X-ray Coal Purity Examination Base on Particle Swarm Optimization Clustering

Long Hao [a], Huo Na[b],Yang Yong[c] ,Yu Bencheng[d]

Xuzhou College of Industrial Technology,Jangsu,China,

[a]longhhao@163.com, [b]longhhao@qq.com, [c]24546046@qq.com, [d]24546046@qq.com.

Keywords: coal purity; particle swarm; clustering analysis; X-ray; fitness function

Abstract: the paper mainly researched on the usage of particle swarm in image clustering. A fitness function is designed based on the principle of minimum in cluster distance and maximum between cluster distance. And particle swarm optimization, which is combined with fuzzy clustering, is used in image clustering. A system of X-ray coal image purity detection is developed and tested. According to the intensity distribution of X-ray images, an area of interest is selected based on Gaussian distribution property which greatly cut down the data to be processed. On the basis of that, particle swarm clustering is applied, and impurities are selected based on the fact that impurities are always tiny intensity elements. Experiments show that the approach is effective.

Foreword

With the rapid development of computer technology, the fieldbus technology and image processing technology, machine vision detection has become an integral part of modern industrial production as a kind of new detection technology. The machine vision' detection on impurities of coal can improve the quality of coal and utilization efficiency of cola, reducing coal emissions and improving the competitive ability of the products[1][2].

Image processing technology and method, as the core of the machine vision detection, directly influence the performance of system decision. To divide the content of the image into different parts is one of the most important steps in visual inspection, and it can be realized through image segmentation and clustering. The clustering method is based on particle swarm optimization and the ability to use of X-ray to penetrate are effective methods to detect the interior composition of the objects . Impurities in the X-ray image of coals show up the gathering of some dark pixels, while the surrounding coal shows shallow color. The impurities are commonly small or the grayscale differs rapidly, so the impurity of coal can be detected by X-ray image[3][4]. According to the principle of minimum-in-cluster-distance and maximum-between-cluster-distance, the paper designs the fitness function of particle swarm method, combining it with fuzzy clustering method, using X-ray coal impurity detection based on the clustering method, and verifying the effectiveness through the simulation experiments.

Particle swarm optimization method

Particle swarm optimization (PSO) algorithm is to realize the optimal solution of the complex space search through collaboration and competition among the individuals. In each generation of the algorithm, the particle will trace two extreme values. One is the optimal solution 'pbest' found by the particles themselves so far, and the other is the optimal solution' gbest' of total population which is found out so far. The mathematical description is: in a n dimensional search space, m particles constitute the population $X=\{x_1,\cdots x_i,\cdots,x_m\}$, and the ith particle's position was set as $x_i=(x_{i1}, x_{i2},\cdots,x_{in})^T$. Its velocity is $v_i=(v_{i1},v_{i2},\cdots,v_{in})^T$ and individual extremum is $p_i=(p_{i1},p_{i2},\cdots,p_{in})^T$, while the global extremum is $p_g=(p_{g1},p_{g2},\cdots,p_{gn})^T$. According to the principle of following the current optimum particles, the particles will change the speed and position based on the following formula:

$$v_{id}^{(t+1)} = v_{id}^{t} + c_1 r_1 (p_{id}^{t} - x_{id}^{t}) + c_2 r_2 (p_{gd}^{t} - x_{id}^{t}) \quad (1) \qquad x_{id}^{(t+1)} = x_{id}^{t} + v_{id}^{(t+1)} \quad (2)$$

Where, d=1, 2, n，i=1,2,⋯m and m represents population size, t represents current evolution algebra, r1and r2 are random numbers within the range [0, 1] ,c1 and c2 are accelerating constants[5][6].

Detection algorithm and implementation of coal impurity

Coal contains a lot of impurities composed of organic and inorganic matters and some wood. The paper uses X-ray to detect various impurities inside the coal. When X-ray goes through the objects with different intensity and density, its attenuation degree varies. Then the strength changes and forms images with different grayscales after induction imaging. Impurities in the X-ray image of coal show the gatherings of some dark pixels, while the surrounding coal shows shallow color and the general shape of the impurities is small. So the impurity can be detected by X-ray image. But the uneven density distribution will also show a deep color pixel gathering which caused difficulties to coal impurities image recognition. Therefore, to distinguish whether the dark edge is caused by the impurities or uneven density is the key to realize the coal impurity detection. In addition, on the premise of guaranteeing detection accuracy, improvement on the detection speed is very important for real-time detection.

Image preprocessing and region of interest selection

Fig 1 shows f coal X-ray image. According to the principle of X-ray imaging, because the impurities density is higher than that of coal, the figure shows the grey dark pixels gatherings. In Figure1, the background is white. The pixel gray value corresponding to thin coal area is higher while the grey value corresponding to the thicker coal area is low. Therefore, the difficulty of impurity detection lies in how to distinguish the dense coal area with similar coal impurities and grey value.

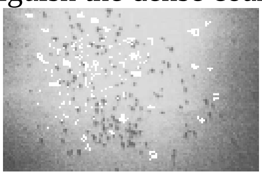

Fig1 coal X-ray image

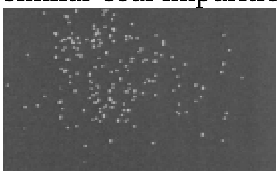

Fig2 region of interest image

Steps of image preprocessing and region of interest selection are as follows: (1) Filter the noise. (2) Mathematical morphology linear enhancement. (3) Choose region of interest.

Because the image grayscale distribution can generally be expressed by a single Gaussian distribution or combination. In view of the tiny proportion of the pixel gray-scale corresponding to the coal impurity, the relations between the area proportion encircled by Gaussian distribution curve and the variance can be mentioned. According to the fig 3, the region with the grayscale average mean $\pm 1.96\sigma$ variance accounted for 95% of the total area of the image, corresponding to larger proportion without impurities. Subtract it from the enhanced image G, then the remaining part is the region of interest containing impurities and mixed region. The fig2 shows the region of interest after processing. As it showed, the interested area accounts for only 5% of the total image which greatly reduce the data to be processed after this treatment.

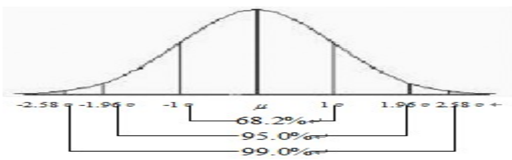

Fig 3 Gaussian distribution curve and relations between the area proportion encircled and the variance

Feature extraction

After the above processing, the location of the area including coal impurities is obtained. Set the coordinates of the pixels in the area as L, and then the analysis is as follows depending on the characteristics of coal impurity:

Firstly, coal area and coal impurities have the spatial continuity, and the coordinates of pixels can reflect the spatial relations between the pixels. The closer pixels form a shape or area, then to a category, and the farther pixels are classified into other different categories. Secondly, coal impurities generally show up as a tiny point with darker gray, and the pixel-value differs apparently compared with that of the surrounding pixel, then the gray level gradient value is large. The gray level gradient value of the dense coal area changes gently and the gradient values are small. So, if selecting the gradient value of gray scale as a feature, we can distinguish between coal impurity and dense area. So, the position and gray level gradient of pixels in the region of interest are chosen as clustering features.

Coal impurity detection based on clustering method

After feature extraction, the pixel in the region of interest can be clustered. The clustering center is considered as a set of solution which can realize minimum in cluster distance and maximum between cluster distance, and the particle swarm optimization method is used in this process. Each pixel in the region of interest is set as Ai, and the pixels is a vector including two feature s, one is the pixel coordinates L, the other is the pixel gray gradient value G. Then the clustering detection process is as follows:

The first step is initialization. Feature extraction is performed to get the initial clustering center Zit（i=1,2,⋯M）, and the gathering of the initial clustering center is set as the initial particle swarm. Initialize the location VI, velocity xi of each particle, and the individual optimal value pbest and group optimal value gbest are also initialized under the condition that the cycle index t = 1.

Step 2 is to calculate the individual and global fitness values. Fitness function is determined according to the principle of minimum-in-cluster-distance and maximum-between-cluster-distance. Calculate the distance Dpit from each point to the particle in each category with formula (3) and calculate the sum of all the in-cluster-distances Dt, the sum of all between-cluster-distances DBt with formula (4). Formula (5) is used to calculate each particle's individual fitness function value Fpit and global fitness function value Fgt.

$$Dp_i^t = \sum_{k=1}^{k} \left| A_i^k - Z_i^t \right| \qquad (3)$$

$$D^t = \sum_{i=1}^{M} \sum_{k=1}^{K} \left| A_i^k - Z_i^t \right| \qquad DB^t = \sum_{\substack{i \neq j \\ i=1,\, j=1}}^{M} \left| Z_i^t - Z_j^t \right| \qquad (4)$$

$$Fp_i^t = \frac{1}{Dp_i^t} \qquad Fg^t = \frac{1}{D^t - DB^t} \qquad (5)$$

Where K is the number of elements in each category and L_k^i is the Kth points in the ith category.

Step 3 to update the individual optimal values and global optimal value according to the fitness function value abstained. Compare the individual fitness function value of the above particle and that of the individual optimal pbest, if the fitness function value of the article is greater than that of the optimal pbest, this particle is considered as the individual optimal value pbest; And if the global fitness function value is greater than that of the global optimal gbest, the particle is a new global optimal individual.

The fourth step is to update the speed Vi

$$V_i^{(t+1)} = wv_i^t + c_1 r_1(t)(pbest^t - Z_i^t) + c_2 r_2(t)(gbest^t - Z_i^t) \qquad (6)$$

Where $\omega \geqslant 0$ is the inertial factor, and the larger ω helps to jump out of local maximum points while the smaller ω is advantageous to the algorithm convergence . c1、c2 area nonnegative constants, becoming the learning factors, and are generally set as $c_1=c_2=2$. $r_1(t)$、$r_2(t)$ are random numbers evenly distributed in the interval (0, 1]. In order to avoid the algorithm premature convergence to local optimal solution caused by the high particle speed, we commonly set constant $V_{max}>0$ in the algorithm according to this problems and ensure the absolute value of each component of $V_i(t)$ shall not exceed V_{max} by intercepting threshold. The larger V_{max} value can guarantee the global searching ability of particle swarm while the smaller $V_{ma\ value}$ is to reinforce the local searching ability.

Step 5 is to update the particle swarm $Z_i^{t+1} = Z_i^t + V_i^{t+1}$ （7）

Step 6 if |gbest(t+1)-gbest(t)|< ε and |gbest(t)-gbest(t-1)|< ε or the number of iterations reached the largest, it is over. Otherwise, go back to Step2.

After clustering, since the coal impurities are generally small, the coal category can be detected. Because the impurities area is smaller and coal area is larger, the categories with huger area can be removed. For the rest of the categories, the point-surface proportion can be calculated by the following formula.

$$ratio_i = \frac{\max(r) - \min(r)}{\max(c) - \min(c)} \qquad (8)$$

Max (r) is the maximum row coordinates while min (r) is the minimum one. Similarly, Max (c) is the maximum column coordinates and min (c) is the minimum column coordinates. By adopting the point-surface proportion, the coal impurities with tiny dot can be detected.

Experimental results and analysis

The experimental data are 3000 X-ray images of coal collected through X-ray imaging device and image acquisition card, and coal impurity detection algorithm research and development is done under offline state. The paper set the clean coal with preliminary washing and detection as samples. In the experiment, because the grey and the background of the impurities differ, the most commonly used threshold method and edge detection method is adapted. However, it was found that the two methods are not suitable for this type of image detection. One is because the impurity part is very tiny which is likely to be regarded as noise or be interfered by the noise points. Besides, due to the influence of the intensity changes, the dense area grayscale is close to that of the coal impurities, so it is difficult to distinguish between them with threshold method.

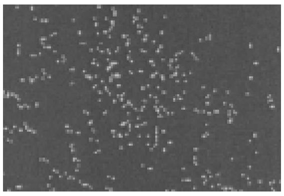

Fig 4 Mean shift detecting results

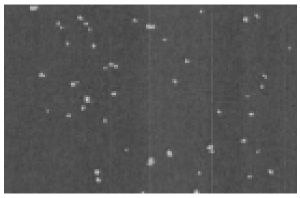

fig 5 PSO clustering detecting results

The experimental result of the PSO and that of mean shift clustering method were compared, and the two results are shown in figure 4 and figure 5. In figure 4, the white part is the testing result. As you can see, just the coal boundary with low grey value and the area with low grey value caused by coal intensity changes are detected, but not the coal impurities. With PSO clustering Method, the impurities are successfully detected and the detecting time lasts just 7.3466 seconds per image with the detection rate of 90%.

Acknowledgements

I want to acknowledge the invaluable help of professor Su and professor Wang, who has given my constant consultant in my paper writing, and has guided me and commented throughout the whole process of the paper.

References

[1] Selim S.Z., Alsultan K. A simulated annealing algorithm for the clustering problem. Pattern recognition, 24(10), 1991, pp. 1003-1008.

[2] Yager R.R., Filev D.P. Approximate clustering via the mountain method. IEEE SMC, 2,1994, pp.267-278.

[3] Christensen S，Oppacher F．What can we learn from No Free LunchA First Attempt to Characterize the Concept of a Searchable Function[A]. Morgan Kaufmann，2001：1219-1226．

[4] Bo Li，Zhi-Yan Zhao，Ju-Dong Li．A hybrid algorithm for nesting problems[A]．Proceedings of the Second International Conference on Machine and Cybernetics[C]．Xi'an，2003：1424-1429．

[5] Abido M A．Optimal power flow using particle swarm optimizat [J]．Electrical Power and Energy Systems，2002，24，563-571.

[6] Xie XF，Zhang WJ，Yang ZL．A dissipative particle swarm optimization[A]．In：Proc.of the IEEE Int’l Conf.on Evolutionary Computation[C]．Honolulu:IEEE Inc.，2002：1456-1461．