

Analysis Model of Molten Crystallization of Mold Flux based on Image Segmentation Recognition

Sha Yuan

Yan'an University, Yan'an, Shaanxi, 716099, China

Abstract. In this paper, through the research and analysis of the melting and crystallization process of mold flux, the characteristics of steel are improved. In the process of casting billet, the solidification characteristics of continuous casting billet are closely related to the physical and chemical properties of mold flux. The hot wire recognition image caused a great waste of manpower and material resources. Therefore, it is particularly important to develop automatic feature extraction and mathematical modeling techniques for sequence images. In this paper, correlation analysis, image segmentation and extraction technology, time series, discrete wavelet and other related models are used to study the images in the melting and crystallization process of mold flux, and analyze the differences of mold flux in the melting and crystallization process, so as to solve the kinetic relationship between the melting and crystallization process of mold flux.

Keywords: Correlation Analysis; Image Segmentation Extraction Technology; Monte Carlo Algorithm; Discrete Wavelets; Multivariate Linear Regression.

1. Restatement of Issues

1.1 Problem Background

Since the reform and opening up, with the enhancement of China's economic strength, the steel industry has entered a stage of steady and rapid development, which has played a great role in supporting and promoting national defense, petroleum, shipbuilding, construction, equipment manufacturing and so on. Among them, the solidification characteristics of continuous casting billet are closely related to the physicochemical properties of mold flux. The protective slag in the continuous casting process has the functions of preventing the secondary oxidation of molten steel, reducing the heat loss of molten steel by thermal insulation, and absorbing and dissolving non-metallic inclusions. However, due to the high temperature, transient flow, complex phase transformation and chemical reaction, as well as the mold wall opacity, it is difficult to visually observe a series of development and changes of mold flux. The method of identifying images by hot wire method is extremely wasteful to manpower and is not conducive to the improvement of scientific and technological level. Therefore, it is particularly important to develop automatic feature extraction and mathematical modeling techniques for sequence images.

1.2 Questions Raised

Problem 1: Through image segmentation and recognition and other related technologies, the temperature of No.1 thermocouple and No.2 thermocouple in the upper left corner of each image is automatically extracted from the 562 sequence images of melting and crystallization of mold flux in Annex 1, so as to automatically import the corresponding table of Annex 32, and draw two temperature-time curves: 1 1 line temperature-2 line temperature-time diagram, 2 1 line average temperature-2 line average temperature-time diagram. There are inaccurate data of 1#line or 2#line in the given data, and the related reasons are solved.

Problem 2: By observing the six node images of Fig.1, the digital image processing technology is used to study the images, and the dynamic differences between the adjacent sequence images during the melting and crystallization process of the mold flux are quantitatively analyzed. The time series model is established for the different features between the quantized images, and the curve image of the melting and crystallization process of the mold flux is analyzed and discussed.

2. Problem Analysis

2.1 Analysis of Problem One

For problem one, it is required to establish a mathematical model to extract data to explore and verify the accuracy of the experimental results. In this paper, image segmentation technology is used to segment 562 sequence images of slag melting and crystallization by using Python language os, cv2, numpy third-party library, and pytesseract.image _ to _ string () method in third-party library is used to extract the temperature data of No.1 thermocouple and No.2 thermocouple in the upper left corner, which are automatically imported into the corresponding table in Appendix 2. The temperature-time curve is made by using the data, and the data analysis is carried out to obtain the error of the data measurement.

2.2 Analysis of Problem Two

Aiming at the second problem, firstly, the six node images in Fig.1 are divided into six sequences by digital image processing technology. By studying the correlation between images, the Monte Carlo algorithm is used to calculate the area ratio and image features to quantify the dynamic differences between adjacent sequence images during the melting and crystallization of mold fluxes. By establishing a time series model, the different characteristics of the image are analyzed and studied, and a simulation model is established to study the curve of the melting and crystallization process of the mold flux and analyze the reasons.

2.3 Analysis of Problem Three

For the third problem, under the condition of known temperature and time, we establish a regression model to establish the functional relationship between the change of temperature and time and the melting and crystallization process of the mold flux, and use the numerical simulation results to analyze the mold flux. The relationship between temperature, melting rate and crystallization rate.

3. Problem Hypothesis

1. Assume that the monitoring data provided in the picture is true and reliable.
2. Assuming that the topic in addition to the selected factors, the effect of other factors can be ignored.
3. Assume that there is no chemical reaction in the sample before the test.
4. Assuming that all # 1wire and # 2wire temperature is detected in the same environment conditions.

4. Symbol Description

symbol	symbol description
P_{ij}	The pixel value of the j th pixel of the i th color channel of the digital image P
N	Number of pixels in the imageTeam
X_t	The actual observed value of period t
I_t	The estimated level of time t
b_t	The predicted trend of time t
α	The smoothing parameter of level and trend
a	Dimension
b	Displacement

5. Establishment and Solution of Problem One Model

5.1 The Preparation of Problem One

5.1.1 Image Segmentation Technology

a. Data preprocessing

The corresponding time in the image and the temperature of 1 # wire and 2 # wire are in the upper left corner of the image. The image matrix is read into the computer, and the image is segmented by the method of matrixization. The size of each matrix is equal by the method of matrixization. The length of the matrix is set to 100 and the width is set to 70 by the method of trial and error. Identify and crop out the top left corner in the image. The cut out image is shown in Figures 1 and 2 :

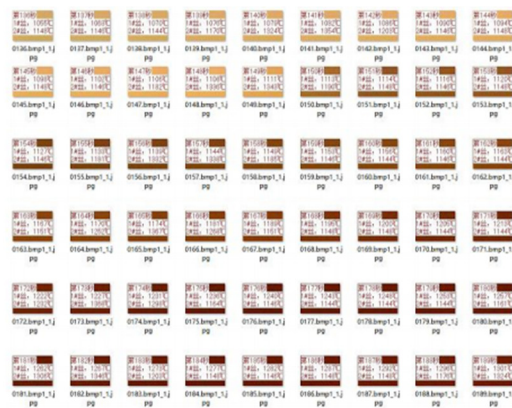


Figure 1. Partial clipped image



Figure 2. 141st seconds image

b. In the cmd command prompt input: pip install os, pip install cv2, pip install numpy commands to download os, cv2, numpy three third-party libraries

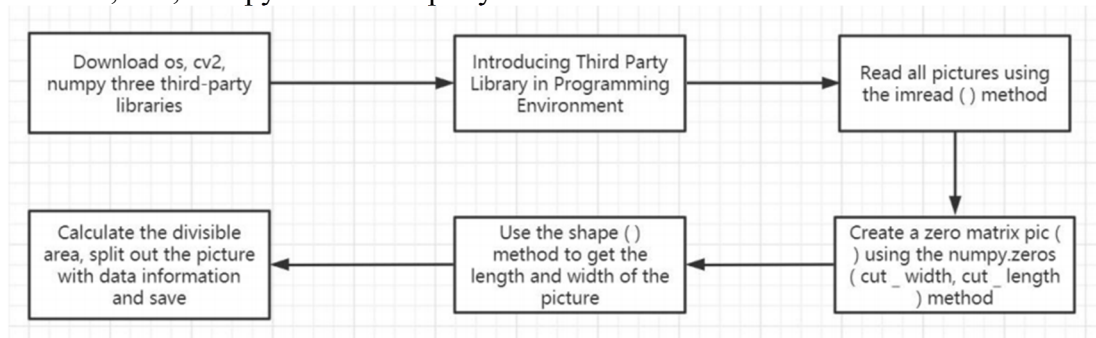


Figure 3. Image segmentation technology flow chart

5.1.2 Image Recognition and Extraction

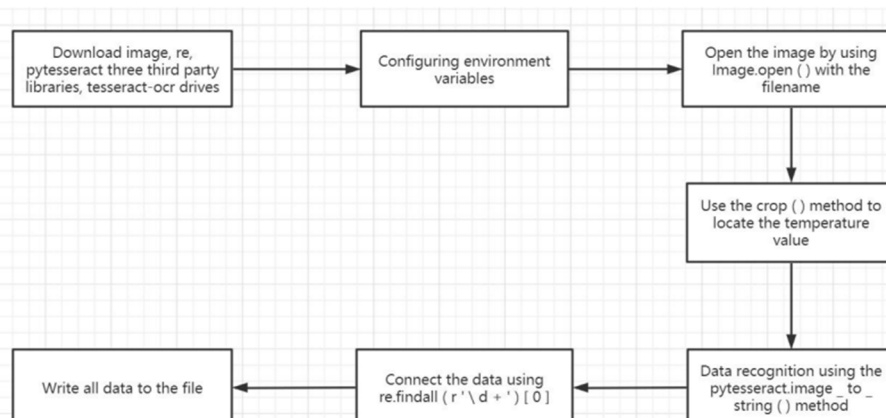


Figure 4. Image recognition and extraction flow chart

In the cmd command prompt, enter: pip install image, pip install re, pip install pytesseract command to download image, re, pytesseract three third-party libraries. Download and install tesseract-ocr using the tesseract website. After the installation is completed, open the advanced system settings- > environment variables- > system variable path, add C: \ Program Files \ Tesseract-OCR to the system variable path, and then add C: \ Program Files \ Tesseract-OCR \ tessdata to a new system variable named TESSDATA _ PREFIX in the environment variable. The configuration installation process ends.

5.2 A Step-by-step Technical Operation Document

5.2.1 Image Segmentation Technology

First, in the Python programming environment, the import statement is used to call the three third-party libraries os, cv2, and numpy to split the graph and assign Attachment 1 to the path _ photo custom variable.

Second, using the os.listdir () method provided by the os library, traverse all filenames under the path _ photo folder and assign each filename to files _ list, traversing each element in files _ list through the for loop.

Third, change each element type to string type, get the relative path of each image by os.path, join () method in os library, use imread () method provided by cv2 to read the relative path, open the photo corresponding to the path and assign the variable picture.

Fourth, define two variables cut_width, cut_length and assign any initial value, and establish a preprocessing zero matrix pic () by numpy 's zeros (cut _ width, cut _ length) method.

Fifth, the length and width of the image are obtained by the shape () method of cv2, and the divisible region is calculated, and the image with data information is segmented and saved.

5.2.2 Image Recognition and Extraction

First, in the Python programming environment, the for loop is used to traverse the file name of each image obtained from the segmentation, and the image is opened using the Image.open () method provided by Image.

Secondly, after opening the picture, the crop () method provided by Image is used to find the location of the temperature value by trial and error. The picture that has been opened and the data location is determined is used to realize data recognition by pytesseract 's pytesseract.image _ to _ string () method.

NO	Time	1# Temperature	2# Temperature
1	110	900	1142
2	111	904	1146
3	112	910	1146
4	113	914	1144
5	114	918	1144
6	115	923	1146
7	116	927	1146
8	117	932	1146
9	118	935	1148
10	119	938	1146
11	120	942	1146
12	121	947	1153
13	122	951	1267
14	123	953	1418
15	124	959	1377
16	125	965	1218
17	126	970	1146
18	127	974	1148

Figure 5. Partial table data

Third, using the `pytesseract.image_to_string()` method to set the language pattern `lang` to 'chi_sim', and set the engine to `psm-6` (using the data as a unified text block), each number in the picture can be extracted and used as a string type. Through the `findall(r'\d+')[0]` provided by the `re` library, the extracted data can be expressed and connected by regular expressions, and this time point and the two-wire temperature of this time point can be merged into one element through string formatting. Write all these elements to `all_data`, create a file named 'data.csv', set to 'w', write all the data to the 'data.csv' file through `write(all_data)`, and the image recognition ends.

5.3 Temperature-time Image of Problem One

Use the excel table to analyze the data extracted in 5.2 and draw a graph, as shown in Figure 6.

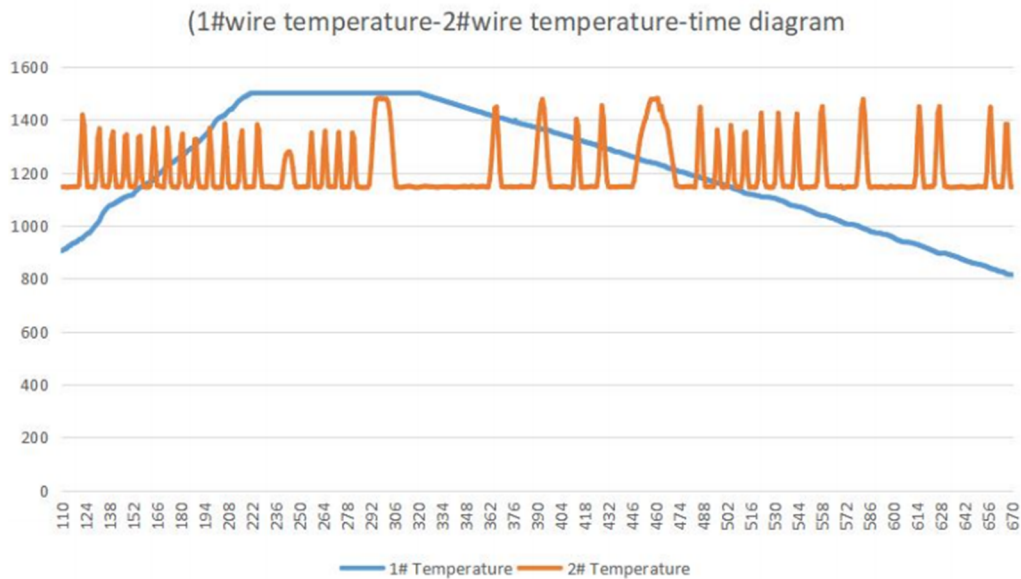


Figure 6. 1#wire temperature-2#wire temperature-time diagram

The average temperature-time diagram is an analogy of the average speed-time relationship, as shown in Figure 7.

Average temperature in the nth second= (The temperature of n - 1 seconds + The temperature of n seconds)/2

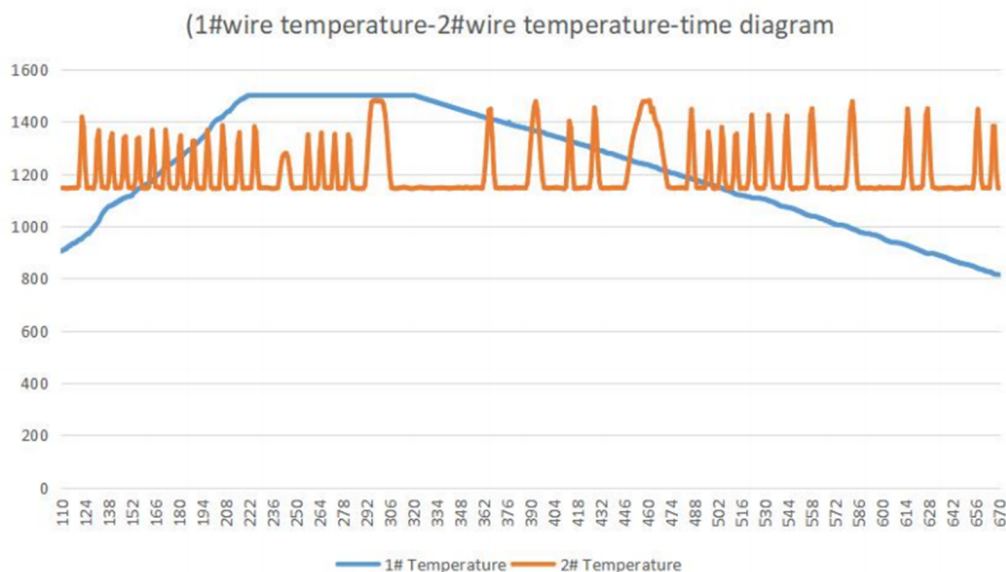


Figure 7. 1#wire average temperature-2#wire average temperature-time diagram

As can be seen from the curve image (Fig. 6) : 1#wire temperature overall rise-stable-down trend. The sample began to melt from 110 s, and the temperature gradually increased to the melting point of molten steel (1500 °C) in 110-222s, and the curve showed an upward trend. In 222-321s, the molten steel has been in the melting point state, the curve was stable trend. At 321-671s, the temperature of molten steel decreased steadily with time until crystallization. The trend of the average temperature-time curve in Fig.7 is basically consistent with the trend of the temperature-time curve in Fig.6.

5.4 Analysis of Data Results of Question One

1#wire temperature in the 322-671s this process continued to decline steadily, but in the 378s this time the temperature suddenly rose, compared to its adjacent sequence image, this point is abnormal, thus determining this point measurement error. Observing the temperature-time image, the overall trend of 1#wire is rising-stationary-falling, and 2#wire is generally stable but fluctuates frequently. SHTT II melting and crystallization temperature tester is double platinum rhodium wire bent into a U-shaped temperature element, but also temperature measuring parts. It is judged that the temperature fluctuation is frequent due to the poor contact of the 2#wire temperature probe.

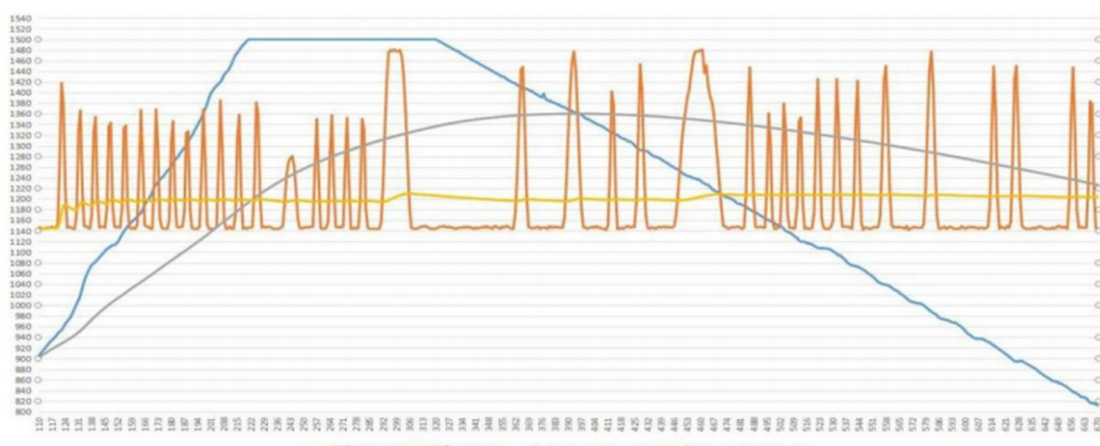


Figure 8. Comparison diagram of temperature-average temperature-time curve

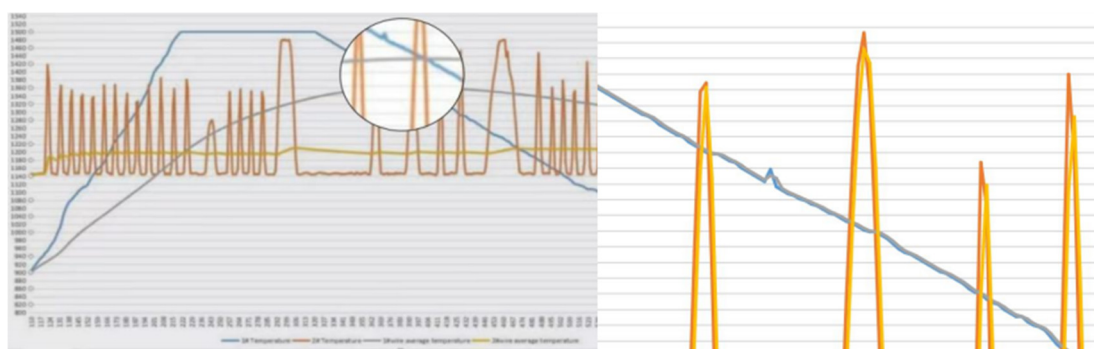


Figure 9. Problem Sample

6. Establishment and Solution of Problem Two Model

6.1 Establishment and Solution of Problem Two Model

6.1.1 Establishment and Solution of Difference Analysis Model of Area Ratio and Time

According to the six node images in the title, we divided 562 mold flux melting crystallization images into six sequences (the first sequence 0s, the second sequence 0-110s, the third sequence 110-

141s, the fourth sequence 141-149s, the fifth sequence 149-150s, the sixth sequence 150-671s), and solved them by Monte Carlo method.

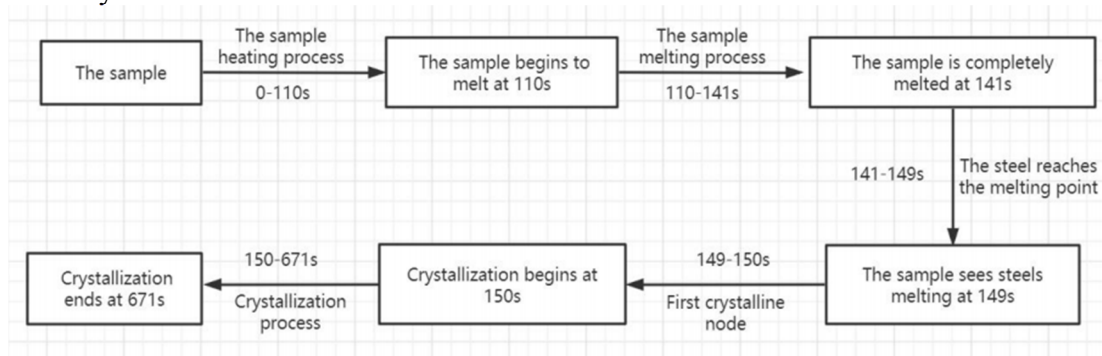


Figure 10. Protective slag melting crystallization sequence flow chart

We mark the left endpoint of the image as a , and the right endpoint as b . On the interval $[a, b]$, we use the computer to uniformly generate n random numbers $x_1, x_2, \dots, x_n$, using the `unifrnd` command of matlab software. Then calculate the value of the integrand $f(x_1), f(x_2), f(x_n)$ corresponding to each random number, calculate the average value of the integrand $x = \frac{\sum_{i=1}^n f(x_i)}{n}$ solving Integral by Monte Carlo Method $\theta = \int_a^b f(x)dx$ the area ratio of the first sequence is 43.23 %, and the area ratio of the second sequence is 72.60 %. The area ratio of the third sequence was 56.31 %, the area ratio of the fourth sequence was 52.76 %, the area ratio of the fifth sequence was 54.46 %, and the area ratio of the sixth sequence was 55.68 %.

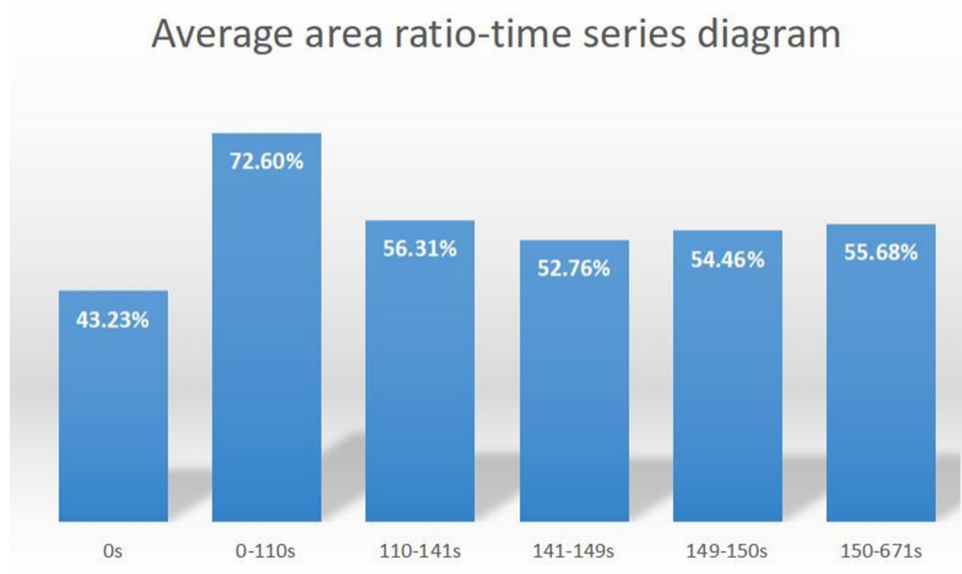


Figure 11. Area Ratio Result Chart

In the melting and crystallization process of mold flux, we use Monte Carlo method to calculate the area ratio of samples in each image. in the process from the second sequence to the third sequence, the average area ratio of the image decreases ; the average area ratio of the third sequence to the sixth sequence is basically consistent, and the difference is not significant. Melting process of protective slag due to temperature rise, from the solid slag through chemical reaction to form a low melting point material and then converted into liquid slag, liquid slag from the liquid steel surface of the slag pool into the gap between the shell and the copper mold wall, area ratio is reduced. The three-layer slag film structure (glass layer, crystal layer and liquid slag layer) formed during the crystallization process is relatively stable, and the area ratio remains unchanged.

6.1.2 Establishment and Solution of Difference Analysis Model of Image Feature and Time

a. Establishment of the model

To find the dynamic difference between adjacent sequence images, we use the method of color, discrete wavelet and texture feature extraction to model the image. Image feature extraction can be regarded as a generalized image transformation, which transforms the image from the original attribute space to the feature attribute space.

1 Extraction of color features

The first moment represents the mean of data distribution, the second moment represents the variance of data distribution, and the third moment represents the deviation of data distribution. For the digital image P, the first-order color moment is defined as: $u_i = \frac{1}{N} \sum_{j=1}^N P_{ij}$. p_{ij} represents the pixel value of the j th pixel of the i th color channel of the digital image P, and N represents the number of pixels in the image.

The second-order color moment is defined as: $\sigma_i = [\frac{1}{N} \sum_{j=1}^N (P_{ij} - u_i)^2]^{\frac{1}{2}}$

The third-order color moment is defined as: $S_i = [\frac{1}{N} \sum_{j=1}^N (P_{ij} - u_i)^3]^{\frac{1}{3}}$

The first-order moment can represent the average response intensity of the color channel, the second-order moment can represent the response variance of the color channel, and the third-order moment can represent the offset of the data distribution of the color channel.

2 Extraction of Discrete Wavelet Features

A tool for signal processing by adaptive window size, providing a 'time-frequency relationship' window that changes with frequency for signal time-frequency analysis and processing.

$$\omega(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) * \varphi\left(\frac{t-b}{a}\right) dt$$

Wavelet transform has two variables : scale a and shift b. The scale a controls the stretching of the wavelet function, and the shift b controls the shifting of the wavelet function. The scale corresponds to the frequency (inverse ratio), and the shift b corresponds to the time.

3 Extracting texture feature

Gabor transform belongs to windowed Fourier transform. Gabor function can extract relevant features in different scales and directions in frequency domain. In addition, the Gabor function is similar to the biological effects of the human eye, so it is often used for texture recognition and has achieved good results. The two-dimensional Gabor function can be expressed as :

$$f(x,y,\sigma_x,\sigma_y,\omega_f,\theta_f) = \frac{1}{2\pi\sigma_y\sigma_x} \exp\left\{-\left[\frac{(x\cos\theta_f+y\sin\theta_f)^2}{2\sigma_x^2} + \frac{(-x\sin\theta_f+y\cos\theta_f)^2}{2\sigma_y^2}\right]\right\} * \exp\{j2\pi(\omega_f x \cos\theta_f + \omega_f y \sin\theta_f)\}$$

where σ_x , σ_y represent the spatial scale factors of horizontal and vertical directions respectively, ω_f and θ_f represent the center frequency and direction respectively. Two real filters can be obtained by decomposing the filter : cosine Gabor filter and sine Gabor filter.

The cosine Gabor filter is expressed as:

$$f(x,y,\sigma_x,\sigma_y,\omega_f,\theta_f) = \frac{1}{2\pi\sigma_y\sigma_x} \exp\left\{-\left[\frac{(x\cos\theta_f+y\sin\theta_f)^2}{2\sigma_x^2} + \frac{(-x\sin\theta_f+y\cos\theta_f)^2}{2\sigma_y^2}\right]\right\} * \cos[2\pi(\omega_f x \cos\theta_f + \omega_f y \sin\theta_f)]$$

The sine Gabor filter is expressed as:

$$f(x,y,\sigma_x,\sigma_y,\omega_f,\theta_f) = \frac{1}{2\pi\sigma_y\sigma_x} \exp\left\{-\left[\frac{(x\cos\theta_f+y\sin\theta_f)^2}{2\sigma_x^2} + \frac{(-x\sin\theta_f+y\cos\theta_f)^2}{2\sigma_y^2}\right]\right\} * \sin[2\pi(\omega_f x \cos\theta_f + \omega_f y \sin\theta_f)]$$

Two-dimensional cosine Gabor filter is a commonly used image feature extraction filter. By Fourier transform of cosine Gabor filter, it can be obtained that:

$$F(\omega_x, \omega_y, \sigma_x, \sigma_y, \omega_f, \theta_f) =$$

$$\frac{1}{2} \exp\{-2\pi^2[\sigma_x^2(\omega_x \cos\theta_f + \omega_y \sin\theta_f + \omega_f)^2 + \sigma_y^2(-\omega_x \sin\theta_f + \omega_y \cos\theta_f)^2]\} +$$

$$\frac{1}{2} \exp\{-2\pi^2[\sigma_x^2(\omega_x \cos\theta_f + \omega_y \sin\theta_f - \omega_f)^2 + \sigma_y^2(-\omega_x \sin\theta_f + \omega_y \cos\theta_f)^2]\}$$

b. Solution of the model

The color moment is used to extract the molten slag crystallization process image. The color moment of the image has a total of nine components, and each color channel has three low-order moments, as shown in Figure 11. Using numpy as np, cv2, pywt, matplotlib.pyplot as plt to construct discrete wavelet feature map, as shown in Fig.13. Using Gabor transform, the relevant texture features of images are extracted in different scales and directions, as shown in Fig. 14.

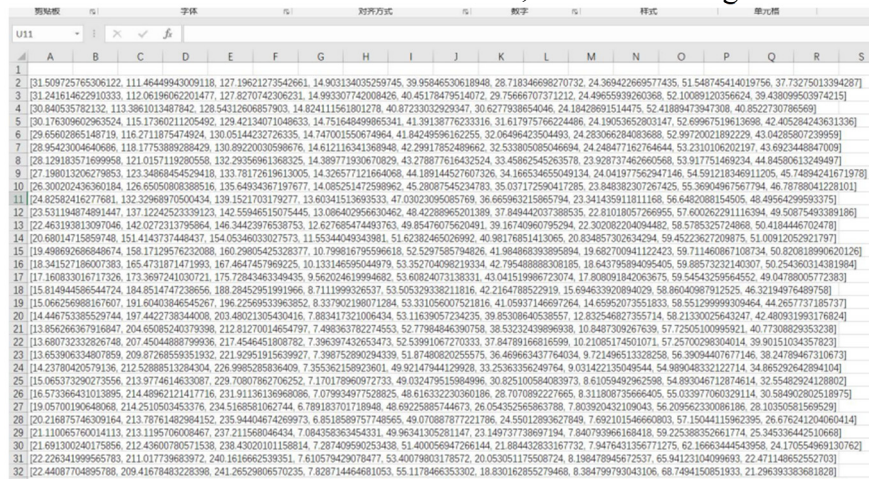


Figure 12. Color data graph

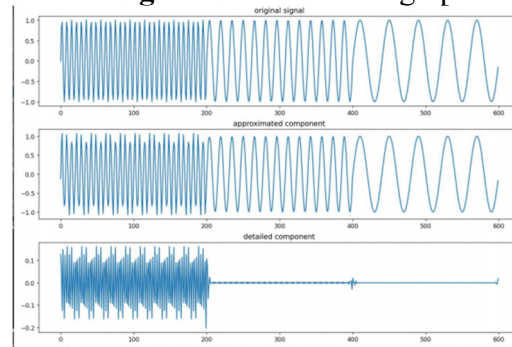


Figure 13. Frequency wave diagram

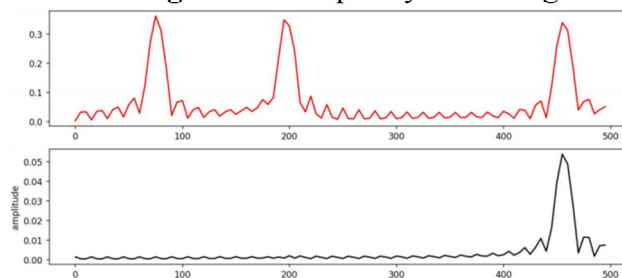


Figure 14. Discrete wavelet curve diagram

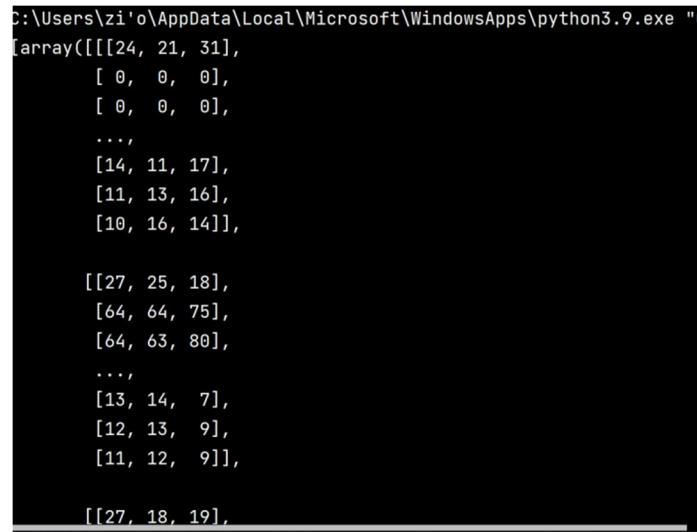


Figure 15. Texture data map

Based on the extracted image feature results, the curve of the slag melting and crystallization process was analyzed and studied by two time series prediction models.

The Brown linear trend model is mainly constructed by horizontal smoothing equation, trend smoothing equation and prediction equation. where h is the prediction step, t is the current period, X_t is the actual observed value of period t , $\hat{X}_t$ is the estimated level of time t , $\hat{X}_{t+h}$ is the predicted trend of time t , and α is the smoothing parameter of level and trend.

$$I_t = \alpha X_t + (1 - \alpha)(I_{t-1} + b_{t-1}) \quad (\text{Horizontal smoothing equation})$$

$$b_t = \alpha(I_t - I_{t-1}) + (1 - \alpha)b_{t-1} \quad (\text{Trend smoothing equation})$$

$$\hat{X}_{t+h} = I_t + h b_t, \quad h=1,2,\dots \quad (\text{forecasting model})$$

The above model explores a stationary time series, and the time series explored by the second question may be a d -order unit root process, so the ARIMA (p, d, q) model is used (p is the number of autoregressive AR items, q is the number of moving average MA items, d is the number of differences made to make the sequence stable). The data need to be differential processed first, and then converted into a stable time series before modeling. Its prediction formula is:

$$y = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_q \epsilon_{t-p}$$

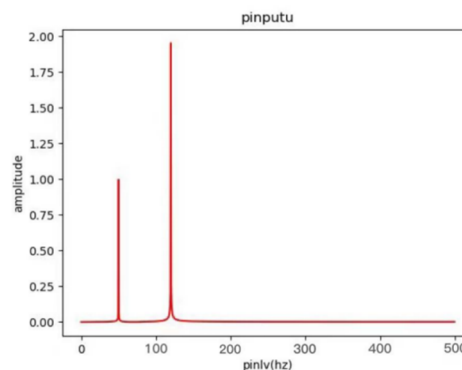


Figure 16. Difference Curve of Melting Crystallization Process of Protective Slag

From Figure 16, it can be seen that the overall trend of the difference curve of the melting and crystallization process of the adjacent sequence slag remains unchanged. The characteristics quantified at the nodes are basically the same.

By fitting the area ratio feature and image feature, the fitting curve is obtained, as shown in Figure 17. The whole process of melting and crystallization of protective slag is taken as a whole and quantified as one unit to draw the progress curve of melting and crystallization process of protective slag, as shown in Fig18.

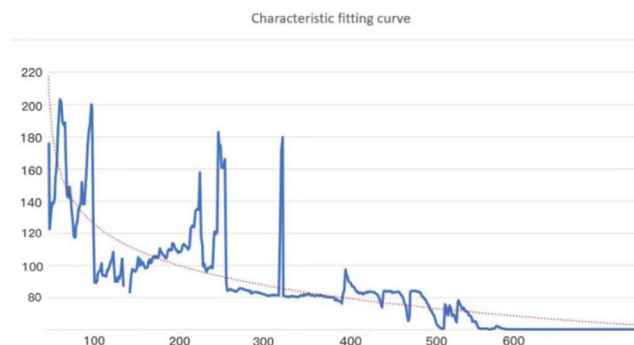


Figure 17. Fitting curve

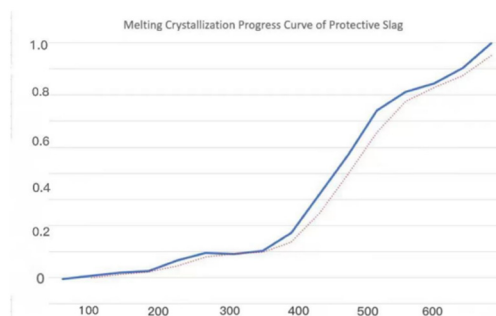


Figure 18. Melting Crystallization Progress Curve of Protective Slag

References

- [1] Wang Yongjing. Application analysis of mathematical modeling thought in probability and statistics teaching [J]. Mathematics learning and research. 2021 (23): 122-123.
- [2] Stricker M A, Orengo M. In Similarity of color images, Storage and retrieval for image and video databases III, International Society for Optics and Photonics: 1995; pp 381-392.
- [3] Chen Xiaojun, Ji Fuxing. Comprehensive Evaluation of Customer Relationship Management, Internal Control and M & A Performance Measurement-Based on Multiple Linear Regression Model Analysis [J]. Management Review, 2021,33 (08): 256-262.
- [4] Jiang Qiyuan, Xie Jinxing, Ye Jun. Mathematical Modeling (5th Edition) [M]. Higher Education Press, 2018.