Automatic Classification of Coarse Aggregate Particle Size Based on Light Gradient Boost Machine

Lili Pei¹, Ting Yu¹, Ruichi Ma²*, Wei Li¹ and Xueli Hao¹

¹School of Information Engineering, Chang’an University, Xi’an, Shaanxi, China
²Gansu Wantai Construction Engineering Co., Ltd., Lanzhou, Gansu, China
*Email: 2543487651@qq.com

Abstract. Aggregate plays an important role in the performance of asphalt mixture. In order to complete the intelligent online monitoring of asphalt aggregate quality, an automatic classification algorithm for coarse aggregate particle size based on LightGBM classification algorithm is proposed. Firstly, OpenCV is used to extract the two-dimensional morphological features of coarse aggregates, and then the correlation between these features and the classification of aggregates is analyzed. Finally, the accurate classification of coarse aggregate particles is achieved through grid search and cross-validation optimization model. The results show that the average classification precision of coarse aggregate can reach 84.7%. Compared with the classification method based only on image processing technology, the precision is increased by about 20%, and the efficiency of classification is greatly improved.

1. Introduction

The amount of aggregate occupies the first place in the composition of asphalt pavement, mainly playing the role of skeleton filling [1-2]. Different morphological features of aggregate have different influences on the durability, compactness, ease and other properties of asphalt pavement [3-5]. In general, standard square hole screens with different specifications are used to screen the aggregate morphological granularity so as to obtain the aggregate particle size and classification type. The domestic and foreign scholars have achieved some research results in the classification of aggregate morphological features and particle size in recent years [6-8].

Li et al. [9] proposed an index for evaluating aggregate particle morphological features, and developed a morphological feature research system for quantitative evaluation of roughness, convexity, angular parameters, fractal dimension and other indexes. Liu [10] et al. improved a three-dimensional high-resolution image Fourier transform interference detection system for the characterization of aggregate morphology, and detected the surface morphological features of aggregate such as sphericity, flatness ratio, elongation rate, angular and so on. Patel et al. [11] designed a conveyor belt based lab-level image collection system, extracted 18 color features from the images, and established support vector machine model for iron ore classification. Geng [12] et al. combined skeleton extraction and support vector machine (SVM) algorithm to propose a method for aggregate particle classification, which can effectively achieve particle size classification. Ji [13] studied the influence of fine aggregate grading on the construction workability and pavement performance of MA asphalt concrete. Pei [14] et al. integrated multiple correlation analysis methods and completed the classification of aggregate shape by optimizing the super-parameter combination training XGBoost classification model. Pei [15] et al. proposed a calculation model of aggregate particle size based on neural network algorithm to realize accurate calculation of coarse aggregate particle size.

The morphological features and particle size classification of aggregates are of great significance.
However, the existing methods have some technical gaps in particle size classification. Therefore, this paper proposed the LightGBM (Light Gradient Boost Machine) model for automatic classification of coarse aggregate particle size.

2. Aggregate Feature Extraction and Data Set Construction

2.1. Image Acquisition and Processing

The aggregate image was collected by the MER-500-14GM camera. The camera's resolution is 2592 × 1944, the pixel size is 2.2μm × 2.2μm, and the image data format is Mono8 / Mono10. A total of 1000 aggregate images were collected, including 300 aggregates for each of the 4.75 mm and 9.5 mm, and 200 aggregates for each of the 13.2 mm and 16 mm.

The collection will inevitably be affected by various factors, which will eventually lead to poor aggregate image quality. Therefore, it is necessary to perform distortion correction, morphological processing and other operations on the collected aggregate images. The specific process is shown in Figure 1.

![Figure 1. Collection and processing of aggregate image](image)

2.2. Image Feature Extraction

The basic features of the aggregate image are extracted, and a total of 30 types of morphological feature information including perimeter, area, circumscribed rectangle and fitted ellipse are obtained, and the original data set of aggregate geometric features is finally established. The schematic diagram of extracting some feature factors such as equivalent ellipse and equivalent rectangle is shown in Figure 2.

![Figure 2. Schematic diagram of aggregates main geometric features extraction](image)

2.3. Feature Selection and Normalization

Due to the uneven distribution of the original data of aggregate geometric features, some of the feature values are overall high, and some feature data fluctuate greatly. Therefore, in order to ensure the robustness of the final model, feature screening and normalization of the data are performed. First, use equation (1) to normalize the data to the interval [0, 1]. Secondly, use Pearson's correlation coefficient
to analyze the correlation between the two-dimensional morphological characteristics of coarse aggregates and the classification of aggregates. The relevant feature factors are obtained as the input of the LightGBM model.

\[ x^* = \frac{x - \min}{\max - \min} \]  

\[ \rho_{x,y} = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} \]  

3. Coarse Aggregate Particle Size Classification Based on LightGBM

3.1. Establishment of Aggregate Particle Size Classification Model

LightGBM algorithm is a distributed integrated learning algorithm based on GBDT data model. In the algorithm, regression tree is used as weak learner. Firstly, histogram algorithm is used to pre-sort the features and node expansion method is used to construct the tree. The current residual regression tree is then obtained by using the residuals of each predicted result and target value as the next learning target. Each tree learns the conclusions and residuals of all the previous trees, adding up the results of the multiple decision trees as the final predictive output. LightGBM model can obtain high precision under the condition of maintaining high computational efficiency.

The aggregate feature data set obtained in 2.3 was taken as samples for model training. The input is mainly the aggregate morphological feature index obtained based on 2d image processing technology, with a total of 19 characteristic factors. The output end is the real aggregate category. The establishment and training process of LightGBM model are shown in Figure 3.

![Figure 3. Automatic classification process of aggregate particle size](image)

Next, the parameters are optimized through grid search and cross validation. The key parameter settings of LightGBM algorithm are shown in Table 1.

| Parameter name         | Value range | Parameter interpretation                                      | Final parameter value |
|------------------------|-------------|--------------------------------------------------------------|-----------------------|
| learning_rate          | 0.01-0.3    | learning rate                                               | 0.2                   |
| max_bin                | >0          | The maximum number of bin that feature will store            | 50                    |
| n_estimators           | >0          | The number of iterations                                     | 700                   |
| bagging_fraction       | 0-1         | The percentage of data used per iteration                    | 0.9                   |
| feature_fraction       | 0.5-0.9     | The percentage of parameters that are randomly selected each iteration | 0.7                   |
3.2. Evaluation and Analysis of Experimental Results

The confusion matrix of the LightGBM model for the classification calculation results of the aggregate particle size is shown in Table 2.

Table 2. Automatic classification results of aggregate particle size test sets

| Confusion matrix | true 4.75 mm | true 9.5 mm | true 13.2 mm | true 16 mm |
|------------------|-------------|-------------|--------------|------------|
| Pred. 4.75mm     | 83          | 10          | 0            | 0          |
| Pred. 9.5mm      | 9           | 52          | 12           | 1          |
| Pred. 13.2mm     | 0           | 7           | 51           | 3          |
| Pred. 16mm       | 0           | 0           | 2            | 70         |

In order to better quantify the results, the confusion matrix can be used to calculate the evaluation indicators of the model classification results, which are Precision, Recall and F1-score, calculated as follows.

\[
\text{Precision} = \frac{nii}{n*} \\
\text{Recall} = \frac{nii}{n*} \\
F1\text{-score} = \frac{2\times\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}
\]

Among them, \(nii\) is the number of positive sample which has been correct predicted; \(n*\) is the number of samples predicted to be positive, \(n*\) is the number of all true positive examples, and \(N\) is the total number of samples. The higher the evaluation index value, the better the classification effect of the model.

The evaluation index values are shown in Figure 4. The abscissa in the figure is the various evaluation indexes, and the ordinate is the value of the evaluation index. The precision of 4.75mm, 9.5mm, 13.2mm and 16.0mm are respectively 93%, 75%, 73%, 97%, the overall classification precision reached 84.7%.

In order to verify the validity of the model, the calculation results of the model output are compared with the existing calculation methods of aggregate classification, such as the results of the equivalent ellipse short axis (Feret) model and the equivalent rectangle short side (Feret) model. The comparison
results of the evaluation indicators obtained are shown in the Figure 5. It can be seen from the figure that the proposed model has the highest classification precision and the smallest error, and all indicators are significantly better than the image-based single geometric model particle size classification method.

![Figure 5. Performance evaluation comparison with single geometric model](image)

### 4. Conclusion

This research has completed the automatic classification of aggregate particle size. By using digital image processing technology to correct image distortion and extract its morphological features, a data set of aggregate features is established. Then analyze the main morphological characteristics that can characterize the aggregate particle size, and finally establish the LightGBM aggregate particle size classification model. The results show that the average classification precision of the four types of aggregate particle size reaches 84.7%. The proposed automatic classification technology of aggregate particle size is of great significance for aggregate virtual screening and online quality monitoring.

### 5. Acknowledgments

The work is supported by National Key Research and Development Program (2018YFB1600202); National Natural Science Foundation Program (51908059, 51978071), Fundamental Research Funds for the Central Universities, CHD (300102240206, 300102249301, 300102249306)

### 6. References

[1] Zhang D, Huang X, Zhao Y. Investigation of the shape, size, angularity and surface texture properties of coarse aggregates[J]. Construction & Building Materials, 2012, 34(none):330-336
[2] Yan P, Zhang J, Fang Q, et al. 3D numerical modelling of solid particles with randomness in shape considering convexity and concavity[J]. Powder Technology, 2016, 301:131-140
[3] Komba J, Mgangira M B, Mohale L. Investigation of the Effects of the Type of Crusher on Coarse Aggregate Shape Properties Using the Three-Dimensional Laser Scanning Technique[C]// Geo-China International Conference. 2016:125-132.
[4] Zhaoyun Sun, Chaofan Wang, Xueli Hao, et al. Quantitative evaluation for shape characteristics of aggregate particles based on 3D point cloud data. 2020, 263
[5] Rong Gao, Zhaoyun Sun, Wei Li, et al. Automatic Coal and Gangue Segmentation Using U-Net Based Fully Convolutional Networks. 2020, 13(4)
[6] Liu Hanye, Sun Zhaoyun, Li Wei, Huyan Ju, Guo Meng, Hao Xueli. Evaluating Angularity of Coarse Aggregates Using the Virtual Cutting Method Based on 3D Point Cloud Images, IEEE Access, 2020.8
[7] Höhner D, Wirtz S, Scherer V. A study on the influence of particle shape and shape approximation on particle mechanics in a rotating drum using the discrete element method[J]. Powder Technology, 2014, 253(2): 256-265
[8] Moaveni M, Wang S, Hart J M, et al. Evaluation of Aggregate Size and Shape by Means of
Segmentation Techniques and Aggregate Image Processing Algorithms [J]. Transportation Research Record Journal of the Transportation Research Board, 2013, 2335(-1): 50-59

[9] Li Xiaoyan, Bu Yin, Wang Hainian, et al. Research on quantitative Evaluation Index of coarse aggregate morphological characteristics [J]. Journal of Building Materials, 2015, 18(03): 524-530

[10] Liu Y, Sun W, Nair H, et al. Quantification of aggregate morphologic characteristics with the correlation to uncompacted void content of coarse aggregates in Virginia [J]. Construction and Building Materials, 2016, 124: 645–655

[11] Patel A K, Chatterjee S, Gorai A K. Development of machine vision-based ore classification model using support vector machine (SVM) algorithm [J]. Arabian Journal of Geosciences, 2017, 10(5): 107

[12] Geng Chao, Bao Jing, Zou Peng, Wang Weibin. Method for particle characterization by using skeleton extraction and SVM classification [J]. China Journal of Highways, 2016, 31(11): 58-65.

[13] Ji Fangli. The Influence of Fine Aggregate Grading to Pavement Performance of MA Asphalt Concrete [J]. Highway traffic technology, 2019, 35(01): 41-46.

[14] Pei Lili, Sun Zhaoyun, Yu Ting, Li Wei et al. Pavement aggregate shape classification based on extreme gradient boosting [J]. Construction and Building Materials. 2020. 256.

[15] Pei Lili, Sun Chaoyun, Hu Yuanjiao, Li Wei, Gao Yao, HAO Xuei. Neural Network Model for Road Aggregate Size Calculation Based on Multiple Features [J]. Journal of South China University of Technology (Natural Science Edition), 2020, 48(06): 77-86.