Intelligent Multi-Drive Inspection Technology for Water Environment of Cable Pipe Gallery Based on Random Forest Algorithm

Lei Wang1, Yuan Cheng1, XiaoJun Li1, Bo Qin1
1State Grid Datong power supply company, Datong, Shanxi, China, 037000

*Corresponding author e-mail: gaoyantougao@163.com

Abstract. Image technology is widely used in intelligent applications. Based on the intelligent multi-drive patrol inspection of the water environment of cable duct corridors, the original technical methods and corresponding algorithms cannot be effectively solved. This paper mainly studies the intelligent multi-drive patrol inspection technology for the water environment of cable duct corridor based on random forest algorithm. In this paper, a feature that is insensitive to changes in illumination is designed and used for image change detection. At the same time, the Haar-like feature is improved according to this feature. The improved Haar-like feature and random forest calculation are used to detect the change area of the image. The experiment in this paper found that the cable fire of underground comprehensive pipe corridor burned more violently during 200 s-600 s. This stage only accounted for 22.3% of the burning time, but contributed 73.4% of the mass loss. The experimental results in this paper show that the intelligent multi-drive patrol inspection technology for the water environment of cable duct corridors based on the random forest algorithm is in line with the actual application standards and has important significance in practical applications.

Keywords: Random Forest Algorithm, Water Environment of Cable Duct Corridor, Multi-Drive Inspection, Image Processing

1. Introduction

There are some new problems in the cable duct Gallery, which can not be solved effectively by the existing theory and technology, and the original theory and technology need to be improved[1]. The existing video surveillance technology has been quite mature, and the target recognition and detection technology based on video usually needs to establish a stable and reliable background model[2-3]. In
practical applications, a camera often needs to monitor multiple targets around, and the targets may be in different directions[4-5]. When the camera turns to the preset point corresponding to the target, the background model reserved at the preset point last time is often different from the image at the current time of the scene. For example, the change in the image caused by the change of illumination conditions causes the detection effect of the algorithm based on continuous video frames is not good[6].

Intelligent target analysis technology contains a lot of knowledge, including image processing, pattern recognition, artificial intelligence, automatic control and so on. It is widely used in military weapon guidance, safety detection, public place monitoring and intelligent transportation[7], which in turn expands the research on target analysis methods[8-9]. Generally, the whole system is required to adapt to the problems of illumination change, sensor noise, occlusion, etc., and accurately detect and track targets in different backgrounds, and deeply analyze their behavior types[10].

In this paper, a new method based on image detection and recognition is proposed by using random forest, which is a machine learning technology. Different from other methods which only stay at the level of image pixel processing, this paper gives the model the ability of classification and discrimination of objects, and uses image feature information of deep mining, supervised learning architecture and a variety of image technology.

2. Intelligent Multi-Drive Inspection Technology for Water Environment of Cable Duct Corridor

2.1. Random Forest Algorithm

In the case of a given test sample, first find the K nearest neighbor of each test sample, still adopt the repeated sampling method in the K nearest neighbor, and add the extracted neighbor sample to Bag. The sample sampling consists of two parts: one part is the sample taken by the traditional Bagging method, and the other part is a certain proportion of the sample from the neighbor set of the test sample. This sampling method not only ensures the diversity of random forest samples, but also incorporates the characteristics of the nearest neighbor classification method.

In the process of building a random forest, the decision tree adopts CART tree, and each random forest has 100 decision trees. The CART tree adopts the form of complete growth without pruning, and the construction process includes sample sampling, feature space sampling and optimal split attribute selection. Feature sampling adopts random sampling method, assuming that the training set has a total of F attributes. Generally, \( f = \left\lceil \log_2 F + 1 \right\rceil \) attributes are selected as the split attribute candidate space. Generally, \( f \) is much smaller than \( F \). When choosing the best split point. The CART tree uses the Gini index (Gini) as the splitting criterion. The principle of Gini is to partition the training data set, as far as possible that the samples in the same partition belong to the same class. The definition of Gini index is shown in formula (1).

\[
Gini(D) = 1 - \sum_{i=1}^{m} p_i^2
\]  

(1)
Among them, $P_i = \frac{|C_i D|}{|D|}$ is the probability that the sample in D belongs to the $C_i$ class. In the data set, attributes are divided into general discrete type and continuity. For both, the Gini index considers binary division.

Assuming that A (the value of $\{a_1, a_2, \ldots, a_r\}$) is a discrete attribute, there are 2v-2 A-based methods for dividing the D data set into two partitions. For example, when the binary division of A divides the data set D into $D_1$ and $D_2$, the calculation method of the Gini index of the division D based on the attribute A is shown in formula (2):

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2)$$

(2)

3. Experimental Design of Intelligent Multi-Drive Inspection Technology for Water Environment

3.1. Experimental Equipment

During the experiment, LCM300 strain-type tensile pressure sensor, self-made weighing instrument, Aglient 34970A, K-type armored thermocouple, high-definition network camera, SLR camera and other experimental equipment were used. Aglient 34970A data collector was used to record the temperature data during the experiment. The instrument comes with measurement and analysis software, which is convenient for test configuration, and can display and record data in real time.

3.2. Experimental Design

According to the similarity theory, this paper establishes a 1:3.6 size model of a comprehensive pipe corridor. The model is mainly composed of barrel concrete pipe corridor, temperature information collection system, quality information collection system, image information collection system, computer and so on. The pipe corridor is 10 m long, with an inner diameter of 1.5 m and an outer diameter of 1.8 m. Since the cable compartment is located at 1/4 above the side of the pipe corridor and is divided by a solid wall, the thermocouple only needs to be arranged in the cable compartment, and no other location is required. The cable fire experiment platform of underground comprehensive pipe gallery is shown in Figure 1.

Table 1. Parameter setting of reduced size model of underground comprehensive pipe corridor

| Name                                           | Outer diameter | Length | Material | Area ratio | Proportion of cable compartment area |
|------------------------------------------------|----------------|--------|----------|------------|--------------------------------------|
| Reduced size model of underground comprehensive pipe corridor | 1.8 m          | 10 m   | Concrete | 1:3.6      | 1:4                                  |

4. Discussion and Analysis of Water Environment Intelligent Multi-Drive Inspection Technology
4.1. Training and Testing Framework Analysis

The training sample map contains the target image and the background image, and records the center value of the target image that has been manually marked. First, the training sample graph is decomposed into blocks of uniform size, such as the method used in the Hough Forest, and the image block is preprocessed, which mainly includes three aspects: extracting the feature information of the image block, including the Haar direction gradient descriptor, Lab color space information, the first and second derivatives of pixels; mark the category of the image block, if the information of the target area is marked as a positive sample, otherwise it is processed as a negative sample; record the center of each image block to the target. The vector value of the center. Then, according to the feature information of the image block, an incremental random forest is constructed, and each node randomly selects the category label uncertainty or center offset value as the splitting criterion, and continuously splits to generate child nodes. Finally, each leaf node stores the category information and center offset of the image block, and retains the best segmentation function information in each branch node. When the number of image blocks of the node is less than the threshold or reaches the maximum depth of the forest, the construction of the forest model is completed. In the Meanshift phase of target detection, this paper does not use the maximum posterior probability greedy algorithm and the non-maximum suppression algorithm of the frame, but finds the extreme point of the target area density estimation, and can also achieve accurate target positioning.

4.2. Analysis of Inspection Results

Put the intelligent inspection into use in a distribution room. Set up automatic inspection tasks. 2 inspections per day, set in the morning and afternoon of peak electricity consumption respectively. Among them, it takes about 2 hours for the robot to complete a patrol inspection, and there are 1,265 equipment monitoring points on site. A six-month operation was carried out on site. During the inspection, the robot passed a remote monitoring platform installed in the main control building of the substation. It can automatically identify and analyze the indoor indicator light status, related instrument (pointer type, digital display) readings, switch status, etc.; at the same time, it can detect the heating state of the device in real time. The fault database and historical database can be set in the background, which can analyze the type of fault in real time, view historical data, and realize automatic sound and light alarm through the background software.

The quality change data collected by the LCM300 strain-type tensile pressure sensor in the experiment was processed to obtain the quality change of the cable material, as shown in Figure 1. In the three cases, the mass loss curve is as follows: 1# fire source combustibles loses about 395g at most, and 3# fire source combustibles loses about 155g at least. It can be seen from the graph trend that the speed of mass loss generally slows first and then accelerates, and the speed decreases again after reaching the inflection point, and finally the speed is extremely low close to zero.
4.3. Performance Comparison on Standard Data Sets

Based on the experimental data of semi-supervised data adjusted based on three standard UCI data, the proportion of labeled samples in all training data gradually increases. The accuracy of the traditional random forest is only about 55% when the labeling rate is 10%, mainly because the training data samples are small, resulting in a low generalization ability of the obtained classification model. Due to the weak generalization ability of the algorithm, after introducing semi-supervised and data editing methods, the algorithm can be improved to a certain extent, but the generalization ability of these two algorithms cannot be improved well. As the proportion of labeled samples increases, the generalization ability of the traditional random forest algorithm gradually improves. When the labeling rate reaches 40.2%, the generalization ability of the classification model reaches about 76.8%. Basically, the data can be well planned. Together. The accuracy of the SRF and DSRF algorithms has also gradually improved, and can reach about 84.2%, indicating that the classification models obtained by these two algorithms have good generalization capabilities. It can be concluded that the DSRF algorithm can obtain a classification model with good generalization ability when the labeling rate reaches 30% to 40%. At the same time, it is proved that the semi-supervised random forest algorithm with data editing is better than the semi-supervised random forest algorithm.

The purpose of this experiment is to determine whether the classification performance of the DSRF algorithm can fit the NIR spectral data well as the proportion of labeled samples in the training data set gradually increases. The training data in this experiment adopts near-infrared spectroscopy data, through PCA method to reduce the dimension, select the main component with the cumulative contribution rate of the main component greater than 99% as the attribute set of the training data, and the quality grade data corresponding to the cigarette as the target set of the training data. The experimental results show that with the gradual increase of unlabeled sample data in the training sample data, the accuracy of the DSRF algorithm can be improved very well. When the ratio increases to 30%, the classifier performance reaches 71%, indicating that in this case Next, a classifier with better classification performance can be trained. As the ratio of unlabeled sample data increases, the performance of the classifier tends to be stable. When the proportion of unlabeled samples reaches 40%, it reaches 74%, and the classification performance is better.

The results of the experiment are considered comprehensively. It can be obtained that when the labeled data samples reach 30% to 40% of the total number of samples, the algorithm in this paper can
use the training data to establish a better performance classification model. Therefore, DSRF can be used to establish a near-infrared spectral model of cigarette quality grades, and then predict the unknown samples, which can greatly reduce the evaluation time of the inspection quality assessment and improve the evaluation efficiency.

5. Conclusion

In this paper, the chi-square test is used as a measure of the dependence between the two class labels. Through the weight calculation and layer-by-layer screening of a large number of class label sets, high-weight and high-differentiated class label sets are selected. It avoids the problem of inter-class correlation that is often ignored in existing multi-class classification algorithms. This paper proposes a multi-class standard random forest classification method. According to the characteristics of the correlation between the multi-category data sets, the multi-category classification problem is divided into several relatively independent sub-problems, and the multi-category problem conversion algorithm and the improved traditional random forest algorithm are used in a step-by-step manner. Solve the sub-problems and then integrate to obtain the optimal solution. Finally, the superiority of the algorithm is proved through experiments.

References

[1] Polan D , Brady S , Kaufman R . Tissue segmentation of computed tomography images using a Random Forest algorithm: a feasibility study[J]. Medical Physics, 2016, 43(6):3330-3331.

[2] Zhang Q , Sun X , Feng K , et al. Predicting citrullination sites in protein sequences using mRMR method and random forest algorithm[J]. Comb Chem High Throughput Screen, 2017, 20(2):164-173.

[3] Liu C , Deng N , Wang J T L , et al. Predicting Solar Flares Using SDO/HMI Vector Magnetic Data Product and Random Forest Algorithm[J]. Astrophysical Journal, 2017, 843(2):13104.

[4] Xu Y , Zhang J , Gong X , et al. A method of real-time traffic classification in secure access of the power enterprise based on improved random forest algorithm[J]. Power System Protection and Control, 2016, 44(24):82-89.

[5] Joshua A , Sugumaran V . Fault Diagnosis for Wind Turbine Blade through Vibration Signals Using Statistical Features and Random Forest Algorithm[J]. International Journal of Pharmacy and Technology, 2017, 9(1):28684-28696.

[6] Xiaonan W , Shihong D , Chen-Chieh F , et al. Interpreting the Fuzzy Semantics of Natural-Language Spatial Relation Terms with the Fuzzy Random Forest Algorithm[J]. ISPRS International Journal of Geo-Information, 2018, 7(2):58.

[7] Wang Y , Wu G , Deng L , et al. Prediction of aboveground grassland biomass on the Loess Plateau, China, using a random forest algorithm[J]. Rep, 2017, 7(1):6940.

[8] Canovas-Garcia F , Alonso-Sarria F , Gomariz-Castillo F , et al. Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery[J]. Computers & geoences, 2017, 103(JUN.):1-11.
[9] Byeon H. Is the Random Forest Algorithm Suitable for Predicting Parkinson's Disease with Mild Cognitive Impairment out of Parkinson's Disease with Normal Cognition?[J]. International Journal of Environmental Research and Public Health, 2020, 17(7):2594.

[10] Osco L P, Ramos A P M, Pereira D R, et al. Predicting Canopy Nitrogen Content in Citrus-Trees Using Random Forest Algorithm Associated to Spectral Vegetation Indices from UAV-Imagery[J]. Remote Sensing, 2019, 11(24):2925.