Eng Int Syst (2021) 4: 235–239

Engineering
Intelligent Systems

Bank Abnormal Behavior Recognition Technology Based on Deep Learning

Yueyun Du*

Computer Department, Shangqiu Vocational and Technical College, Shangqiu 476000, China

To resolve the problem of low recognition accuracy of existing bank abnormal behavior recognition methods, a bank abnormal behavior recognition method based on deep learning is proposed. The static background of a bank surveillance video image is acquired by the Mixture-of-Gaussians (MoG) model. The static background is extracted by background subtraction, and the foreground image is filtered to make the moving human target as clear as possible. The video frame image is divided into blocks to obtain the motion effect map of the foreground area, and the motion effect map features of each space-time block are extracted. After obtaining the depth features of moving objects, the sparse reconstruction method is used to introduce a coefficient learning dictionary and a sparse coding vector to judge whether the behavior is normal or not according to the sparse reconstruction cost. The experimental results show that compared with traditional methods, the proposed method based on deep learning is more accurate and has a higher application value.

Keywords: MoG model, Background subtraction, Feature extraction, Coefficient reconstruction, Abnormal behavior recognition

1. INTRODUCTION

A bank is a financial institution established according to law to operate monetary and credit business. It is also a supervisory system for the development of commodity and monetary economy, as well as a financial supervisory unit. In the supervision system of modern banks, a monitoring system is added. The monitoring system mainly focuses on video recordings of camera equipment. The image captured by the camera is analyzed to ensure the safety of the operation of the banking system. However, commercial banks have the problem of unattended Automated Teller Machines (ATM), and the monitoring system cannot intelligently judge whether there are suspicious people loitering around ATMs when customers withdraw money. To a certain extent, it is difficult to guarantee the safety of bank customers. Effective bank ATM monitoring systems should automatically raise an alarm after identifying robbery, fighting, suspected loitering and other acts, and take the initiative to suspend cash payments, deactivate cards or lock the door of the protective room, so as to effectively prevent crimes before they occur. Therefore, it is particularly important to study the abnormal behavior recognition of bank ATM monitoring systems, which has become the focus of industry research and has received extensive attention. At present, there are many methods in this field.

Literature [1] presents a method for identifying abnormal behavior of bank intelligent monitoring equipment. This method uses smart cameras to collect front-end video information, and can quickly complete the detection and recognition process of different behaviors. Intelligent cameras are responsible for collecting all-day on-site video signals and transferring them to the system for software analysis, to determine whether abnormal behavior information is collected, and to raise alarms quickly. However, there are some problems in this method, such as poor real-time detection of abnormal behavior, low recognition rate of classification algorithms and fewer features.

*Corresponding Author e-mail: yjm1314926@126.com

Literature [2] proposes a 3D DenseNet detection algorithm for abnormal behavior of small and medium-sized populations based on bank intelligent monitoring. Firstly, the fast population density detection algorithm is used to extract the population change information. Secondly, the average kinetic energy, direction entropy and distance potential energy of the crowd in the video are extracted. Finally, the extreme learning machine algorithm is used to classify the crowd behavior, and then the common data set is used for testing. This method has good real-time performance in detecting abnormal behaviors of the bank population, and the 3D DenseNet detection algorithm for intelligent monitoring of abnormal behavior of bank population can effectively improve the amount of detection. However, this method has the problem of low accuracy in identifying abnormal bank behavior.

Aiming at the problem of low recognition accuracy in existing methods, a bank abnormal behavior recognition method based on deep learning is proposed. Experiments show that the method proposed in this paper has a higher accuracy in identifying abnormal bank behavior, can effectively detect the information of dangerous bank behavior, and maintain the safety of bank operations.

2. THE METHOD OF IDENTIFYING BANK ABNORMAL BEHAVIOR

2.1 Foreground Image Extraction and Filtering

The background image of video surveillance in bank is acquired by the MoG model. The static image excluding moving objects is obtained. The static background image is then extracted by the background subtraction method, and the extracted foreground image is filtered. Thus, the target image obtained has good integrity and is suitable for further targeted behavior recognition [3].

Before extracting the static background image, the static background image of the surveillance video needs to be modeled by using MoG, the video frame can then be traversed to check for a match between the pixel value and the Mixture Gauss background model. The successful match is considered the background, and the inverse is the foreground.

MoG background modeling can extract the foreground of a surveillance video from the background, based mainly on the following two factors. First, through traversing the entire video frame, it is found that most of the pixels support the background distribution, while the foreground target image has little influence on the distribution, generally located in the Gauss model as other than K+1. Second, when the pixel value of the moving object is close to the background pixel value, the relative position of the moving object's pixel value changes significantly, and it is therefore easy to extract [4]. The MoG background modeling process is as follows:

$$P(x_t) = \sum_{i=1}^{K} w_{i,t} * (x_t; \mu_i, \delta_{i,t})t \tag{1}$$

In Formula (1), the value of pixel x in video frame at time t is expressed by x. If the background color is RGB, then x is a three-dimensional vector: $x_t = (R_t, G_t, B_t)$. The weights of the i-th Gauss distribution of time t are expressed in $w_{i,t}$, $\mu_{i,t} = \left(\mu_{i,t}^R \mu_{i,t}^G \mu_{i,t}^B\right)$, in a probability density function, $\mu_{i,t}$ and $\delta_{i,t}$ in i mean vector and a covariance matrix of time t, respectively.

Among them:

$$\mu_{i,t} = \left(\mu_{i,t}^R \mu_{i,t}^G \mu_{i,t}^B\right) \tag{2}$$

$$\delta_{i,t} = \begin{bmatrix} \sigma_R^2 & 0 & 0 \\ 0 & \sigma_G^2 & 0 \\ 0 & 0 & \sigma_B^2 \end{bmatrix} \tag{3}$$

Due to the influence of background noise in video, the extracted moving object contains noise points, so the background subtraction method is used to extract the bank abnormal behavior subtraction image [5].

The background subtraction method is based on the assumption that the background of the surveillance video changes slightly. It can extract the foreground moving object behavior by subtracting the static background from the video frame image [6]. If the noise generated by the surveillance video is ignored, the small video frame $I(x, y, t)$ can be regarded as two parts: the video image $b(x, y, t)$ and the image foreground moving target $m(x, y, t)$.

$$m(x, y, t) = I(x, y, t) - b(x, y, t) \tag{4}$$

Firstly, the captured video frame image is preprocessed to determine the abnormal area. The abnormal behavior difference image $d(x, y, t)$ of the bank is then obtained by dividing the video frame sequence from the background. The process is as follows:

$$d(x, y, t) = \frac{I(x, y, t) - b(x, y, t)}{m(x, y, t)} \tag{5}$$

In order to achieve the desired result of a moving human target, a morphological filtering operation is also needed for the video frame image after subtraction [7]. Because the video frame image has rotation invariance, that is, the descending gradient of its distribution in all directions is the same, it is necessary to ensure that the gradient of the image pixel value is the same when the filtering operation is carried out.

As the two-dimensional Gauss distribution is a monotone descent function in one direction, the same as the weighted average filtering function, it is very suitable for image filtering operations [8]. By increasing the pixel value of the foreground image pixels, the image becomes highlighted. The minimum pixel value in the region is then obtained and filtered to obtain a complete and clear foreground target image [9]. This process is expressed as follows:

$$dst(x, y, y) = \max src(x, y, t) + \min src(x, y, t) \tag{6}$$

2.2 Extracting Depth Characteristics of Human Target Motion

To improve the accuracy of bank abnormal behavior detection, it is necessary to determine the description of an object's motion characteristics [10]. According to the actual human

motion model, the different effects of abnormal behavior on the surrounding environment are analyzed in depth, and the effects of moving prospects on the surrounding environment are described.

Firstly, the foreground motion is extracted from the video frame sequence by using the adaptive Gaussian mixture model, and the foreground motion block is obtained by partitioning the space. The motion vectors of the moving foreground blocks are then obtained according to the dense optical flow of video frames. Finally, the effect maps of all of the moving foreground blocks are obtained by calculating the effect of all the moving foreground blocks, and the effect characteristics of space-time blocks are extracted by using the spatial sub-blocks of continuous multi-frames as space-time blocks.

Firstly, the foreground image extracted by the adaptive Gaussian mixture model is segmented into $W \times L$ image blocks in space, and the size of each image block is N *Pixel* $\times$ N *pixel*. Since the effects of various parts of a moving object, such as the effect of human hand swing, should be analyzed in detail, the criterion for determining the foreground is that a single foreground block can be expressed as a part of a moving foreground target. Secondly, image blocks are preprocessed [11]. This is because there may be sporadic foreground spots in the foreground image extracted by the adaptive Mixture Gauss model. These foreground points may be caused by noise, subtle background changes or small motion amplitude of moving objects, which cannot clearly represent the motion behavior of an object. Therefore, only when the number of foreground points of the image reaches a certain threshold, can they be used as moving foreground blocks. The segmentation of each image block is processed as follows:

Assuming that $B_j(1 \le J \le W \times L)$ represents the jth image block and b_j is the number of foreground points, it can only be retained as a moving foreground block if Formula (7) is satisfied, otherwise the block will be removed as noise.

$$G_i = B_j \quad if \frac{b_j}{N \times N} > \lambda \tag{7}$$

Among them, $G_j(1 \le i \le S)$ means that when the image block B_j satisfies the condition, it can be used as the ith moving foreground block, which is the threshold value of foreground points, and the value is $[0.1 - 0.4]$.

For the foreground image with noise points, noise interference mainly arises from the process of image acquisition and transmission, which causes some background pixels to change greatly in the time series and form a new Gauss distribution over a short time. The adaptive Gauss mixture background model misjudges it as foreground point, which leads to the generation of isolated foreground points. The small number and random distribution of outliers will show these points to be noise rather than foreground. Preprocessing the foreground image and removing the small proportion of foreground points from the background can effectively remove sporadic foreground spots, so that the extracted moving foreground blocks can more effectively represent the real motion state of the object.

A moving foreground block can represent a part of the entity of a moving object such as a pedestrian or a car. The motion effect of the foreground block is described by extracting the optical flow information of the pixels in the foreground block.

Firstly, the dense optical flow algorithm is used to calculate the optical flow vector of each pixel in each frame in the original image sequence. Assuming $\{G_1, G_2, \ldots, G_s\}$ is the extracted moving foreground block, the average value of optical flow vectors of all pixels is extracted as the optical flow vectors of the current block, as shown in Formula (11).

$$g_i = \frac{1}{J} \sum_i o_i^j \tag{8}$$

Among them, g_i represents the optical flow vector of the i moving foreground block and o_i^j represents the optical flow vector of the j pixel in the i moving foreground block. In order to facilitate subsequent processing, assume that $\|g_i\|$ and $\measuredangle g_i$ represent the magnitude and direction of the optical flow of the ith moving foreground block, respectively.

The results of the motion foreground effect map can show that the direction of motion of objects is easily affected by many factors, and the influence of these surrounding factors on the movement of pedestrians affects the movement of the whole crowd [12]. In this paper, the effect of pedestrians or other moving objects on the surrounding space is called motion effect. In order to measure the influence of moving foreground block G_i on surrounding space block B_i, two index variables γ_{ij}^d and γ_{ij}^d are defined. The calculation process is as follows:

$$\gamma_{ij}^d = \begin{cases} 1 & if\, d(i,j) < \delta_d \\ 0 & else \end{cases}$$
$$\gamma_{ij}^d = \begin{cases} 1 & if\, F^- < \theta_{ij} < F_i^+ \\ 0 & else \end{cases} \tag{9}$$

Among them, $d(i, j)$ represents the Euclidean distance between the moving foreground block G_i and the space block B_j, δ_d, is the distance value, θ_{ij} is the angle between the optical flow of the vector G_i from the foreground block G_i to the space block B_j, $[F_i^- F_i^+]$ represents the visual field range of the moving foreground block. Therefore, the weights of the effect of the moving foreground block G_i on spatial block B_j can be defined as:

$$W_i = \left\{ \begin{array}{l} \gamma_{ij}^d \gamma_{ij}^\theta \left(-\frac{d(i,j)}{\|g_i\|}\right) \\ \gamma_{ij}^d \gamma_{ij}^\theta \left(-\frac{d(i,j)}{\|g_i\|}\right) \end{array} \right\} \quad if\, G_i \ne B_j \tag{10}$$

Because space block B_j is affected by foreground blocks with different moving directions, the effect of foreground blocks should be counted according to different moving directions, so that the extracted features can be more distinguished. In order to facilitate calculation, the motion direction of moving foreground blocks is quantified as follows.

$$q(\measuredangle g_i) = k \quad if (k-1) \times \frac{2\pi}{p} < \measuredangle g_i \le k \times \frac{2\pi}{p} \tag{11}$$

Among them, $k \in \{1, 2, 3, \ldots, p\}$ is the total number of quantized direction intervals. According to the quantization direction of optical flow of the moving foreground block, the effect of the foreground block to space block B_i is counted as follows:

$$h^{j(k_i)} = \sum w_{ij} \quad if\, q(\measuredangle g_i) = k_i \tag{12}$$

Table 1 Composition Structure of Experimental Data Sets.

Abnormal behavior category	Training set/g	Verification set/g	Test set/g
Fighting	400	30	70
Loitering	400	30	70
Robbing	400	30	70
Normal	400	30	70

Among them, k_i represents the quantized direction index value of the optical flow of the i moving foreground block. After calculating the motion foreground effect weights of all spatial blocks B_i, the motion foreground effect of a frame image in a video sequence can be constructed. Each space block of the motion foreground effect map can be represented by p dimension vector. In order to calculate the prospects of motion for a longer period of time, the space blocks of continuous t frames are taken as a space-time block, and the feature vectors of the space blocks of t frames are superimposed as the features of the whole space-time block [13].

2.3 Bank Abnormal Behavior Detection Based on Sparse Reconstruction

After obtaining the depth features of moving objects, the sparse reconstruction method is used to detect abnormal behavior. The basic idea is that any behavior can be represented by a sparse linear combination of a set of normal training samples. Generally speaking, for normal behavior, the cost of sparse reconstruction is smaller, while the cost of sparse reconstruction for abnormal behavior is larger. Therefore, we can detect abnormal behavior according to reconstruction error [14].

There are C class normal behaviors, each behavior is represented by the above eigenvectors, and $D = [D_1, D_2, \ldots, D_C]$ represents a sparse dictionary. D_i is a sub-dictionary composed of K class i behaviors. For the test sample y, it can be expressed as:

$$y = D_\alpha \tag{13}$$

Among them, $\alpha = [\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_C]^T$ is a sparse coding vector.

The key to sparse reconstruction is dictionary learning and sparse coding. In dictionary learning, given a training sample set $Y = [y_1, y_2, \ldots, y_N] \in R^{m \times N}$, $y_i \in R^m$ denotes the eigenvector of the i normal sample, and defines the learning dictionary D and sparse coding vector α, so that Y can be reconstructed by weighting dictionary, that is, $Y = D_\alpha$, in order to solve the following optimization problems:

$$\min_{D,\alpha} \|Y - D\alpha\|_F + \lambda\|\alpha\|_{2,1} \tag{14}$$

Among them, λ is the control parameter, the first one is a reconstruction error, and the second one is a sparsity constraint [15]. Obviously, this is currently a non-convex optimization problem, however, if any of the values in D and α are fixed, it will become a linear problem. Therefore, by fixing D and α in turn, specific D and α values can be derived.

For abnormal behavior detection, given a dictionary D, the test sample y can be represented by Formula (15), where sparse coding α can be obtained by solving the following formula:

$$\alpha^* = \min_{\alpha} \|y - D\alpha\|_2 + \|\alpha\|_1 \tag{15}$$

The cost of sparse reconstruction is calculated based on the optimal sparse code α^*.

$$S(y, \alpha^*, D) = \|y - D\alpha\|_2 + \lambda\|\alpha^*\|_1 \tag{16}$$

For normal behavior, the cost of sparse reconstruction is relatively small, while the cost of sparse reconstruction for abnormal behavior is relatively high, thus abnormal behavior and normal behavior can be distinguished.

$$S(y, \alpha^*, D) > \varepsilon \tag{17}$$

Among them, y is abnormal behavior, ε is a pre-set threshold.

3. EXPERIMENTAL RESULTS AND ANALYSIS

A simulation data set of ATM behavior recognition is established, divided into four categories. In order to distinguish normal behavior from abnormal behavior, normal withdrawal behavior was selected as the control. Each category is divided into training set, verification set and test set, which contains video clips in g, each video frequency band lasts for 10S, and the data set structure is shown in Table 1.

The experiment adopts the form of comparison, and puts the methods proposed in this paper and those in reference [2] under the same conditions, and counts the accuracy of the two methods in predicting abnormal bank behavior. The specific statistical results are shown in Table 2.

Table 2 shows that the recognition accuracy of abnormal bank behavior recognition method based on in-depth learning is higher than that of literature [2]. This is because this method extracts and fuses image depth appearance features and depth motion features by using the weighted correlation method, improves the classification ability of fusion features, and uses the sparse reconstruction method to detect abnormal behavior, which makes the accuracy of abnormal bank behavior recognition of the proposed method higher than that of the existing methods, and solves the problem of abnormal bank behavior recognition.

4. CONCLUDING REMARKS

A bank anomaly detection method based on deep learning is proposed. The background subtraction method is used to

Table 2 Comparison of Results by Method.

		Paper method		Literature [2] method	
Behavior category	Test number/g	Mistaken number/g	Accuracy rate/%	Mistaken number/g	Accuracy rate/%
Fighting	30	0	100.0	1	96.7
Loitering	30	2	93.3	5	83.3
Robbing	30	4	86.7	6	80.0
Normal	30	3	90.0	3	90.0

extract the foreground of the static background after MoG modeling, and the image is then filtered. The foreground image block is segmented using the preprocessing method of foreground selection, and the effect map of the moving foreground is calculated to extract the features of the motion foreground effect map. Finally, we use the method of judging the sparse reconstruction cost to determine whether the bank's behavior is normal or not. The experimental results show that the proposed method can effectively improve the accuracy of bank abnormal behavior recognition, effectively maintain the normal and safe operation of bank affairs, and is overall more advantageous.

ACKNOWLEDGMENTS

The work was supported by Key R&D and Promotion Projects of Henan Science and Technology Department (No. 182102210480).

REFERENCES

1. Yuan, Y., Lu, Y., & Wang, Q. (2017). Tracking as a Whole: Multi-Target Tracking by Modeling Group Behavior with Sequential Detection. *IEEE T Intell Transp*, 1–11.
2. Zheng, X., Zhang, X., & Yu, Y. (2016). ELM-based Spammer Detection in Social Networks. *J. Supercomputing* (72), 2991–3005.
3. Shen, F., Vecchio, D., Mohaisen, J.A., et al. (2018). Android Malware Detection using Complex-Flows. *IEEE T Mobile Comput 1*, 1.
4. Ameli, A., Hooshyar, A., El-Saadany, E., et al. (2018). Attack Detection and Identification for Automatic Generation Control Systems. *IEEE T Power Syst 1*, 1.
5. Fanaee, T.H., Gama, J. (2016). Tensor-based Anomaly Detection: An Interdisciplinary Survey. *Knowl-Based Syst* (98), 130–147.
6. Candy, J.V., Franco, S.N., Ruggiero, E.L., et al. (2017). Anomaly Detection for a Vibrating Structure: A Subspace Identification/Tracking Approach. *J Acoust Soc Am* (142), 680–696.
7. Li, B., Jing, Y., & Xu, W.A. (2017). Generic Waveform Abnormality Detection Method for Utility Equipment Condition Monitoring. *IEEE T Power Deliver* (32), 162–171.
8. Sharma, M., Deb, D., & Acharya, U.R. (2018). A Novel Three-Band Orthogonal Wavelet Filter Bank Method for an Automated Identification of Alcoholic EEG signals. *Appl Intell* (48), 1368–1378.
9. Mandadi, K., & Kumar, B.K. (2016). Identification of Inter-Area Oscillations Using Zolotarev Polynomial Based Filter Bank with Eigen Realization Algorithm. *IEEE T Power Syst* (31), 4650–4659.
10. Wang, H., Hao, Q., Cao, J., et al. (2018). Target Recognition Method on Retina-Like Laser Detection and Ranging Images. *Appl Optics* (57), B135.
11. Shameem, K.M.M., Choudhari, K.S. & Bankapur, A. (2017). A Hybrid LIBS–Raman System Combined with Chemometrics: An Efficient Tool for Plastic Identification and Sorting. *Anal Bioanal Chem* (409), 1–10.
12. Ambusaidi, M., He, X., Nanda, P., et al. (2016). Building an Intrusion Detection System Using a Filter-based Feature Selection Algorithm. *IEEE T Comput* (65), 2986–2998.
13. Gerhardt, M., Vennet, R.V., et al. (2017). European Bank Stress Test and Sovereign Exposures. *Applied Economics Letters* (24), 1–5.
14. Le, C.H.A. (2016). Macro-financial Linkages and Bank Behaviour: Evidence from the Second-Round Effects of the Global Financial Crisis on East Asia. *Eurasian Economic Review, 6* (3), 1–23.
15. Zhang, Z., Mei, X., & Xiao, B. (2016). Abnormal Event Detection via Compact Low-Rank Sparse Learning. *IEEE Intell Syst.* (31), 29–36.