu Yang, et al. An image segmentation method based on fuzzy clustering [J]. Computer research and development, 1997 (7): 536-541

[4] Hongtao Lu, Zhang Qinchuan. A review on the application of deep convolution neural network in computer vision [J]. Data acquisition and processing, 2016, 31 (1): 1-17

[5] Qing He. Research progress of fuzzy cluster analysis theory and application [J]. Fuzzy system and mathematics, 1998 (2): 89-94

[6] Ke Xu. Research on the application of convolutional neural network in image recognition [D]. Hangzhou: Zhejiang University, 2012

[7] Yandong Li, Zongbo Hao, Hang Lei. A review of convolutional neural networks [J]. Computer applications, 2016, 36 (9): 2508-2515

[8] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.