Preliminary Study on Sapphire Color Grading Method Based on Automatic Clustering Algorithm of Color Space Features

Liujun Lin1,*

1Henan College Of Industry & Information Technology, Jiaozuo, Henan, 454000, China

*Corresponding author e-mail: gxxyllj1@haut.edu.cn

Abstract. Traditionally, the color grading of sapphire is mainly based on the naked eye judgment of the appraiser. This judgment standard is not clear enough, and the judgment result has a greater subjective influence, which affects the accuracy of the classification. In this study, the GEM-3000 ultraviolet-visible spectrophotometer was selected, and the color features of 180 sapphire samples were extracted and classified using the CIE1976 color space of the device. The Kmeans algorithm was used to cluster analysis of 140 samples, and the separability of the color space features of different color levels was verified, and the center sample of each color level was obtained. The Euclidean distance between the centers of the remaining 40 samples is calculated, and each color grade prediction label is determined, and the sapphire color is automatically classified based on this. The experimental results show that the accuracy of sapphire color classification using the above method is 97.5%, which confirms the effect and accuracy of the artificial intelligence method in sapphire color classification.

Keywords: CIE1976, Kmeans Clustering, Euclidean Distance, Sapphire Color Grading

1. Introduction

The current grading standard of the sapphire industry is a standard jointly issued by the State Administration of Supervision, Inspection and Quarantine and the National Standardization Administration in 2017. This standard clearly stipulates the specific classification standards of sapphire's hue, chroma, and brightness. To classify the color of sapphire, the colorimetric method under the prescribed standard lighting source is generally adopted. Sapphire hue characteristics can be divided into blue, greenish blue, and purple blue, sapphire lightness characteristics can be divided into three levels: bright, brightener, and general, usually expressed as V1, V2, V3. Sapphire chroma
characteristics can be divided into five levels, such as dark blue, brilliant blue, deep blue, blue, and light blue. If the sapphire chroma is brilliant blue or deep blue, but there is no slight purple hue in the sample, in this case Call it Royal Blue. If the two grades of sapphires show a slight purple hue, they are called Cornflower Blue. The chroma of the sapphire sample is consistent with a certain grade standard, then the sample chroma is the chroma grade of this grade. If the chroma of the sapphire sample is between the two grades of standard samples, then the sample has a lower chroma grade. If the chroma of the sapphire sample exceeds the highest chroma standard of the existing classification, the sample is still the highest chroma grade. If the chroma characteristics of the sapphire sample are lower than the current lowest chroma grade standard, the sample is determined to be colorless and has no chroma grade [1].

The color grading of sapphire in the sapphire industry is currently based on the subjective judgment of the appraiser. The appraiser's experience level, the color grading operation site environment and other factors will affect the judgment results, resulting in different results for the same sample due to different appraisers or different judgment environments. The appraiser’s color grading of sapphire needs to rely on the brain to direct the eye nerves to perceive the light reflected by the sapphire, and then analyze the light received by the eye and the refraction during the formation of the light, and finally get the color grading judgment result. In this process, many objective factors may affect the appraiser's perception and analysis, and ultimately affect the grading results. In recent years, scientific research institutions have developed ultraviolet-visible spectrophotometers, which can be used to measure the color of sapphire more objectively, and the results of grading the color of sapphire are more accurate, so it has been widely used. The UV-visible spectrophotometer can measure the wavelength of 360-780nm with the support of the UV correction system, and the time will not exceed 1s, which minimizes the interference of external factors during the color measurement process, and the measurement results are more objective and accurate. It can be seen that the sapphire color grading method selected in this paper is feasible. On the other hand, artificial intelligence technology has been widely used in various fields of society. The use of machine learning algorithms in the jewelry inspection industry can greatly improve the accuracy of the detection and save a lot of computing time. In this paper, the cluster analysis method is applied to the sapphire color feature analysis process to realize the automation and intelligence of sapphire color grading, and to ensure the accuracy of sapphire grading [2].

2. Sample color measurement and cluster analysis

2.1. Sample collection

In this study, 180 sapphires were selected as the research samples. In order to ensure the diversity of the samples and the validity of the research, all samples include different color grades, and the sample weight is controlled between 0.76-3.32 carats, including round, oval, rectangular, cushion. Each sample does not contain obvious colored inclusions. Among them, 140 sapphire samples were color graded by the author's naked eyes, including all the grades of sapphire existing color grading. The remaining 40 samples have been appraised by professional organizations such as GRS and Gübelin, and have a certificate of color grading appraisal [3].

2.2. Color measurement
The GEM-3000 ultraviolet-visible spectrophotometer was selected to construct the three-dimensional space of the CIE1976 standard colorimetric system and the GEM3-B1372 detector to perform color measurement on all sapphire samples. During the measurement, the reflection measurement mode is selected, the light source is CIE standard light source D50, and the observation angle should be 10. First scan the baseline with an integrating sphere, and then perform a measurement on each of the three sides of each sample, take the average of the three measurement results, and complete the spatial color spectrum collection of all sapphire samples [4].

In this study, the sapphire color measurement system is selected as a three-dimensional space containing uniform colors. Any point in the space represents a different color. The spatial distance between two points is the color difference of the two colors. If the spatial distance is equal, it means the color difference. In the same way, this is the meaning of the discussion of placing colors in a uniform color space. The three elements of color including hue, chroma, and lightness correspond to the three chromaticity parameters of h*, C*, and L*, and also conform to the description of color in gemology. Therefore, the CIE1976 standard chromaticity system can be used for this study Sapphire samples were characterized [5].

Construct a three-dimensional coordinate system for color description: L* represents the lightness of the color, the value is between 0-100, if the color is black, L* is 0, if the color is white, L* is 100; h* represents the hue of the color, a*Positive value is red, a* negative value is green, b* positive value is yellow, b* is blue. C* means chroma, which refers to the depth of color on the surface of an object. The values of a* and b* depend on hue (h*) and chroma (C*). A three-dimensional color object must contain three elements: hue, lightness and chroma. The circle in the horizontal range of the coordinate system represents the hue and coordinates of the object The vertical main axis of the system represents the brightness of the object, and the horizontal circle radius represents the chroma of the object. According to the characteristics of the coordinate system, three sets of data of L*, a*, and b* of the color of the tested sample can be obtained, that is, the brightness, hue and chroma of the color of the object. On this basis, cluster analysis of the sapphire color can be performed [6].

2.3. Sample cluster analysis

The basis of the K-means clustering algorithm is a hierarchical clustering algorithm based on distance, which can divide the data into K categories and minimize the error [7]. The index of this algorithm to measure the similarity is the distance, the increase of the distance is the decrease of the similarity, and the decrease of the distance is the increase of the similarity. This indicator can be expressed as:

\[ d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \]. The process of using the above algorithm is as follows:

1) Determine the range of data samples, select K samples from N data samples in this range to form the initial cluster center;

2) Calculate the distance between the selected K data samples and the cluster center, determine the closest cluster to the cluster center, and put all samples into it;

3) Perform clustering center calculation again;
4) Compare the calculation results of the two cluster centers. If the results are inconsistent, repeat the calculation. If the results are consistent, the iterative output stops.

In this study, 180 samples were selected and 140 of them were selected for cluster calculation. First, perform color feature extraction on 140 samples in sequence which is \( x_n = \{L^*, a^*, b^*\} \), construct a color feature training data set with the three sets of data, which is expressed as \( X = \{x_1, x_2, \cdots, x_{140}\} \). Then, normalized preprocessing of the data set to obtain \( x'_n = (x_n - \text{mean})/\text{std} \), thereby forming a new data set expressed as \( X' = \{x'_1, x'_2, \cdots, x'_{140}\} \). Use the Kmeans clustering algorithm to analyze this new data set. In the case of \( K=5 \), this data set will be automatically divided into 5 clusters, and each cluster has a cluster center, which is expressed as \( U = \{\mu_1, \mu_2, \cdots, \mu_5\} \). There were 25 color samples, 38 brilliant blue samples, 42 deep blue samples, 24 blue samples, and 11 light blue samples. After normalizing the center of the above five clusters, the results are as follows [8].

| Category label | L* | a*   | b*   |
|----------------|----|------|------|
| Dark blue      | 40 | 1.06 | 1.89 |
| Brilliant blue | 42 | 2.77 | -26.98|
| Deep blue      | 40 | 1.45 | -9.50|
| Blue           | 48 | -1.24| -8.35|
| Light blue     | 82 | -L77 | -8.23|

The remaining 40 samples with identification results are selected to form the test set. Figure 1 shows the color grade distribution of the test set samples. The normalized preprocessing is performed on the color feature value of each sample, and it can be obtained that \( y'_n = (y_n - \text{mean})/\text{std} \). The euclidean distance between the centers of the 5 clusters is calculated, and the formula is: \( d = \|y_n - \mu_i\| \), \( i=1,2,\ldots,s \), and finally the predicted label is determined to be the smallest after calculation distance clusters. After the above steps are completed, the color grading of the test set samples is automatically completed. Comparing the results of automatic grading and manual grading, there is only one case where the grading does not match. The classification accuracy of this paper is 97.5%. The analysis found that the Euclidean distance between the centers of the wrong sample clusters is at the boundary of deep blue and brilliant blue. Inaccurate prediction is a normal phenomenon [9].
In order to improve the practical application value of the results of this study, figure 2 shows the automatic grading of sapphire color using this method. Following this process, the color of sapphire samples that are closer to the center of the color cluster can be automatically graded. For samples that are very close to the cluster boundary, manual operation can be combined [10].

3. Results and discussion

The sapphire samples selected in this study have a large dispersion in different color grades, and support automatic color grade grading. On this basis, the Kmeans cluster analysis method can effectively ensure the accuracy of grading and work efficiency, with strong repeatability, and provide a guiding theory for sapphire color grading.

In the process of collecting the color characteristic data of the sapphire sample in this study, the error of the measured values on the three sides of the sample did not exceed 0.1. The error rate was
rubies gradually verify insufficient. samples analysis show all reduced by increasing the number of measurements to obtain the average value. The characteristics of all samples are close to normal distribution, so there is statistical significance. The research results show that the intra-class difference of the samples does not exceed the inter-class difference, indicating that the Kmeans cluster analysis is very little affected by measurement errors and the analysis results are valid. However, due to the limitations of objective conditions, the number of samples in this study is limited, and the analysis of other factors affecting the color of the samples is insufficient. It needs to be further discussed in the next step. More samples need to be collected to verify the effectiveness of this classification method, and more The feasibility of the algorithm has gradually confirmed the application value of the clustering algorithm in the automatic color grading of rubies and other types of gems.

References

[1] Lu Shujing(2019). Gemstone identification method and empirical analysis based on convolutional neural network. Chengdu University of Technology.

[2] Zhang Jian, Bao Wei, Lu Taijin, etc.(2017). GB/T32862-2016 Sapphire Classification. Beijing: China Standard Press.

[3] Li Xiaoxiao(2015). The color origin and spectrum characteristics of sapphire in Changle, Shandong [J]. Postgraduate Journal of Sun Yat-sen University (Natural Science. Medical Edition), vol.36, no.4, pp: 60-73.

[4] Zhang Jian, Lu Taijin, Chen Hua(2015). Analysis of the National Standard of "Sapphire Grading" [C]. Proceedings of China Jewelry Academic Exchange Conference. Jewelry Management Center of Ministry of Land and Resources of China Jewelry and Jade Industry Association, pp:62 -64.

[5] Han Wen, Bi Lijun, Ke Jie, etc.(2017). Artificial intelligence + jewelry detection: laser-induced breakdown spectroscopy combined with support vector machine method to determine the origin of Hetian jade. Proceedings of China Jewelry Academic Exchange Conference 2017. Beijing: Geological Publishing House , pp: 335-336.

[6] Zhang Lili, Yuan Xinqiang(2017). Computer color grading of green jadeite based on HSL chromaticity parameters. Journal of Gems and Gemology, vol. 19, no.3, pp:45-51.

[7] Zhou Zhihua(2016). Machine learning. Beijing: Tsinghua University Press, pp:144-152.

[8] Li Jian, Jiang Nan, Baoyinbat, Zhang Fan, Zhang Weijian, Wang Wei(2020). Spatial color clustering algorithm and its application in image feature extraction. Journal of Jilin University (Science Edition), vol.58, no.3, pp:627-633.

[9] Sun Xiujuan(2019). Research on Canopy-Kmeans Parallel Clustering Algorithm on Cloud Computing Platform. Modern Electronic Technology, vol.42, no.19, pp: 78-81.

[10] Bian Yongming, Gao Fei, Li Mengru, Li Qiao, Ma Xiaoyang(2020). Flame detection method combining Kmeans++ clustering and color geometric features. Chinese Journal of Engineering Machinery, vol.18, no.1, pp: 1-6.