Research on Detection Method of Internal Defects of Metal Materials Based on Computer Vision

Li Zhang(✉) and Ying Zhao

Nanchang Institute of Science and Technology, Nanchang 330108, China

Abstract. In order to improve the accuracy of detecting internal defects of metal materials, a method of detecting internal defects of metal materials is designed based on computer vision. First, computer vision methods are used to collect internal images of metal materials, and then the images are processed and image features are extracted. Finally, accurate detection of internal defects of metal materials was carried out. The experimental results show that, compared with the traditional detection methods, the detection accuracy of the metal material internal defect detection method based on computer vision is high, and the detection time is short, which proves that it has high practical application significance.

Keywords: Computer vision · Metal materials · Internal defect detection · Image feature extraction

1 Introduction

Metal materials generally refer to pure metals or alloys in industrial applications. There are about 70 kinds of pure metals in nature, of which the common ones are iron, copper, aluminum, tin, nickel, gold, silver, lead, zinc, etc. But the alloy often refers to two or more kinds of metals or the combination of metal and non-metal materials and has metal characteristics. At present, a variety of metal materials have become an important material basis for the development of human society. However, due to the influence of the production process, metal materials are prone to internal defects in the manufacturing process. Therefore, it is particularly important to design an effective detection method for internal defects of metal materials.

Computer vision is also often called machine vision. The occurrence and development of this technology has a history of decades. Computer vision is a subject that studies the observation of the surrounding world through image or video data. It mainly uses the image or video taken by the camera as the original data to extract the things that can be observed in the image or video. The problem to be solved by this subject is very similar to the visual perception function of humans observing the world through eyes. Machine vision emphasizes the processing of vision problems with a system composed of machines, while computer vision emphasizes the processing of visual computing problems. The core component is a computer with powerful computing capabilities.

W. Fu et al. (Eds.): ICMTEL 2021, LNICST 387, pp. 27–38, 2021.
https://doi.org/10.1007/978-3-030-82562-1_3

Therefore, this study applied computer vision to metal material internal defect detection process, and designed a metal material internal defect detection method based on computer vision. In this method, the internal images of metal materials are collected by computer vision method, and then the images are processed and the image features are extracted. Finally, the internal defects of metal materials are accurately detected. The experimental results also prove that this method has the advantages of high detection accuracy and detection time.

2 Computer Vision-Based Internal Defect Detection Process of Metal Materials

In the process of metal material internal defect detection based on computer vision, image acquisition technology is used to obtain complete metal material image data, and the image data is associated with metal material information to form a complete metal material defect data record. Then, image processing and pattern recognition technology are used to analyze the image, realize the automatic recognition of the internal defects of metal materials, and record the location information of defects, so as to facilitate the inspection and maintenance of relevant staff [1]. The core part of the system is related technologies and methods in the field of computer vision. The overall process is shown in Fig. 1.

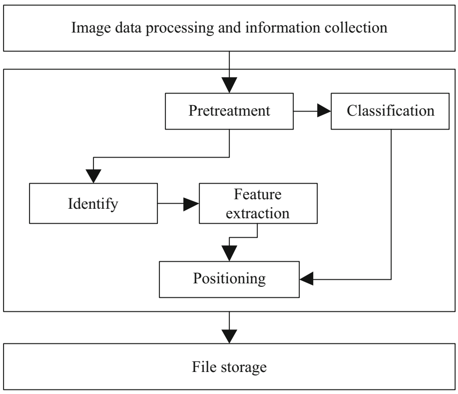

Fig. 1. The internal defect detection process of metal materials based on computer vision

Due to the uneven illumination, occlusion and interference target, the railway image collected in metal materials has poor image quality. Therefore, it is necessary to preprocess the original image to facilitate the subsequent effective image segmentation, feature extraction and recognition [2]. Secondly, in order to automatically identify the state of metal materials through images, it is necessary to extract features that can effectively describe the state of metal materials. There are many methods of image feature extraction. Considering the characteristics of railway fastener itself, the gradient direction histogram feature and local binary mode are selected to describe the metal material image. Finally, according to the known sample features extracted, combined with support vector machine (SVM), the classifier is trained to detect the internal defect state of metal materials.

2.1 Internal Image Processing of Metal Materials

Firstly, the detected image is divided into blocks, and the block size is determined according to the size distribution of the detected object. Firstly, the "block size" must be larger than the largest metal material size in the "image" to ensure the integrity of the detection target in the region containing some blocks; Secondly, after the image is divided into small pieces, it must be rich enough. And the small pieces must contain the metal material in the picture.

Based on the above analysis, the block is set to be square, and the side length is twice the edge of the largest metal material. The block diagram is shown in Fig. 2.

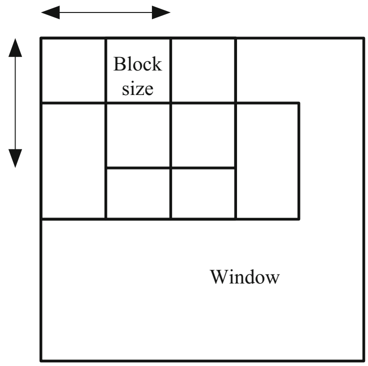

Fig. 2. Block diagram

Then use the back propagation algorithm to pre-train the image, and the cost function formula is as follows:

$$J_d = f + \beta \sum_i w(\rho) \tag{1}$$

In formula (1), J_d is the cost function of a layer of back-propagation network, β is the sparsity cost function, ρ is the activation degree of the training image, $\sum_i w$ is the number of hidden neurons, and f is the expected activation degree of the image. Then, it uses deep learning algorithm to calculate the cost function layer by layer, and carries out unsupervised learning. The process is shown in the following formula:

$$J(w, d) = \frac{q}{m} \sum_f q(x_i, y_i) \tag{2}$$

In formula (2), x_i, y_i represents the training image sample set, $J(w, d)$ represents the hidden layer output parameter, m represents the weight attenuation parameter, q

represents the network layer number, and $\sum_{f} q$ represents the node number of the q layer.

The image is pre-trained through the above process, and the image is processed on this basis.

In the process of image shooting and transmission, due to the influence of various interference factors in the scene, it is easy to lead to the actual image can not meet the needs of viewing, recognition and understanding [3]. The main purpose of image enhancement technology is to purposefully enhance the features of the interested parts of the image according to different needs of people, while suppressing the unnecessary features, and expanding the differences between different objects in the image. Its purpose is to improve the visual effect of the image, improve the clarity, or transform the image into a form more suitable for analysis and processing, and contain as much useful information as possible.

Image enhancement technology is divided into spatial domain processing method and frequency domain processing method according to the different space in the enhancement process. The spatial domain processing method is to directly process the values of the pixels of the collected image, based on the gray-scale mapping transformation, which mainly includes gray-scale correction, image smoothing and image sharpening. The frequency domain processing method is to process the coefficients through a certain transformation, and finally change back to the original space. In order to improve the quality of the original image to facilitate subsequent defect location and recognition, it is necessary to first perform certain enhancement processing on the images collected on the actual railway line.

Based on the analysis of a large number of metal material images, histogram equalization is used to process the original image [4]. Histogram equalization is to transform the original gray distribution of the input image into a gray distribution image with approximately the same number of pixels in each gray level. In the new image obtained by histogram equalization, pixels occupy as many gray levels as possible and are approximately evenly distributed, and the contrast and dynamic range of the image are increased.

The gray distribution of the image is generally represented by a histogram. The histogram describes the number of pixels of each gray level in the image or the frequency of the gray level pixels, and reflects the gray level distribution of the image. The histogram of the image whose gray level is in the range of [0, 1] is a discrete function, and its form is as follows:

$$h(r_k) = n_k \tag{3}$$

In formula (3), r_k represents the gray level of level $_k$, and n_k represents the number of pixels with gray level r_k in the image.

Divide the total number of pixels contained in the image (indicated by n) by the number of pixels in each gray level to get a normalized histogram:

$$h'(r_k) = \frac{n_k}{n} \tag{4}$$

In order to facilitate the analysis of histogram equalization principle, firstly, it is assumed that the gray value range of the image is continuously distributed on the interval

[0, 1]. According to the properties of the probability density function, there are:

$$\int_0^1 p(x)dx = 1 \tag{5}$$

Assuming that the probability density of the image before transformation is $p_r(r)$, the probability density function of the image after transformation is $p_s(s)$, and the transformation function (that is, the gray-scale mapping relationship) is $s = f(r)$, we can get from the probability theory:

$$p_s(s) = p_r(r) \cdot \frac{dr}{ds} \tag{6}$$

For discrete gray values, the corresponding transformation formula is as follows:

$$D_B = f(D_A) = \frac{D_{aax}}{N} \sum_{D_1}^{j=0} n_i \tag{7}$$

In formula (7), n_i represents the number of pixels of the $_i$ grayscale; N represents the total number of pixels of the image; f represents the transformation function.

The transform function is a monotone increasing function to ensure that the gray level of the image before and after the transformation will not be reversed. For the original image, the above transformation formula can be used to equalize it, and the original gray level can be mapped to a new gray level, so as to obtain an approximately uniform histogram.

On this basis, image edge detection is carried out. Edge feature is one of the basic features of image, which often contains a lot of useful information. Image edge refers to the discontinuity of image local characteristics, such as the mutation of pixel value and texture structure [5]. Edge exists widely between different objects, which is an important basis for image segmentation. At the same time, the edge is also the symbol of contour and position. In the process of texture feature extraction, shape feature extraction and image understanding and recognition, it is often necessary to detect the edge of the image first. The formula is as follows:

$$S(x, y) = [f(x + n, y) - f(x, y)]^2 \tag{8}$$

In formula (8), $f(x, y)$ is the gray distribution function of the edge image to be detected, and $(x + n, y)$ is the gradient value of the image edge.

Image binarization is a typical threshold segmentation method. The specific processing method is: set an appropriate threshold for an input image with multiple gray levels, and set the gray values of all pixels in the image one by one Comparing with this threshold, pixels larger than this threshold are all set to 1, and pixels smaller than the threshold are set to 0.

Assuming that the gray distribution function of the original image is $f(x, y)$, and the image distribution function after binarization is $g(x, y)$, then:

$$g(x, y) = \begin{cases} 0, f(x, y) < \text{Threshold} \\ 1, f(x, y) \geq \text{Threshold} \end{cases} \tag{9}$$

Among them, the threshold is the scale to distinguish the target from the background, and reasonable selection of threshold is the key to achieve satisfactory results of binarization processing. It is required to retain as much useful information as possible and reduce the interference of noise and background [6].

In the process of image binarization, there are usually two strategies for threshold selection. One is the global threshold method, that is, only one threshold is used in the binarization process, which is suitable for the image with obvious distinction between the target and the background, and the histogram presents bimodal characteristics; the other is the local threshold method, which uses different segmentation thresholds for different pixel regions, which belongs to the dynamic adaptive binarization processing method, which is suitable for the background gray change or serious noise in the image The sound quality is poor.

2.2 Feature Extraction of Internal Image of Metal

The first step in automatic image recognition is to extract useful data and information from the image, that is, image features. Through the training process, the computer "understands" these features, so as to achieve the purpose of classifying and identifying the input unknown image. This requires that the features we choose not only describe images well, but also distinguish images belonging to different categories.

Each image contains features that can be distinguished from other types of images, some of which can be intuitively felt by humans, and some of the features need to be obtained by transforming or processing the original image data, such as matrix, Histogram etc. In general, feature extraction should be analyzed by specific issues, and the evaluation criteria are also subjective. In general, the feature extraction process should be relatively easy, and it should be weighed against the feature classification ability. In addition, the extracted features should be insensitive to noise and irrelevant transformations. For example, in the recognition of the license plate number, the shooting angle of each image may be inconsistent. But we are only concerned about the above numbers, so we need to obtain feature descriptors that are not sensitive to geometric distortion and deformation, so as to obtain features that are not subject to rotation or projection distortion.

The local binary mode is an operator used for texture description, which can measure and extract the texture information of local regions in grayscale images [7]. The local binary mode texture unit is shown in Fig. 3.

When using the local binary pattern to extract the texture features of an image, firstly, it is necessary to calculate the binary relationship between the gray value of each pixel in the image and the local neighborhood pixels on the gray level [8]; secondly, the obtained binary relationship is weighted according to certain rules to obtain the local binary pattern; finally, the histogram sequence of multiple regions is used as the binary mode of the whole image, and the calculation is introduced below The mathematical analysis of sub.

For the local texture T of an image, it can be regarded as the joint distribution density of the gray level in the local area of the image

$$T = t\left(g_c, g_0, \cdots, g_{p-1}\right) \tag{10}$$

6	5	2
7	6	1
9	8	7

(a)

1	0	0
1		0
1	1	1

(b)

8	4	2
16		1
32	64	128

(c)

Fig. 3. Local binary pattern texture unit

In formula (10), g_c represents the gray value of the central pixel in the local neighborhood of the image. Without losing the texture information of the local area, subtracting the gray value of the center pixel of the neighborhood from the neighborhood point, the calculation expression is:

$$T = t\left(g_c, g_0 - g_c, \cdots, g_{p-1} - g_c\right) \quad (11)$$

In formula (11), g_c represents the center pixel of the neighborhood.

Performing differential operations on all pixels in the neighborhood can make larger and smaller gray levels uniform (especially under uneven lighting conditions), with uniform brightness invariance, and translation invariance characteristics within the gray range [9]. The difference between the central pixel and the neighboring pixel is independent of the gray value of the central pixel, so the above formula is converted to:

$$T = t(g_c)t\left(g_0 - g_c, \cdots, g_{p-1} - g_c\right) \quad (12)$$

Finally, a coding value is used to represent the local binary pattern of texture units, namely texture characteristics. The correlation coefficient is defined as: 1

$$r_{i,j} = \frac{E[(F_i - F_j)(F_{i2} - F_{j2})]}{\sigma_i \sigma_j} \tag{13}$$

In formula (13), σ_i and σ_j represent the standard deviations of features F and F_j respectively. The closer $r_{i,j}$ and 1, the higher the correlation between the two features.

In addition, feature stability is also an important factor affecting feature selection. Therefore, the following formula is used to measure the stability of the target feature, which is defined as follows:

$$\rho_i = \frac{E[F_{iz}^{(w)z}] - E^2[F_{i2}^{(w)}]}{E[F_i^{(w)2}]} \tag{14}$$

In formula (14), $E[F_{iz}^{(w)z}]$ and $E^2[F_{i2}^{(w)}]$ represent the mean and mean square values of the current characteristic F_i respectively. w represents the label of classification type. The smaller ρ_i is, the more stable the feature is.

The method of thresholding the neighborhood elements with the gray value of the central unit as the threshold can accurately reflect the change characteristics of the gray values of the elements in the neighborhood, and has certain gray invariance and stable output.

Because when the neighborhood of the image rotates, the neighborhood elements will move in the circumferential direction around the center pixel of the unit. When calculating the encoding of the neighborhood, the default will start from the element to the right of the center pixel of the unit as the first element, and then rotate counterclockwise to get other elements in turn, and the corresponding weights will increase in turn. Big. As a result, for the same texture unit, the coded value obtained when the rotation changes occurs often, and the coded value will be different with different rotation angles.

In order to make the local binary pattern be able to describe the element with rotation change effectively, it is necessary to improve the descriptor to make it rotation invariant. The specific processing method is as follows: for the neighborhood unit, a group of coding sequences are obtained according to the original thresholding method, and then the coding sequence is circularly shifted to minimize the corresponding decimal result value, and the minimum value is taken as the coding value of the unit [10]. In this way, no matter how much the rotation angle changes, a coding mode will eventually appear.

2.3 Extraction of Internal Defects of Metallic Materials Based on Computer Vision

On the basis of the above image preprocessing, the edge segmentation line is combined with the image threshold region of interest, and the image is rotated and

$$\text{translated}\, P(X) = A(X) + b \tag{15}$$

In formula (15), X represents the spatial position of the image pixel, A represents the rotation matrix of the two images, and b represents the translation vector.

Among them, the matrix A needs to meet the constraints:

$$A^t A = I, \det I = 1 \tag{16}$$

In formula (16), A^t represents the rotation value of matrix A and I represents the identity matrix.

After optimizing the spatial transformation, the similarity function is defined to measure the similarity between the image and the reference image, and the similarity measurement function is optimized by continuously adjusting the transformation parameters. The calculation formula is as follows:

$$Z(a, b) = \frac{h(a, b)}{\sum_{i=1} k(a, b)} \tag{17}$$

In formula (17), $h(a, b)$ represents the gray value corresponding to the image, and k represents the image joint similarity coefficient.

Image registration defines a three-dimensional coordinate system. The x axis scans along the image row direction, the y axis scans along the column direction, and the z axis scans along the aspect of layer growth. The expression of the image to be registered is:

$$C = \frac{\left(C_x, C_y, C_z\right)}{\sum_{y} g(x, y, z)} \tag{18}$$

In formula (18), $g(x, y, z)$ represents the gray level of the pixel whose position is (x, y, z) in the image.

On this basis, the position coordinates of the reference image are discretized sufficiently to make the reference image cover the whole style element space to avoid collapse. The expression is as follows:

$$u = \frac{z}{\exp(v/m)} \tag{19}$$

In formula (19), v represents an ordinary image, m represents the content feature of the image, and z represents the spatial position of the style element.

Finally, combined with the background difference method, the possible defects are initially extracted, and then the connected regions are used to extract edge defects based on the area, roundness, and rectangularity. The algorithm flow of edge micro defect extraction is shown in Fig. 4.

In this way, the internal defect detection of metal materials based on computer vision is completed through the above process.

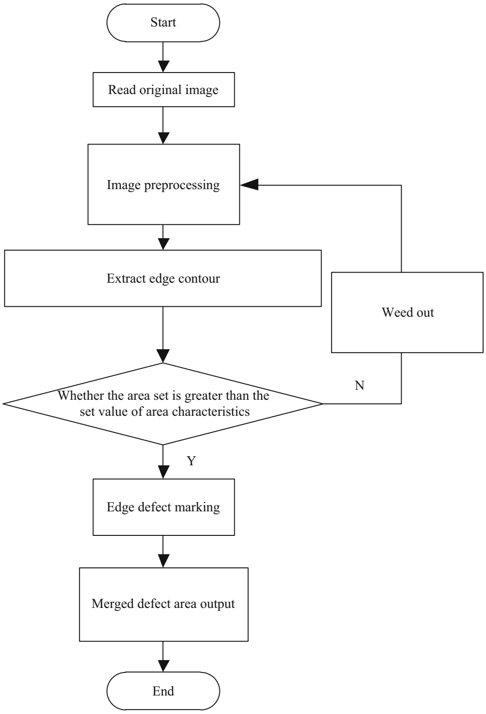

Fig. 4. Process flow of edge defect treatment

3 Experiment

In order to verify the effectiveness of the internal defect detection method of metal materials based on computer vision, experiments are carried out and the traditional internal defect detection method is compared with the method in this paper, and the detection accuracy and detection time of the two methods are compared.

3.1 Comparison of Detection Accuracy

The comparison results of detection accuracy between this method and traditional detection methods are shown in Fig. 5.

According to Fig. 5, compared with the traditional detection methods, the detection accuracy of the metal material internal defects detection method based on computer vision is higher, and the detection accuracy can be controlled above 90%.

3.2 Comparison of Detection Time

The comparison result of the detection time between the traditional detection method and the method in this paper is shown in Fig. 6.

It can be seen from the analysis of Fig. 6 that the detection time curve of the proposed method is below that of the traditional method in many tests, which proves that the

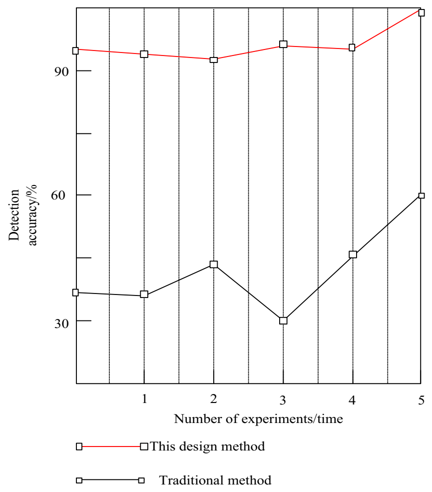

Fig. 5. Comparison of detection accuracy

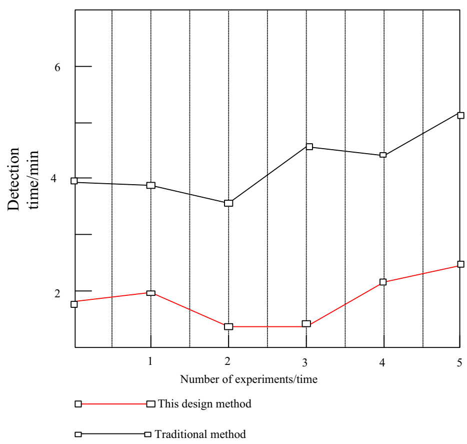

Fig. 6. Comparison of detection time

proposed method consumes less detection time than the traditional method and the detection time is always less than 3 s, thus proving that the proposed method has high timeliness.

4 Conclusion

In this study, an internal defect detection method of metal materials based on computer vision is designed. Firstly, the internal images of metal materials are collected by computer vision method, and then the images are processed and the image features are extracted. Finally, the internal defects of metal materials are accurately detected. In addition, this study also verified the application advantages of this method with high detection accuracy and detection time through experiments. However, due to the limitation of research time, this method still has some shortcomings. To this end, in the follow-up research, will do more in-depth research.

References

1. Haomin, Y., et al.: A coarse-to-fine model for rail surface defect detection. IEEE Trans. Instrum. Meas. **68**(3), 656–666 (2019)
2. Vrana, J., Goldammer, M.: Defect detection mechanisms with induction and conduction thermography: current flow and defect-specific warming. Quant. Infrared Thermography J. **17**(2), 130–151 (2020)
3. Sun, Y., Li, X., Xiao, J.: A cascaded Mura defect detection method based on mean shift and level set algorithm for active-matrix OLED display panel. J. Soc. Inform. Display **27**(1–3), 13–20 (2019)
4. Yao, M.-H., Ma, M., Liu, J.-M.: Defect detection method of magnetic disk image based on improved convolutional neural network. Optoelectron. Lett. **16**(5), 396–400 (2020)
5. He, Z., Wang, H., He, Y., et al.: Joint scanning laser thermography defect detection method for carbon fiber reinforced polymer. IEEE Sens. J. **20**(1), 328–336 (2019)
6. Liu, S., Li, Z., Zhang, Y., Cheng, X.: Introduction of key problems in long-distance learning and training. Mobile Netw. Appl. **24**(1), 1–4 (2018)
7. Shuai, L., Dongye, L., Gautam, S., et al.: Overview and methods of correlation filter algorithms in object tracking. Complex Intell. Syst. **3**, 1–23 (2020) https://doi.org/10.1007/s40747-020-00161-4
8. Lib, C., Gao, G., Liu, Z., et al.: Defect detection for patterned fabric images based on GHOG and low-rank decomposition. IEEE Access **7**(99), 83962–83973 (2019)
9. Shuai, L., Weiling, B., Nianyin, Z., et al.: A fast fractal based compression for MRI images. IEEE Access **7**, 62412–62420 (2019)
10. Wu, Y., Guo, D., Liu, H., et al.: An end-to-end learning method for industrial defect detection. Assem. Autom. **40**(1), 31–39 (2019)