Recognition Method of Abnormal Behavior in Electric Power Violation Monitoring Video Based on Computer Vision

Mancheng Yi(✉), Zhiguo An, Jianxin Liu, Sifan Yu, Weirong Huang, and Zheng Peng

Guangzhou Power Supply Bureau of Guangdong Power Grid Co., Ltd., Guangzhou 510000, China
anzhiguo123444@163.com

Abstract. In order to improve the accuracy of abnormal behavior recognition in electric power illegal behavior monitoring, a method of abnormal behavior recognition in electric power illegal video based on computer vision is proposed. The monitoring image of electric power violations is collected by sensors, and the monitoring video image is preprocessed based on mathematical morphology and neighborhood average filtering; The static target detection method and background difference method in computer vision technology are used to separate the background and moving foreground in the video frame sequence; Locate the staff in the video image and track their movement track; Fusing FAST corner and SIFT algorithm to extract corner features and texture features of staff action behavior in the monitoring image; The above features are input into the long and short memory recurrent neural network to realize the recognition of abnormal behavior in the electric power illegal monitoring video. The results show that the Kappa coefficient between the method and the measured results remains above 80%, which proves that the recognition method improves the accuracy of abnormal behavior recognition.

Keywords: Computer Vision · Abnormal Behavior of Electric Power Violation · Surveillance Video · Identification Method Middle Figure Classification

1 Introduction

The operating environment of electric energy is complex, facing challenges of high voltage and high voltage operation. To ensure the safety of designers during power operation, the energy industry requires energy operators to monitor operators and operational processes in real-time. In the past, human supervision was often used, not only wasting a lot of work, but also due to the fatigue, laxity, and other personal reasons of superiors, which often led to inadequate supervision, and human supervision could not simultaneously balance the global situation. With the development of high-definition video monitoring, high-definition cameras are widely used in power construction sites for all-round, large-scale and long-term supervision. At present, security personnel at power operation sites

B. Wang et al. (Eds.): ICMTEL 2023, LNICST 533, pp. 168–182, 2024.
https://doi.org/10.1007/978-3-031-50574-4_12

often use high-resolution cameras to capture on-site information, and then supervisors inspect the returned videos to determine whether the operator has engaged in illegal activities. However, the observation range of the human eye is limited, and long-term observation may lead to visual fatigue. However, using computer vision technology to detect risks on returned images can aid in security monitoring. When risk information is detected, an alarm is issued, and then security monitoring personnel conduct inspections based on video information, greatly improving the work efficiency of security monitoring personnel and adding a "dual guarantee" for risk detection in the energy operation center. Reference [1] proposes a video anomaly detection method considering crowd density. Design and generate the framework of the confrontation network; According to the scene density and the object of the behavior, a video anomaly detection model based on the generation of confrontation network is constructed from two aspects of individual behavior anomaly and group anomaly, and the individual abnormal behavior and group abnormal behavior are detected based on reconstruction and prediction methods. Reference [2] designed a method for detecting abnormal behavior through feature extraction based on the characteristics of different application scenarios. Analyze the characteristics and categories of abnormal human behavior; From the perspective of identifying and detecting abnormal behavior, a method for identifying abnormal behavior has been designed to supplement the identification and detection of abnormal behavior. The above traditional target detection methods can only detect the target by extracting the simple features of the target. However, power operation is usually in an open and complex outdoor environment. There are a lot of buildings, infrastructure and tools in the monitoring image. There are many people coming and going, and they are often blocked by obstacles. Different light intensities also bring different colors. The above traditional target detection methods are difficult to extract more complex and comprehensive features to identify the target.

With the rapid development of computer vision, there are better applications in the image recognition field. A computer video is a multidisciplinary scientific field that studies how computers gain a high level of understanding of digital or video images. From a design perspective, he is looking for automatic tasks that can be performed by human vision systems. Under this background, a method of abnormal behavior recognition based on computer vision in power violation monitoring video is studied.

2 Research on the Identification Method of Electric Power Violations

2.1 Key Technologies

Computer vision is the automatic extraction, analysis, and understanding of useful information from a single image or sequence of images. It involves developing theoretical and algorithmic foundations to achieve automatic visual understanding. As a scientific discipline, computer vision is related to the theory behind artificial systems that extract information from images. Image data can take various forms, such as video clips, views from multiple cameras, or multidimensional data from medical scanners. As a technical discipline, computer vision attempts to apply its theories and models to the construction

of computer vision systems [3]. The application range ranges from industrial image processing systems to artificial intelligence research, to computers or robots that can understand the surrounding world. One.

Automatic inspection, e.g. in manufacturing applications;

Assisting humans in carrying out identification tasks, such as the species identification system;

Process control, such as industrial robots;

Detect events, such as visual observation or personnel counting;

Interaction, e.g. as an interactive device input computer human;

Facility or environmental modelling, such as medical image analysis or terrain modelling;

Navigation, such as autonomous vehicle or mobile robot;

Organization information, such as a database for indexing images and image sequences.

The organization of computer vision systems largely depends on their application. Some systems are independent applications that solve specific measurement or detection problems, while others are larger subsystems. This subsystem includes subsystems such as mechanical actuator control, design, database, human-machine interface, etc. The specific implementation of a computer vision system also depends on whether its functions are predefined or whether its parts can be learned or modified during operation. Many features are unique to applications. However, many computer vision systems have typical functions.

(1) Image Acquisition - Digital images are generated by one or more image sensors. In addition to different types of photosensitive cameras, they also include distance sensors, tomography equipment, radar, ultrasonic cameras, etc. Depending on the type of sensor, the obtained image data is a regular 2D image, 3D volume, or image sequence. Pixel values typically correspond to the intensity of light in one or more spectral bands (gray or colored), but can also be related to different physical measurements, such as the depth, absorption, or reflection of sound or electromagnetic waves or nuclear magnetic resonance [4].

(2) Before applying image processing methods to image data to extract specific information, it is usually necessary to process the data to ensure that it meets certain assumptions implicit in the method. For example:

Resample to ensure the correct image coordinate system.

Reduce noise and ensure that sensor noise does not provide incorrect information.

Enhance contrast to ensure relevant information is recognizable.

Scaling space means that the image structure is enhanced at a local appropriate scale.

(3) Extract image features of different complexity levels from image data. A typical example of this function is:

Lines, edges, and burrs.

Local attractions, such as corners, points, or points.

More complex features may be related to texture, shape, or motion.

Detection/segmentation: At a certain point in the processing, determine which pixels or regions of the image are related to subsequent processing.

The segmentation of one or more image regions containing specific objects of interest.

Divide one or more videos into a series of prominent masks per frame while maintaining their temporal semantic continuity.

(4) Identification - In this step, the input is usually a small group of data to obtain the identification results, such as:

Verify that the data meets the assumptions based on the model and specific applications.

Evaluate application specific parameters, such as the pose of the object or the size of the object.

Image recognition fonts categorize the detected objects into different categories.

2.2 Video Image Preprocessing of Power Violation Monitoring

Digital image processing is an image processing process and technology that utilizes digital processing techniques, such as noise reduction, image enhancement, image restoration, image segmentation, etc. In today's society, with the rapid development of information technology, computer processing skills have significantly improved, and mathematics has developed rapidly. The demand for applications in various fields of production and life is also constantly increasing. Images are the most intuitive way to obtain information and are still the focus of research to this day. These factors have led to the rapid development of digital image processing technology [5]. Digital image processing technology is committed to helping people understand the world more accurately through images. The task of image processing is.

2.2.1 Mathematical Morphology Image Processing

Morphology generally refers to an important branch of biology. It is a discipline to study the structure of animals and plants. Its important function is to obtain topology and structure information of animals and plants. Mathematical Morphology is a mathematical processing method based on set theory and topology. As an image processing tool, it processes images to extract valuable information that can describe the shape of the region, such as region edges and bones. The idea achieved is to simplify image data with specific shaped structural components, maintain the foreground shape, fill holes in the target area to extract or eliminate false boundaries caused by noise, and thereby improve detection accuracy. Its four main operations are expansion, corrosion, opening, and closing [6]. These four operations can also be derived and combined into various image processing algorithms, such as image segmentation, edge detection, image enhancement, etc., to process and analyze the shape or structure of the target area in the image.

(1) Bulge: Bulge is to connect the separated parts of the object in the image to make the object more complete. Expansion is to expand or coarsen the target. The calculation formula is as follows:

$$F_1 = A \oplus B = \{C(D)_C \cap A \neq \emptyset\} \tag{1}$$

among,

$$C = \begin{pmatrix} 1\ 0\ 1 \\ 0\ 0\ 0 \\ 1\ 1\ 1 \end{pmatrix} \tag{2}$$

In formula, F_1 represents the expanded power violation monitoring video image; A represents the pixel set of the original video image; C represents the structural elements; $(D)_C$ represents the structural element template; and $\oplus$ represents the expansion operator.

(2) Corrosion: The corrosion operation is to corrode the edges of the binary image. The operation formula is as follows:

$$F_2 = A\Theta B = \{C(D)_C \subseteq A\} \tag{3}$$

In formula, F_2 represents the power violation monitoring video image after corrosion; Θ represents the corrosion operator; $(D)_C$ represents the element set that the structural element C moves over the plane.

The formula (3) indicates that the point set obtained by corrosion B corrosion to A is composed of points in B translated by C and included in A . The erosion operation can remove the boundary points of the target area, making the target boundary shrink from outside to inside. Therefore, the area of the image target area will be smaller after the erosion operation, and some small targets and isolated noise points can be effectively removed. The protrusions around the target after corrosion have been eliminated.

(3) Open operation: the open operation is to first erode and then expand the image to make the image contour smooth, break the narrow neck and eliminate thin protrusions; After this operation, the position and shape of the image will not change. The closing operation is to expand the image before etching, which can make the image contour smooth [7]. The calculation formula is as follows:

$$F_3 = A \circ E = (A\Theta E) \oplus E \tag{4}$$

among,

$$E = \begin{pmatrix} 0\ 0\ 1 \\ 1\ 1\ 0 \\ 1\ 0\ 0 \end{pmatrix} \tag{5}$$

In, F_3 represents the power violation monitoring video image after the operation; E represents the structural element; $\circ$ represents the operation symbol.

(4) Closed operation: closed operation is to expand the image and then corrosion, can make the outline of the image become smooth, but unlike open operations, it can bridge narrow, intermittent, and elongated grooves, eliminate small holes, and fill cracks on the contour line, and can make the position and shape of the image will not change [5]. The operation formula is as follows:

$$F_4 = A \bullet E = (A\Theta \oplus E)\Theta E \tag{6}$$

In formula, F_4 represents the power violation monitoring video image after closed operation; $\bullet$ represents the closed operation symbol.

2.2.2 Image Noise Processing

Image noise processing is the unnecessary or excessive interference information in an image, which seriously affects image quality. Generally, the noise of the image can be eliminated by filtering processing. Common methods include neighborhood average filtering, median filtering, Gaussian filtering, etc. The average neighborhood filtering method can also be called the average value filtering method. The implementation process involves first crossing pixels with all their surrounding pixels, and then replacing the corresponding pixels with the average values in the output image to achieve filter smoothing. The simplest neighborhood average method is to take the same value for all template coefficients. For example, take the template coefficient as 1, which is also called the Box template. The commonly used templates are the following two template types with the size of 3×3 and 5×5 .

$$\frac{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}}{9} \tag{7}$$

$$\frac{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}}{25} \tag{8}$$

The calculation formula of the proposed method is as follows:

$$A(i,j) = \frac{\sum\limits_{(i,j)\in R} A'(i,j)}{N} \tag{9}$$

In formula, $A(i,j)$ represents the value at pixel point (i,j) after the mean filtering of the power violation surveillance video image; $A'(i,j)$ represents the original pixel value at pixel point (i,j); N represents the number of pixels in the template R and R represents the set of pixels in the template.

2.3 Positioning and Tracking of Operators

The positioning and tracking of operators in the electric power violation monitoring video image is the research basis for the analysis and detection of abnormal behavior in the monitoring video. Moving target positioning refers to the use of digital image processing technology to separate the background and moving foreground in the video frame sequence, so as to extract the moving operators in the image. Tracking is a process based on extracting target features, used to track and retrieve motion paths, so as to detect the anomaly of some specific subsequent behaviors. The positioning and tracking of

operators will directly affect the subsequent analysis and detection results of abnormal behaviors of target personnel. Therefore, it is essential to do a good job in the basic work of positioning and tracking.

2.3.1 Target Positioning of Operators

Operator localization is the extraction of moving target characters from video images. Moving target location is the basis of target recognition, tracking and understanding of target behavior, so the accuracy of target location seriously affects the next step of video processing.

The operator positioning can be divided into two types according to the state of the background: one is static target detection, in which case the camera is static relative to its field of view; The other is dynamic target detection. In this case, the camera needs to follow the target movement to obtain a larger field of view. The field of view of the video collected in this paper is static, so the method of static object detection is adopted. This article mainly uses the background difference method in static object detection methods to locate the operator's target.

Background difference method is a very common operator location method. Its basic principle is to obtain the foreground target by subtracting the pixel brightness of the current video frame and the background image at the same position. The specific steps are:

Step 1: Establish a background model as the background image according to the current scene;

Step 2: Update the background image according to the changes of the background model, and differentiate the current video frame from the background image;

Step 3: Perform threshold processing for the differential image to extract moving targets.

With the current frame video image representing $S_t(i, j)$ and the background frame representing $Q_t(i, j)$, the differential image $\Delta T_t(i, j)$ can be represented as:

$$\Delta T_t(i, j) = |S_t(i, j) - Q_t(i, j)| \tag{10}$$

the expression (i, j) represents the spatial coordinates of the pixels. Thresholding of the differential image yields the foreground target $U_t(i, j)$, expressed as:

$$U_t(i, j) = \begin{cases} 1, \Delta T_t(i, j) \geq W, \text{target} \\ 0, \Delta T_t(i, j) < W, \text{ background} \end{cases} \tag{11}$$

In formula, W is the segmentation threshold, when the pixel $\Delta T_k(i, j)$ of the difference image is greater than the threshold, the point is marked as the foreground pixel, otherwise marked as the background pixel. The foreground target obtained from the above formula is the target positioning result of the operator, and the result is used for real-time tracking of the operator's motion process.

2.3.2 Motion Tracking of Operators

The operator's motion tracking is to track and locate the moving target in the sequence image in real time after detecting the moving target person, and obtain the position

coordinate, motion speed, direction and other motion information of the moving person in real time through tracking, laying the foundation for the subsequent motion target behavior analysis. At present, many scientists both domestically and internationally are conducting extensive research and analysis on the problem of moving target tracking. And have also successfully implemented many moving target tracking algorithms. The existing algorithms mainly focus on two analysis ideas: first, detect the moving target from the sequence image, and then locate and identify the target to achieve target tracking; The second is to first establish the target template according to the acquired target prior information, and then match the template in the sequence image to find the moving target and achieve real-time tracking. At present, a lot of research has been done on target tracking methods. According to the image edge, contour, shape, texture, region, etc. and similarity measurement (Euclidean distance, etc.) of the tracked moving target, it can be roughly summarized into the following categories: tracking methods based on active contour, region based, feature-based, and model-based. The quality of the target tracking method lies in whether it can accurately and real-time locate the target position in the sequence image to achieve target tracking. The effective expression of the tracked moving target and the definition of similarity measure are the key to determine the accuracy and stability of the tracking algorithm.

Currently, the main target tracking methods for mobile personnel include model-based tracking, feature-based tracking, region-based tracking, and active contour-based tracking. Among them, the realization process of region based target tracking method is: take the connected area of the moving target detected in the video sequence image as the detection unit, and extract the features of each target region. The extracted features can be color, area, centroid, shape, etc., and match the similarity of the features of the moving region in the adjacent two frames of images, The position of moving target is determined according to the feature similarity of moving region in two images to achieve the tracking of moving target. When calculating the matching degree of the target area in two adjacent frames of the image sequence, the best matching target can also be obtained by combining multiple features of the target. The key of region based target tracking algorithm is to segment the moving target region in the image accurately. Only in this way can the extracted motion features be accurate and effective, the accuracy of subsequent matching be improved, and the tracking effect be improved.

When there are few moving objects in the video images of power violation monitoring, the region based target tracking method has higher tracking accuracy and can track moving objects stably. However, when there are many moving objects in the sequence image, the target area is easy to be occluded, which affects the tracking effect, reduces the tracking accuracy, and easily leads to the loss of the target. However, this method is simple and easy to implement. Selecting specific motion features for different tracking objects can make the tracking effect of moving objects better.

2.4 Staff Motion Feature Extraction

Staff motion feature extraction refers to extracting effective feature data from video sequences to describe human motion state, which is the premise of describing and understanding human behavior in video sequences. Therefore, the rationality of feature selection will directly affect the effect of human behavior recognition. At present,

there are four main methods of features used in human behavior analysis and understanding: appearance shape features, motion features, space-time features, and hybrid features combining motion features and shape features. Among them, appearance shape feature and motion feature are two commonly used features, and spatiotemporal feature has also been widely used in human behavior recognition. Appearance shape features are easy to obtain and relatively stable, and are insensitive to texture changes. Motion features are suitable for behavior description and recognition in the case of long distance or low visibility. With the continuous in-depth study of human behavior analysis and understanding in video sequences, high-level feature information in video image sequences has been widely studied and can more accurately describe human behavior. Considering from the aspect of obtaining human behavior feature data, we can avoid the difficulty of accurately estimating human behavior by extracting the moving human body area in each frame of video sequence, and then using the features of the moving area or contour sequence to describe and identify human behavior [8].

Corners, as an important local feature of images with rotational invariance, do not change with changes in lighting. Extracting key features can reduce the computational complexity of data and shorten processing time without losing image information. This chapter proposes an image feature extraction algorithm based on FAST corners, combining FAST and SIFT algorithms. The system diagram of this algorithm is shown in Fig. 1.

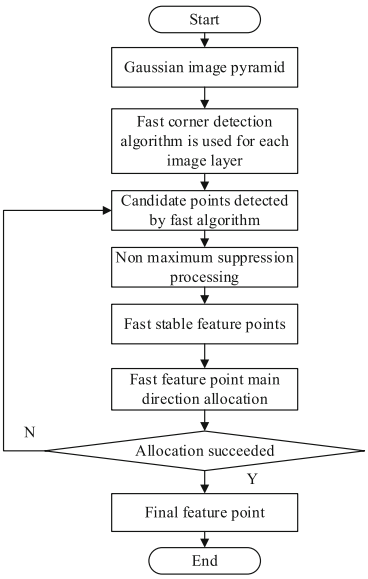

Fig. 1. Image feature extraction based on FAST angles

Aiming at the corner features of FAST, the feature descriptor of FAST is constructed by SIFT algorithm. The specific implementation process is as follows:

Step 1: In order to have rotation invariance, within the neighborhood of any FAST feature point, according to the main direction a of the point, rotate the initial coordinate system by an angle a, and then establish a new coordinate system.

Step 2: Take the feature point as the center, take the 16 × 16 pixel template window as the neighborhood of the feature descriptor, divide the pixel neighborhood window into 16 4 × 4 sub regions, and then establish a gradient histogram in 8 directions in each sub region, and count the amplitude of each direction in the region.

Step 3: Arrange the 8 directional gradient histograms in 4 × 4 sub regions in order of position. Since there are 4x4 sub regions, there are 4 × 4 × 8 = 128 data in total, and finally 128 dimensional feature vectors are formed, as shown in Fig. 3.5.

Step 4: The feature vector is weighted by a Gaussian standard function with a variance of 6 and then normalized. This can eliminate the influence of light changes, so that the feature vectors with light changes do not change to a certain degree, thereby reducing sensitivity to brightness changes.

Step 5: In order to improve the discriminability of corner features of FAST, normalize the feature vector again.

The information represented by a single feature is limited, so LBP algorithm is used here to extract texture features. LBP algorithm was originally used to describe texture features. The basic LBP algorithm works on the 3x3 neighborhood of the central pixel and uses the grayscale value of the central pixel in the neighborhood as a threshold, then compares the grayscale values of the pixels in the 8-neighborhoods with the grayscale values of the central pixel. If the grayscale value of pixels in the neighborhood is greater than the grayscale value of the middle pixel, mark the pixel point value as 1, otherwise mark it as. Finally, you get an 8-bit binary number and convert the sum of binary numbers with different weights of each position into a decimal number, so the decimal number is the LBP value of the neighborhood. Equation (12) is the definition of the basic LBP algorithm.

$$Z(L, b) = \sum_{i=0}^{L-1} f(h_i - h)2^L \tag{12}$$

Among,

$$f(h_i - h) = \begin{cases} 1, h_i - h \geq 0 \\ 0, h_i - h < 0 \end{cases} \tag{13}$$

In the formula, L is the number of domain pixels (the number of sampling points), b represents the radius of the neighborhood, h represents the gray value of the central pixel in the neighborhood, and h_i is the gray value of the i th pixel in the neighborhood.

2.5 Identification of Abnormal Behavior

The feature data can be classified for behavior recognition. For example, we first build our own behavioral feature library by learning typical behaviors, and then extract motion features from real-time behaviors to compare and match with them. There are two key issues: one is to establish a meaningful feature sequence database by analyzing typical

samples; The second is to find an appropriate matching method to compare the similarity between the feature sequence to be detected and the feature sequence database. Up to now, there are many methods for behavior recognition. Here, we use short-term and short-term memory neural networks for recognition [9].

In the traditional neural network algorithm, the nodes in the same hidden layer are not connected. Therefore, the traditional neural network cannot save data information at different times, so it cannot capture the relationship between consecutive video frames. The recurrent neural network can store data information of different times through feedback connection. However, Sepp Hochreiter found the long-term dependence of the recurrent neural network, that is, when learning sequence data, the recurrent neural network will appear gradient disappearance and gradient explosion phenomena, and it is impossible to master the nonlinear connection of a long time span. Long Short Memory Recurrent Neural Network (LSTM) is a very special type of recurrent neural network that can effectively alleviate the problem of RNN gradient disappearance and explosion. The LSTM structure has been carefully designed to avoid long-term dependencies. The LSTM structure includes Forgotten Gates, Input Gates and Output Gates. Unlike the nonlinear transformation of input data in RNN, LSTM structure threshold unit controls information transmission.

Forgotten Gate: The most important element in LSTM structure is the unit state. It learns the dependence of each frame image like a chain, so that information can be transmitted downward. The mathematical formula is described as follows:

$$Y_t = \psi \cdot v(\beta_{t-1}, \alpha_t) + Q \cdot k \tag{14}$$

In the formula, Y_t represents the result of the forgotten gate operation; ψ is the sigmoid function; v is the weight matrix of the forgotten gate joint, with the bottom marker corresponding to the corresponding joint; β_{t-1} is the output of a time node on the hidden layer, α_t is the output of the input layer at all time, and k is the offset value of the forgotten gate.

Entrance Gate: Finally, define how to filter out based on the state of the exit cell. There are two tasks to be done in the input gate: the first input threshold determines what information needs to be updated on the sigmoid level and generates the content that needs to be updated on the Tanh level; The second element updates the device status. The mathematical formula is described as follows:

$$\lambda_t = Y_t \cdot \lambda_{t-1} + O_t \cdot \hat{\lambda}_t \tag{15}$$

In the formula, O_t is the cell state of the input gate at time t ; $\hat{\lambda}_t$ is the temporary state with new candidate values; λ_t represents the new cell state; λ_{t-1} represents the new old cell state.

Output gate: The output gate is the last threshold through which information flows. When information passes through the forgetting and entrance gate, it reaches the exit gate. Its main function is to output the final value. There are two more steps: the first step is to determine the status of the community; Step 2: Hide the output of cells. Enter the cell state in the Tanh function (convert the value to between -1 and 1), and then multiply it by the sigmoid threshold to get the output. The mathematical expression reads as follows:

$$\beta_t = \upsilon_t \cdot \mathrm{T}anh\lambda_t \tag{16}$$

In the formula, υ_t represents the output gate cell state at time t
The basic process of abnormal behavior identification based on LSTM is as follows:

(1) As the input of LSTM at time t, the behavior characteristics output the calculation results.
(2) The input gate, exit gate, and forgetting gate are used to control the gate, process the previous input data and input time t, and check the output and unit status of time t. Output to the next LSTM unit up to the last layer.
(3) The output layer neurons are used to process the information learned by the last layer LSTM unit.
(4) The data loss function of the calculated output layer is updated by the gradient descent method until the conditions are met.

Based on the above research, a computer-aided method for identifying abnormal behaviors in videos for monitoring power breaches was completed.

3 Method Test

3.1 Video Image Acquisition for Action Monitoring of Power Maintenance Personnel

Take the image of maintenance personnel's work site collected by the on-site video monitoring system in the past as an example to test the method. A total of 1200 video images were obtained, of which 600 were abnormal and the rest 600 were normal. Some of the images are shown in Fig. 2.

Fig. 2. Example of the action monitoring video image of the electric power maintenance personnel

To test the feasibility of this algorithm, the experimental platform used in this article is a personal computer with a CPU frequency of 2.53 GHz and an NVIDIA ForceGTX1070 graphics card with 4 GB of memory and 1 GB of video memory. This algorithm was developed by MATLAB R2016A. All processed images are stored in an HDFS database with an image size of 256 x 256.

3.2 Positioning of Operators

For example, a behavioral motion image uses differential background method to achieve target localization, as shown in Fig. 3.

Fig. 3. Target positioning

3.3 Feature Vectors

Two features are extracted for each action monitoring video image of the power maintenance personnel, and some of the results are shown in Table 1.

Table 1. Feature vector extraction table

Image	Corner/°	Texture/dpi
1	1.2623	15.8754
2	2.154	14.5365
3	0.231	16.8754
4	1.2645	14.7452
5	2.1225	11.8765
6	0.0862	13.8645
7	1.0861	11.7452
8	0.5521	12.8422
9	0.4553	13.8645
10	0.4222	15.5122
11	1.7452	13.5312
12	1.3011	14.4222
13	1.8952	13.8654
14	1.0252	15.2102

3.4 Image Recognition Results

For the detection results, calculate the Kappa coefficient between the measured results and the measured results, and judge the accuracy of the method identification. The results are shown in Fig. 4.

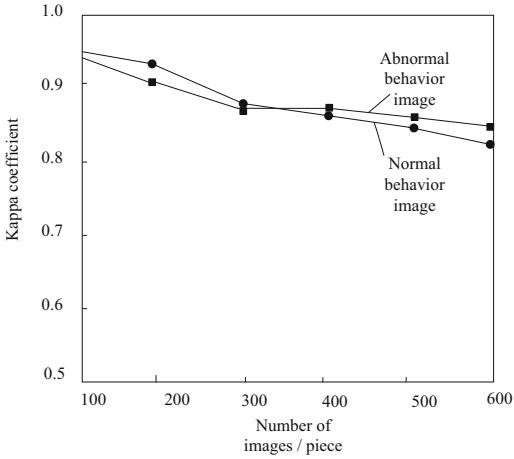

Fig. 4. The Kappa coefficient

As can be seen from Fig. 4, under the Kappa coefficient of the studied method, the K a p p a coefficient between the measured results decreases with the number of video images, but it remains above 80%, proving the accuracy of the identification method.

4 Conclusion

Whether or not the construction behaviour of the energy conservation industry is regulated, it is related to the personal safety of workers and is crucial for the development of the energy industry. The behavioral analysis of employees is an important research direction in the field of modern deep learning. Currently, video surveillance is mainly implemented in maintenance projects in the energy industry. Through the analysis of these videos, we can obtain relevant information about the construction of staff. In view of the illegal operation of electric power maintenance personnel, the illegal behavior of electric power maintenance site is identified from the perspective of computer vision. The results are as follows:

1) Data enhancement and image enhancement technology can increase the number of pictures and improve the training effect to a certain extent;
2) The proposed algorithm can effectively identify the types of illegal operations in power operation scenarios and specific scenarios;
3) The power operation environment is complex, and there are often people blocking and different distances that affect the detection effect. In the future, the applicability of the model can be improved by continuing to expand training samples and modify training models.

Aknowledgement. Science and Technology Project of China Southern Power Grid Co., Ltd. (GZHKJXM20200058).

References

1. Shen X., Li, C., Li, H.: Overview on video abnormal behavior detection of GAN via human density. Comput. Eng. Appl. **58**(7), 21–30 (2022)
2. Xiaoping, Z., Jiahui, J., Li, W., et al.: Overview of video based human abnormal behavior recognition and detection methods. Contr. Decision **37**(1), 14–27 (2022)
3. Daradkeh, Y.I., Tvoroshenko, I., Gorokhovatskyi, V., et al.: Development of effective methods for structural image recognition using the principles of data granulation and apparatus of fuzzy logic. IEEE Access **9**(99), 13417–13428 (2021)
4. Liu, S., Wang, S., Liu, X., Gandomi, A.H., Daneshmand, M., Muhammad, K.: Victor hugo c de albuquerque, human memory update strategy: a multi-layer template update mechanism for remote visual monitoring. IEEE Trans. Multimed. **23**, 2188–2198 (2021)
5. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective template update mechanism in visual tracking with background clutter. Neurocomputing **458**, 615–625 (2021)
6. Shuai, L., Shuai, W., Xinyu, L., et al.: Fuzzy Detection aided real-time and robust visual tracking under complex environments. IEEE Trans. Fuzzy Syst. **29**(1), 90–102 (2021)
7. Gao, P., Zhao, D., Chen, X.: Multi-dimensional data modelling of video image action recognition and motion capture in deep learning framework. IET Image Proc. **14**(7), 1257–1264 (2020)
8. Shatalin, R.A., Fidelman, V.R., Ovchinnikov, P.E.: Incremental learning of an abnormal behavior detection algorithm based on principal components. Comput. Opt. Opt. **44**(3), 476–481 (2020)
9. Zerkouk, M., Chikhaoui, B.: Spatio-temporal abnormal behavior prediction in elderly persons using deep learning models. Sensors **20**(8), 2359 (2020)