Image monitoring and recognition processing based on neural network

L. Min[1], Y. Zhengkun [2]

Changsha Vocational & Technical College, Changsha 410217, Hunan Province, China

[1] ORCID: 0000-0003-2481-3789, lminlanm@yeah.net
[2] ORCID: 0000-0003-0032-311X

Abstract

With the development of economy and the abundance of material, people tend to travel. In the peak season of tourism, the scenic spots are crowded and easy to cause trample and safety problems. The traditional monitoring methods are rigid and have low recognition accuracy. This paper briefly introduced the image monitoring and recognition system and the back-propagation (BP) neural network used for identifying the trampling risk areas in the monitoring images. After that, the image monitoring and recognition system was simulated by using MATLAB software, and it was compared with the traditional entropy method and state-of-the-art CNN. The results showed that the three methods could identify the area with trampling risk in the image, but the image monitoring and recognition system designed in this study was more comprehensive and had lower false alarm rate and shorter recognition time than the traditional information entropy method and state-of-the-art CNN. In summary, the image monitoring and recognition system designed in this study can efficiently and accurately identify the trampling risk areas in the monitoring images.

Keywords: image recognition, information entropy, back-propagation neural network.

1. Introduction

With the development of economy and the popularization of transportation, it is more and more convenient to travel around the world. Moreover, material satisfaction also makes people tend to travel outdoors [1]. In tourist attractions, collisions and frictions between people are inevitable. In off season, such collisions may not be frequent, but in peak season or popular tourist attractions, the frequency of collisions between people will rise dramatically. The increase of tourist density will not only bring more collisions, but also affect the mood of tourists, thus increasing the probability of collisions between tourists and affecting personal safety [2]. Moreover, dense tourist groups will increase the probability of trampling accidents, if not handled in time, it will cause greater personal safety. Therefore, tourist attractions need effective monitoring and timely identification of areas where trample has happened or is going to happen. Traditionally, trampling prevention methods include on-site command, dispatch, and observation of designated areas through surveillance cameras. However, the former is time-consuming and laborious and may not be able to present the whole picture in the dispatching process, but may also be involved in trampling events. The latter also requires manual identification of surveillance images transmitted by cameras. Although it is safer, it is also time-consuming, laborious, and difficult to maintain stable accuracy [3]. Zhong et al. [4] proposed a multi-mode deep learning method, which integrates different visual features as a form of image recognition optimization. The simulation results showed that the scheme had

good performance in heterogeneous visual feature integration of image recognition optimization. Desai et al. [5] recognized the human body posture using the neural network of the restricted Boltzmann machine and realized the intelligent opening of the monitoring equipment, which reduced energy consumption and data volume, and verified the feasibility of the scheme by simulation experiments. Training [6] extracted the features of the images collected using the method of color distribution pattern measurement and verifies its recognition accuracy through the simulation experiment of the ImageLab pedestrian recognition data set. This paper briefly introduces the image monitoring and recognition system, as well as the back-propagation (BP) neural network used for identifying the trampling risk area in the monitoring image. Then the simulation analysis was carried out using MATLAB software, and the image monitoring and recognition system designed in this study was compared with the state-of-the-art CNN and traditional information entropy method.

2. Image monitoring and recognition

The image monitoring and recognition system collects surveillance images through cameras and then transmits them to remote servers through wireless communication networks for analysis and processing. In this study, BP Neural Network [11] was used to identify and analyze the monitoring image. BP Neural Network can effectively fit the non-linear function and can be used for image target recognition. The structure of BP neural network is usually divided into input layer, hidden layer and output layer. Its work flow in the image recognition system is shown in Figure 1.
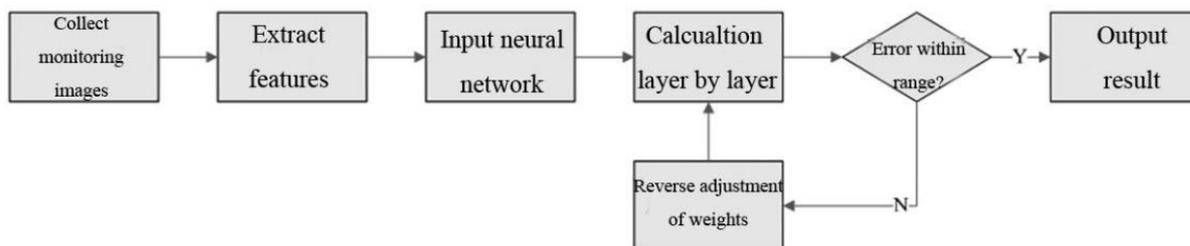

Fig. 1. The flow of image monitoring and recognition

The recognition process contains following main steps.
(1) The surveillance images are collected by cameras distributed in densely populated scenic spots and then transmitted to remote servers through wireless communication networks.
(2) After receiving the surveillance image, feature extraction is carried out, and the surveillance image is set as the matrix of $f(m,n)$, where m,n is the coordinates of the pixel of the image. The feature extraction formula is:

$$
\begin{cases}
u_{pq} = \sum_{m=1}^{M} \sum_{n=1}^{N} (m-x)^p (n-y)^q f(m,n) \\[4pt]
x = \dfrac{\sum_{m=1}^{M} \sum_{n=1}^{N} mf(m,n)}{\sum_{m=1}^{M} \sum_{n=1}^{N} f(m,n)} \\[4pt]
y = \dfrac{\sum_{m=1}^{M} \sum_{n=1}^{N} nf(m,n)}{\sum_{m=1}^{M} \sum_{n=1}^{N} f(m,n)} \\[4pt]
\eta_{pq} = \dfrac{u_{pq}}{u_{pq}^{(p+q)/2+1}}
\end{cases}
, \tag{1}
$$

where u_{pq} stands for the image center moment; p,q stands for the moment of the image matrix, its value is $0,1,2,\cdots$ and moreover $p+q = 2,3,4,\cdots$; M,N are the maximum values of m,n; x,y is the center of gravity of the image; η_{pq} is the central moment after normalization. The corresponding features can be obtained by using the normalized central moments and their second and third order central moments.

(3) The feature matrix that is composed of image features obtained through calculation is taken as the input vector of BP neural network, and then computed layer by layer. The calculation formula [12] is as follows:

$$
\begin{cases}
o_j = f\left(\sum_{i=1}^{n} \omega_{ij} x_i - \alpha_j \right) \\[8pt]
y_k = g\left(\sum_{j=1}^{n} \omega_{jk} o_j - \beta_k \right)
\end{cases}
, \tag{2}
$$

where o_j is the output vector of the hidden layer, α_j is the adjustment item of the hidden layer, $f(\bullet)$ is the activation function of the hidden layer, y_k is the output vector of the output layer, β_k is the adjustment item of the output layer, $g(\bullet)$ is the activation function of the output layer, ω_{ij} and ω_{jk} are the weights of the hidden layer and output layer, respectively.

(4) After calculating step by step, the output vectors obtained are compared with the actual output vectors of training samples, and the error function is obtained:

$$
E = \frac{\sum_{k=1}^{t} (d_k - y_k)^2}{2}
, \tag{3}
$$

where E stands for the error between the calculated output vector and actual output vector and t is the dimension of the output vector.

(5) If the error function is within the prescribed range, the result is output. The weight is reversely adjusted if the error function exceeds the prescribed range. The formula is as follows:

$$\begin{cases} \omega'_{jk} = \omega_{jk} + \eta(1 - y_k)o_j \\ \omega'_{ij} = \omega_{ij} + \eta o_j(1 - o_j)\sum_{k=1}^{t}\sigma_k\omega_{jk}o_j \end{cases},$$

(4)

where η is the learning rate of the adjusted weight and σ_k is the error between the actual vector and expected vector in the k-dimensional output layer.

(6) When the error function reaches the prescribed range or the maximum number of iterations, the training ends, and then the image that needs to be detected is input. Otherwise, procedure (3), (4) and (5) are repeated.

3. Simulation experiment

3.1 Experimental environment

BP neural network model algorithm was compiled using MATLAB software [13]. The experiment was carried out on the laboratory server with configuration of Windows 7 system, i7 processor and 16 G memory.

3.2 Experimental setting

Thirty videos were shot using cameras in densely populated scenic areas. Each video was 10 seconds long. They were uploaded to the cloud through wireless communication module and received by the laboratory server through the cloud. The camera contains 2 Mpx matrix, uses H.264 codec to compress the video, supports wired/wireless transmission, and has 0.008-0.128 m focal length. The wireless communication module used iNET300, the wireless frequency band was 344 MHz, and the maximum transmission rate was 512 kb/s.

The parameters of BP neural network written by MATLAB software in the laboratory server were as follows. The initial weights of the hidden layer and output layer were random numbers in $(-1,1)$; learning rate η of the adjusted weight was 0.1; the number of nodes in the input layer was 7; the number of nodes in the hidden layer was 10; the number of nodes in the output layer was 7.

Moreover, the method of information entropy and convolutional neural network (CNN) was used for comparison to verify the effectiveness of the image monitoring and recognition system. The information entropy method is to transform the image into black and white by gray level method and binarization method, calculate its information entropy, and recognize the crowded trampling area in the image according to the distribution of information entropy in the image. The formula for calculating information entropy [14] is:

$$H(X) = -\sum P(X_i)\log(2, P(X_i)),$$

(5)

where $H(X)$ is the information entropy of an image and $P(X_i)$ represents the output probability function of a variable. Information entropy can reflect the orderliness of image information; the larger the information entropy is, the more confused the information reflected by the image is.

CNN is often used in image recognition. CNN consists of input layer, convolution layer, pooling layer and output layer, in which the convolution layer and pooling layer are equivalent to the hidden layer in the conventional neural network. One of the characteristics of CNN in image recognition is that it does not need additional image feature extraction steps, because the convolution by the convolution kernel in convolution layer is equivalent to the image feature extraction. Relevant parameters of CNN which was used for comparison included one intput layer, one output layer, four convolution layers and four pooling layers. The convolution layer included 96 convolution kernels in a size of 10×10. The pooling layer adopted mean-pooling, the size of the pooling box was 3×3, and the moving step length was 3. Relu function was selected as the activation function in the convolution layer.

As mentioned above, 30 videos were used as experimental data. Among them, 20 videos were randomly selected as training samples for the recognition of three kinds of monitoring images. The remaining 10 videos were used as test samples.

3.3 Judgement standard

In this study, the accuracy, false alarm rate and recognition time are used to judge the performance of the two recognition methods. The calculation formula [15] is:

$$\begin{cases} ACC = \dfrac{TP+TN}{TP+TN+FP+FN} \\ FAR = \dfrac{FN}{TP+FN} \end{cases}, \tag{6}$$

where ACC and FAR stand for the accuracy rate and false alarm rate respectively, TP stands for frame number which is identified as having trampling risk and actually has trampling risk, TN stands for the frame number which is identified as not having trampling risk and actually does not have trampling risk, FP stands for the frame number which is identified as not having trampling risk but actually has trampling risk, and FN stands for the frame number which is identified as having trampling risk but actually does not have trampling risk. In the equation, the trampling risks include trample which is going to happen and trample which has happened. The recognition time refers to the time from inputting video to receiving the recognition result by the monitoring terminal.

3.4 Experimental results

Due to the limitation of space and the huge number of frames in video, only part of the image recognition results is displayed. The monitoring image recognition results of the method proposed in this study are shown in Figure 2, the monitoring image recognition results of the method based on information entropy are shown in Figure 3, and the monitoring image recognition results of the method based on CNN are shown in Figure 4. The comparison of

Figure 2, 3 and 4 demonstrated that both methods could identify the area trampled by disputes effectively (shown in the red box), but after comparison, it was found that the area identified by the method proposed in this study was more comprehensive and accurate, while the area identified by the information entropy based method and CNN based method was relatively small and the key area had deviation.

Fig. 2. The image monitoring and recognition results of the method proposed in this study

Fig. 3. The image monitoring and recognition results of the method based on information entropy

Fig. 4. The image monitoring and recognition results of the method based on CNN

As shown in Figure 5, the accuracy of the method proposed in this study was over 93%, up to 99.2%, with an average accuracy of 96.4%, the accuracy of the information entropy based image surveillance and recognition was between 70% and 85%, with an average accuracy of 78.5%, and the accuracy of the method based on CNN was between 80% and 90%, with an average accuracy of 84.0%. It can be clearly seen from Figure 5 that the accuracy of the method proposed in this study and based on CNN was higher than that of the information entropy

based image monitoring and recognition. The accuracy of the recognition based on the method proposed in this study was significantly higher than that of the method based on CNN.
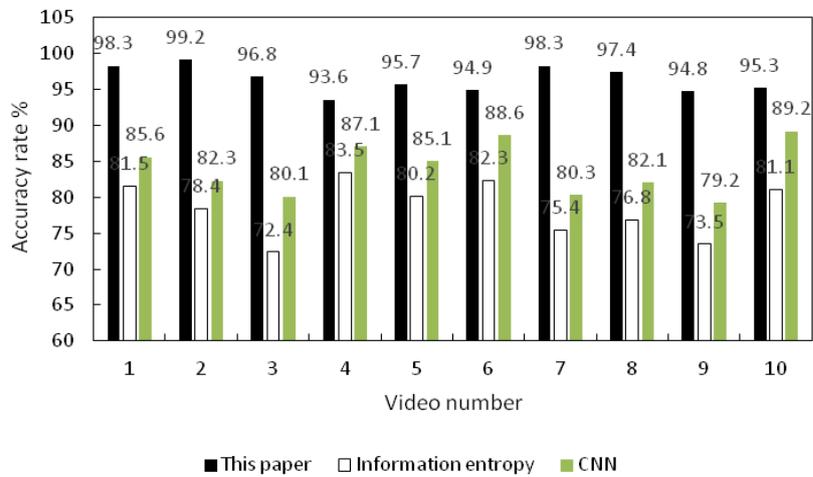

Fig. 5. Recognition accuracy of three methods

Recognition false alarm rate refers to the probability of identifying images without trampling risk as images with trampling risk. High false alarm rate will lead to human waste in maintaining the order of scenic spots. Therefore, for the image monitoring and recognition system of scenic spots, the smaller the false alarm rate is, the better the result is. As shown in Figure 6, in the testing of 10 videos, the false alarm rate of the image recognition system proposed in this study fluctuated from 5% to 6%, with an average false alarm rate of 5.1%; the false alarm rate of the image recognition system based on information entropy fluctuated from 10% to 16%, showing a large amplitude of variation, with an average false alarm rate of 13.2%; the false alarm rate of the image recognition system based on CNN was between 7.8% and 10.3%, with an average false alarm rate of 8.9%. It was seen from the false alarm rate of the three kinds of recognition system in recognizing the video with the same number that the false alarm rate of the image recognition method proposed in this study was significantly lower than that of the recognition based on information entropy and CNN.
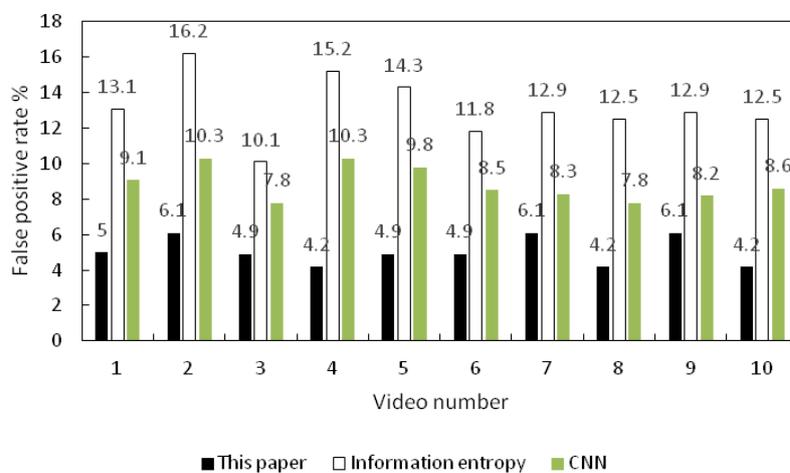

Fig. 6. Recognition false alarm rate of three methods

As shown in Figure 7, in the testing of the 10 videos, the time required the image monitoring and recognition method proposed in this study ranged from 1.98 s to 3.55 s, with an average recognition time of 5.06 s; the time required by the information entropy based image monitoring and recognition ranged from 5.26 s to 7.35 s, with an average recognition time of 13.15 s; the time required by the CNN based image monitoring and recognition ranged from 3.56 s to 5.23 s, with an average recognition time of 4.70 s. By comparing the recognition time of the three methods in testing the video with the same number, it was found that the time needed by the image monitoring and recognition method proposed in this study was obviously less than that by information entropy and CNN based image monitoring and recognition. In conclusion, the efficiency of monitoring and identification has been greatly improved after the adoption of the method proposed in this study.
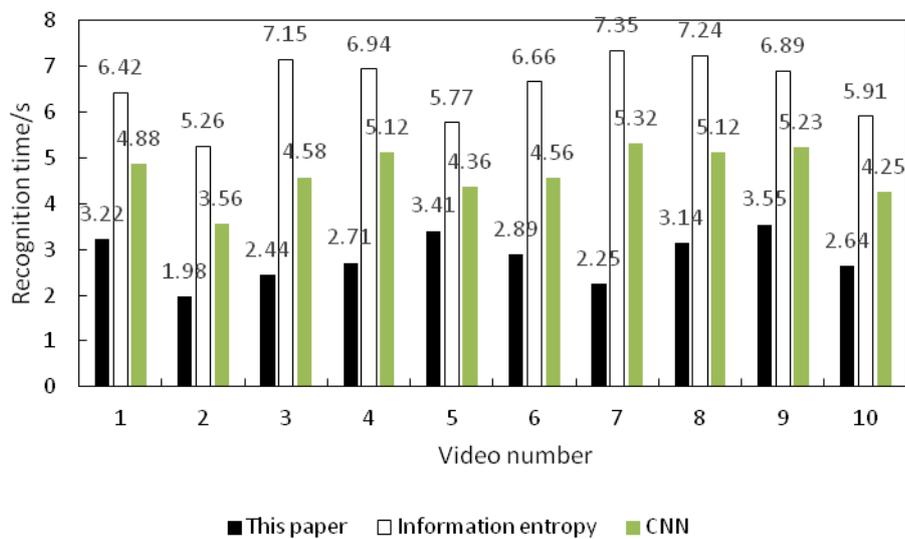

Fig. 7. Recognition time of three methods

4. Conclusion

This paper briefly introduced the image monitoring and recognition system and the BP neural network used for identifying the trampling risk area in the monitoring image. Then the simulation analysis was carried out using MATLAB software and compared with the traditional information entropy method and state-of-the-art CNN. The results are as follows. The three methods could identify the trampling risk area in the monitoring image. However, the trampling risk areas identified by the image monitoring and recognition system proposed in this study were more comprehensive and accurate. In terms of recognition accuracy, the average accuracy of the recognition system proposed in this study was 96.4%, the average accuracy of the information entropy based recognition system was 78.5%, and the average accuracy of the CNN based recognition system was 84.0%. In terms of false alarm rate, the average false alarm rate of the recognition system proposed in this study was 5.1%, and the average false alarm rate of the information entropy based recognition system was 13.2%. In terms of recognition time, the average recognition time of the recognition system proposed in this study was 5.06 s, the average recognition time of the information entropy based recognition system was

13.15 s, and average recognition time of the CNN based recognition system was 4.70 s. In this study, Hu invariant moment was used as the recognition feature of monitoring image, and BP neural network was combined to identify the stampede risk area in the image quickly. The recognition effect of the proposed method was verified by the simulation experiment. The future research direction is to further improve the accuracy of surveillance image recognition system and reduce the recognition time. The purpose of the image recognition and monitoring system studied in this study is to monitor the flow density and release pre-warning in crowded scenic areas.

Acknowledgement

References

[1] Fukada H., Kasai K., Shou O. A Field Test of System to Provide Tourism Information Using Image Recognition Type AR Technology. Lecture Notes in Electrical Engineering, Vol. 312, pp. 381-387, 2015.

[2] Elliot S., Papadopoulos N. Of products and tourism destinations: an integrative, cross-national study of place image. Journal of Business Research, Vol. 69, No. 3, pp. 1157-1165, 2016.

[3] Zhou L., Li Q., Huo G., Zhu G. Face Image Recognition Method Based on the NSCT and Bionic Pattern. Laser & Optoelectronics Progress, Vol. 52, No. 3, pp. 126-133, 2015.

[4] Zhong F., Chen Z., Ning Z., Min G., Hu Y. Heterogeneous Visual Features Integration for Image Recognition Optimization in Internet of Things. Journal of Computational Science, pp. S18777503163076.

[5] Desai S., Mohammed S., Raychowdhury A. An ultra-low power, "always-on" camera front-end for posture detection in body worn cameras using Restricted Boltzman Machines. IEEE Transactions on Multi-Scale Computing Systems, Vol. 1, No. 4, pp. 187-194, 2015.

[6] Ye Y.S., Zhang X.M., Ng W.Y. Color Distribution Pattern Metric for Person Reidentification. Wireless Communications and Mobile Computing, Vol. 2017, pp. 1-11, 2017.

[7] Hamedani K., Seyyedsalehi S.A., Ahamdi R. Video-based face recognition and image synthesis from rotating head frames using nonlinear manifold learning by neural networks. Neural Computing & Applications, Vol. 27, No. 6, pp. 1761-1769, 2016.

[8] Dong S., Yuan Z., Gu C., Yang F. Research on intelligent agricultural machinery control platform based on multi-discipline technology integration. Transactions of the Chinese Society of Agricultural Engineering, Vol. 33, No. 8, pp. 1-11, 2017.

[9] Fiore G.D., Mainetti L., Mighali V., Patrono L., Alletto S., Cucchiara R., Serra G. A Location-Aware Architecture for an IoT-Based Smart Museum. International Journal of Electronic Government Research, Vol. 12, No. 2, pp. 39-55, 2016.

[10] Bondi L., Baroffio L., Cesana M., Redondi A., Tagliasacchi M. Open-source and flexible framework for visual sensor networks. IEEE Internet of Things Journal, Vol. 3, No. 5, pp. 767-778, 2017.

[11] Liao B., Wang H. The Optimization of SIFT Feature Matching Algorithms on Face Recognition Based on BP Neural Network. Applied Mechanics & Materials, Vol. 743, pp. 359-364, 2015.

[12] Li Q.H., Liu D. Aluminum Plate Surface Defects Classification Based on the BP Neural Network. Applied Mechanics & Materials, Vol. 734, pp. 543-547, 2015.

[13] Su J.H., Piao Y.C., Luo Z., Yan B. Modeling Habitat Suitability of Migratory Birds from Remote Sensing Images Using Convolutional Neural Networks. Animals, Vol. 8, No. 5, pp. 66, 2018.

[14] Raju P., Rao B.P., Rao V.M. Gray Wolf Optimization-Based Artificial Neural Network for Classification of Kidney Images. Journal of Circuits Systems & Computers, pp. S0218126618502316, 2018.

[15] Zhou J., Wang Q., Yi M., Wang S. Acoustic Emission Signal Recognition Based on Wavelet Transform and BP Neural Network. Journal of Qingdao University of Science and Technology, Vol. 8, No. 3, pp. 80-85, 2015.