Research on Parallel Mining Method of Massive Image Data Based on AI

Shuang-cheng Jia(✉) and Feng-ping Yang

Alibaba Network Technology Co., Ltd., Beijing 100102, China
tomjia1980@126.com

Abstract. Parallel mining of image data is based on the extraction of internal rules and detail features of image. Combined with image edge detection to realize parallel mining of image data, a parallel mining algorithm of image data based on AI is proposed. Firstly, the multidimensional parallel eigenvalues of image data are extracted by the gray feature extraction algorithm of massive images, and then the template matching and information fusion of massive image data are carried out by using Map/Reduce model. According to the matching results, the parallel mining results of image data are obtained. Finally, the simulation experiment of image data parallel mining is realized by using Matlab software. The results show that compared with other image data parallel mining algorithms, this algorithm reduces the parallel mining time of image data and improves the speed of image data parallel mining, especially for large-scale image data parallel mining.

Keywords: AI · Massive image data · Parallel mining · Template matching

1 Introduction

The mining and processing of large-scale images is the basis of image recognition. With the development of digital image information processing technology, higher requirements are put forward for the accuracy of image mining. It can effectively realize image target recognition, mechanical fault diagnosis, condition monitoring and pattern recognition. The research of image data parallel mining technology will show good application value in the fields of industrial production and military strike. The research of parallel mining algorithms for related image data has been paid more and more attention [1].

Image data parallel mining is based on the internal rules and detail feature extraction of the image. Combined with image edge detection, image noise reduction, image enhancement and image segmentation algorithm, the regular feature extraction of the image is realized [2]. By extracting the feature information which can reflect the attributes of the target image, the parallel mining of image data is realized by combining the excavator. Traditionally, the parallel mining method of image data is mainly based on a single board computer, such as support vector machine algorithm, BP neural network algorithm and so on [3]. In reference [4], a method of radar imaging feature extraction and image data parallel mining based on edge feature fusion is proposed to realize the mining and recognition of aerial target images, and the mining accuracy of

G. Gui and L. Yun (Eds.): ADHIP 2019, LNICST 302, pp. 118–127, 2019.
https://doi.org/10.1007/978-3-030-36405-2_13

aircraft and other target images is good. However, with the increase of interference intensity, the accuracy of mining is limited, and the computational cost of the algorithm is large. In reference [5], a parallel image data mining and recognition algorithm based on autocorrelation matching detection is proposed and applied to mechanical fault detection. By extracting CT scanned image under mechanical fault state and mining by BP neural network, To realize the recognition of image fault attribute mining and improve the performance of fault diagnosis, but the algorithm adopts a single board computer image data parallel mining method, which cannot integrate and mine a large number of large-scale images. The efficiency of image batch processing is not good, the efficiency is low, and the application value is not good.

Artificial Intelligence (AI) is a new technical science that studies and develops the theory, method, technology and application system used to simulate, extend and extend human Intelligence. It is widely used in the field of computer vision and image processing. As a result, to solve the above problems, this paper designs an artificial intelligence-based parallel mining algorithm for image data, with the following ideas: firstly, the multi-dimensional parallel characteristic value of the image data is extracted, the mass image data is matched and the information fusion is carried out on the mass image data by adopting a Map/Reduce model, the parallel mining result of the image data is obtained according to the matching result, and finally, the simulation experiment of the parallel mining of the image data is realized by using the Matlab software, The effectiveness of this method is proved by the experimental results, and the effectiveness conclusion is shown, and the superiority of the image data parallel mining algorithm designed in this paper in the realization of image cloud computing is shown.

2 Image Pre-processing and Feature Extraction

2.1 Collection and Noise Reduction Preprocessing of Massive Image Data

In order to realize that parallel mining and processing of large-scale mass image data, the property of the mass image data is judged, the acquisition and noise reduction preprocessing of the mass image data is required, and the data information input is provided for the mass image data parallel mining model [6]. According to the scale invariance of the mass image data, the feature detection and acquisition of mass image data are carried out by four different methods such as transverse scanning, longitudinal scanning and oblique scanning and block scanning, and the input mass image data is assumed to be $f(x, y)$, the mass image data acquisition process is interfered by the additive noise term $\eta(x, y)$, and the acquired pixel value of the mass image data is obtained according to the change characteristics of the density and the noise intensity of the imaging in the mass image data acquisition process:

$$g(x, y) = h(x, y) * f(x, y) + \eta(x, y) \tag{1}$$

Feature acquisition and scanning process can enlarge the scale of image data acquisition. The process of grid computing for image data is as follows: cloud computing batch processing technology is adopted to obtain large-scale image data, and the pixel sequence output after cloud computing processing is shown as follows:

$$g(x,y) = f(x,y) + \eta(x,y) \tag{2}$$

On the basis of the above massive image data acquisition, it is found that the massive image data acquisition process is interfered by the jitter of the acquisition equipment and the disturbance error of the medium, resulting in noise. The attribute characteristics of massive image data are not obvious [7]. In order to improve the accuracy of mining, it is necessary to reduce the noise of massive image data, and extract the statistical features of the massive image data from the output of denoising processing through cloud computing. In order to improve the efficiency of parallel mining of massive image data, it is necessary to construct batch processing model of massive image data. Finally, an excavator is constructed in Map/Reduce model to realize accurate mining of massive image data. According to the above analysis, the block diagram of the parallel mining algorithm for massive image data designed in this paper is shown in Fig. 1.

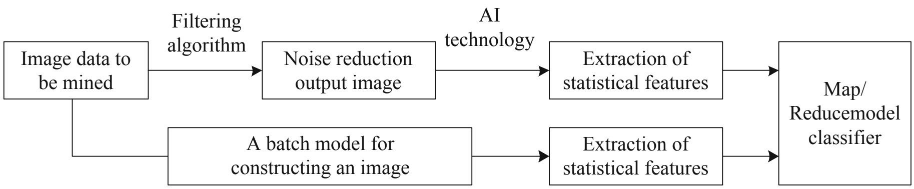


Fig. 1. Implementation flow of parallel mining algorithm for massive image data

According to the overall design idea and the flow of the algorithm, combined with the acquisition result of the mass image data, the noise reduction pre-processing of the mass image data is first carried out, the wavelet noise reduction algorithm is designed to perform the noise reduction of the mass image data, firstly, the edge information distortion amplitude after being subjected to noise interference of the mass image data to be excavated is given as follows:

$$\hat{f}(x,y) = \beta F(x,y) + (1-\beta)m_l \tag{3}$$

Wherein, $F(x,y)$ is the envelope error of white balance distortion of massive image data to be excavated. The pixel value of noisy massive image data at (x,y) point is β, m_l is the embedding dimension of massive image data denoising along the gradient direction, and the parent wavelet basis function is constructed as:

$$F = \tilde{p}(x,y) = p(x,y)\left(\frac{v(x)}{v(y)}\right)^{1/2} \tag{4}$$

Wherein

$$p(x,y) = \frac{k(x,y)}{v(x)}, v(x) = \sum_{y} k(x,y) \tag{5}$$

The above represents the color difference and gradient pixel difference of the massive image data, combined with the dark primary color feature information of the massive image data, the noise in the massive image data is adaptively filtered, and the filtering function is described as follows:

$$\hat{f}(x,y) = \begin{cases} g(x,y) - 1, & if \quad g(x,y) - \hat{f}_{Lee}(x,y) \geq t \\ g(x,y) + 1, & if \quad g(x,y) - \hat{f}_{Lee}(x,y) < t \\ g(x,y), & else \end{cases} \tag{6}$$

Wherein, $\hat{f}_{Lee}(x,y)$ represents the estimation function of continuous wavelet time scale decomposition, and the adaptive focusing of noise points of massive image data is realized by wavelet scale decomposition and adaptive local noise filtering [8]. The focused noise is erased adaptively by autocorrelation detector, and the filtered massive image data is obtained as follows:

$$g(x,y) = f(x,y) + \eta_m(x,y) \tag{7}$$

The Markov chain model of massive image data is constructed, and the output of principal component gray histogram after denoising of massive image data is obtained as follows:

$$p(\eta_m(x,y)) = \begin{cases} \frac{r}{4}, & \eta_m(x,y) = -1 \\ 1 - \frac{r}{2}, & \eta_m(x,y) = 0 \\ \frac{r}{4}, & \eta_m(x,y) = 1 \end{cases} \tag{8}$$

Thus, the collection and noise reduction preprocessing of massive image data is realized, which is used as data input to extract features.

2.2 Feature Extraction of Massive Image Data

A variety of feature extraction and fusion of mass image data are the basis for realizing parallel mining of mass image data. The high-order moment characteristics of mass image data are extracted to carry out parallel mining of mass image data, and a cloud computing technology is adopted, and the degradation model of the gray histogram after noise reduction and output is expressed as follows:

$$g(x,y) = f(x,y) + \varepsilon(x,y) \tag{9}$$

The noise obeys $n \in N(0, \sigma_n^2)$ distribution, it performs high-order cumulant segmentation on the gray histogram of massive image data output. In the cloud computing environment, the high-order cumulant segmentation model of massive image data is obtained as follows:

$$\begin{cases} J_1(w,e) = \frac{\mu}{2} w^T w + \frac{1}{2}\gamma \sum_{i=1}^{N} e_i^2 \\ s.t. y_i = w^T \varphi(x_i) + b + e_i, \quad i = 1, \cdots, N \end{cases} \tag{10}$$

Under the condition of constant frame difference, the second order invariant moments of massive image data are fused, multiple feature extraction and affine invariant moments which can reflect the feature properties of massive image are carried out. The data to be mined are as follows:

$$F = \tilde{p}(x,y) = p(x,y)\left(\frac{v(x)}{v(y)}\right)^{1/2} \tag{11}$$

Where

$$p(x,y) = \frac{k(x,y)}{v(x)}, v(x) = \sum_y k(x,y) \tag{12}$$

The above formula represents the color difference and gradient pixel difference of the feature attributes of the massive image data respectively. The cloud computing is carried out by using the marked watershed method, and the high-order moment feature extraction output of the massive image data is obtained as follows:

$$W_\psi y(a,b) = <y, \psi_{a,b}> = \int_{-\infty}^{+\infty} y(t) \frac{1}{\sqrt{|a|}} \psi^*(\frac{t-b}{a}) dt \tag{13}$$

Where, $<y, \psi_{a,b}>$ is the second invariant moment of massive image data, $y(t)$ is the pixel value of inputting massive image data, b is wavelet scale, ψ^* is the segmentation threshold. Based on the above processing, a variety of features extraction of massive image data is realized.

3 Improved Implementation of Parallel Mining Algorithm for Massive Image Data

3.1 Algorithm Design

On the basis of collecting and denoising massive image data and extracting high-order moment features of massive image data, the parallel mining algorithm of massive image data is optimized [9]. According to the target attribute characteristics of massive image data, the position scale information of massive image data is described as follows: according to the target attribute characteristics of massive image data in Map/Reduce model:

$$\psi_{a,b}(t) = [U(a,b)\psi(t)] = \frac{1}{\sqrt{|a|}}\psi(\frac{t-b}{a}) \tag{14}$$

The Map/Reduce model is used to estimate the gradient value of massive image data moving along the moving average window $(f, \hat{f})$, and the deviation compensation is used to replace $a = 1/s$, $b = \tau$, and the above formula is rewritten as:

$$f_{s,\tau}(t) = [U(1/s,\tau)f(t)] = \sqrt{|s|}f(s(t-\tau)) \tag{15}$$

Under the cloud computing environment, the mass image data is batch processed, and the rapid iterative contraction of the mass image data parallel mining attribute is carried out at the (a, b_m) point on the scale translation plane of the mass image data in parallel, and the obtained mining center is expressed as follows:

$$L(a, b_m) = \log\left(\frac{|V||V_m \cap V_n|}{|V_m||V_n|}\right) \tag{16}$$

The white balance deviation compensation is carried out for the highlight model of mass image data to obtain the gray scale value of the first order field of the mining mass image data, and the iterative process of matching the high-order moment characteristics with scale M is expressed as follows:

$$d_{i+1} = 2F(x_{i+1} + \frac{1}{2}, y_i + 2) \tag{17}$$

For the single scale Harris variable $f = \frac{N - \bar{N}}{N}$, each feature point of massive image data to be excavated as the number of eigenvalues, a fast iterative shrinkage threshold algorithm is adopted to realize the fast search and processing of parallel mining process of massive image data through cloud computing. As a result, the mining efficiency of massive image data is improved [10].

3.2 Algorithm Implementation Flow

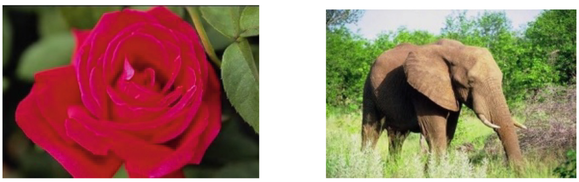
Based on the above-mentioned algorithm design, the implementation of the parallel mining algorithm of mass image data based on AI designed in this paper is described as follows:

Step1: The scale invariance of massive image data is used, the massive image data is collected, and the massive image data input is based on cloud computing batch processing for $f(x, y)$. The output pixel sequence of massive image data processed by cloud computing is obtained.

Step2: The mass image data is filtered and preprocessed, and the wavelet noise reduction algorithm is designed to reduce the noise of mass image data. The autocorrelation detector is used for adaptive erasing the focused noise point, and the filtered output mass image data $g(x,y) = f(x,y) + \eta_m(x,y)$ is obtained.
Step3: The attribute feature analysis and extraction of massive image data are carried out by feature extraction method, and the affine invariant moment $F = \tilde{p}(x,y)$ and high order moment $W_{\psi}y(a,b)$.
Step4: According to the weight of the target feature model, the pixel weights of the target and the candidate region are calculated, and the deviation compensation and weighting of the massive image data are carried out. according to the texture information and high-order moment features of the obtained massive image data, Adaptive Lyapunove functional is used to retrieve the attributes of massive image data mining in parallel.
Step5: The invention uses the AI model to estimate the mass image data along the sliding average window, and is matched with the feature points to obtain a Map/Reduce model of parallel mining of the mass image data to realize the parallel mining of the mass image data.

4 Simulation Experiment and Result Analysis

In order to test the application performance of the algorithm in the parallel mining of mass image data, the simulation experiment is carried out, the experiment is established on the Matlab simulation software, the experimental data set is the Corbel standard mass image database, and the mass image database contains large amount of image data of various attributes, in the process of carrying out the corbel standard mass image database retrieval, a mass image data parallel mining process is needed, the flowers and the animals in the mass image database are taken as test sets, and the original test mass image data is obtained as shown in Fig. 2.



(a) Flower (b) Elephant

Fig. 2. Test image sample set

Taking the above two kinds of massive image data as the test set, the massive image data of Corel standard mass image database is excavated and searched in parallel, and the massive image data is processed by this method. Firstly, the noise reduction and feature extraction of massive image data are carried out, and the high-order moment features and affine invariant moment features of massive image data are extracted by gray feature extraction algorithm of massive image data. Then the Map/Reduce model is used for template matching and information fusion of massive image data, and the mining results are shown in Figs. 3 and 4.

Fig. 3. Output results of parallel mining and retrieval for massive image data of flower class

Fig. 4. Mining results of large-scale image data of elephants

It can be seen from the diagram that the parallel mining of massive image data by using this method can realize the rapid mining and retrieval of massive image data in the whole Corel standard massive image database, and the category attributes of outputting massive image data are accurate.

In order to quantitatively test the efficiency and accuracy of mining algorithm, 1000 monte carlo experiments were used to conduct multi-feature fusion and parallel mining of massive image data. In the experiment, the comparison method is adopted to compare the massive image data parallel mining method based on AI proposed in this paper with the Reference [4] method and the Reference [6] method. The index comparison results of parallel mining of massive image data are shown in Fig. 5.

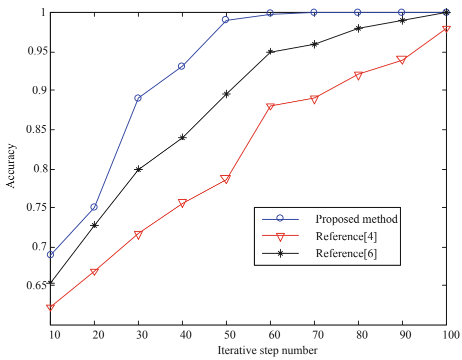


Fig. 5. Comparison of accuracy in parallel mining of massive image data

The analysis of Fig. 5 shows that, with the continuous change in the number of iteration steps of the experiment, the mining accuracy of different methods is also constantly changing. However, the accuracy curve of the method in this paper is always above the accuracy curve of the method in reference [4] and the method in reference [6], which proves that the performance of this method is better than that of traditional methods. This is because the method in this paper achieves parallel mining of massive image data in Map/Reduce model by extracting and integrating multiple features of massive image data, which has the characteristics of high precision, high execution efficiency and short mining time.

5 Conclusions

In this paper, the problem of mining and recognizing massive image data in large-scale image database is studied, and a parallel mining algorithm of massive image data based on AI is proposed to collect and reduce noise of massive image data. To provide data information input for parallel mining model of massive image data, wavelet denoising is used to purify massive image data, and high-order moments and affine invariant moments of massive image data are extracted. Feature fusion and feature matching are realized in Map/Reduce model, and mining optimization is realized. The results show that the parallel mining of massive image data by using this method has high accuracy and good execution efficiency.

References

1. Zhou, S.B., Xu, W.X.: A novel clustering algorithm based on relative density and decision graph. Control Decis. **33**(11), 1921–1930 (2018)
2. Zhu, Y., Zhu, X., Wang, J.: Time series motif discovery algorithm based on subsequence full join and maximum clique. J. Comput. Appl. **39**(2), 414–420 (2019)
3. Wang, Z., Huang, M., et al.: Integrated algorithm based on density peaks and density-based clustering. J. Comput. Appl. **39**(2), 398–402 (2019)

4. Liu, Y., Yang, H., Cai, S., Zhang, C.: Single image super-resolution reconstruction method based on improved convolutional neural network. J. Comput. Appl. **39**(5), 1440–1447 (2019)
5. Xu, R., Zhang, J.G., Huang, K.Q.: Image super-resolution using two-channel convolutional neural networks. J. Image Graph. **21**(5), 556–564 (2016)
6. Megha, G., Yashpal, L., Vivek, L.: Analytical relation & comparison of PSNR and SSIM on Babbon image and human eye perception using Matlab. Int. J. Adv. Res. Eng. Appl. Sci. **4**(5), 108–119 (2015)
7. Li, Y.F., Fu, R.D., Jin, W., et al.: Image super-resolution using multi-channel convolution. J. Image Graph. **22**(12), 1690–1700 (2017)
8. Dong, X.W., Yao, S.M., Wang, Y.W., et al.: Virtual sample image set based multi manifold discriminant learning algorithm. Appl. Res. Comput. **35**(6), 1871–1878 (2018)
9. Gui, J., Sun, Z.N., Jia, W., et al.: Discriminant sparse neighborhood preserving embedding for face recognition. Pattern Recognit. **45**(8), 2884–2893 (2012)
10. Dong, B., Chen, D., Jing, W.: Salient region detection method based on symmetric region filtering. Comput. Eng. **45**(5), 216–221 (2019)