

# 基于 K-均值聚类的岩芯偏振显微图像粒径分析

陈本廷<sup>1</sup>, 周 骛<sup>1</sup>, 蔡小舒<sup>1</sup>, 徐喜庆<sup>2</sup>

(1. 上海理工大学 能源与动力工程学院, 上海 200093;

2. 中国石油大庆油田有限责任公司 勘探开发研究院, 大庆 163712)

**摘要:** 针对岩芯图像的粒径分析提出了一种基于 K-均值聚类算法的半自动分割算法,并编写了一套颗粒粒度图像处理程序. 首先将超像素处理概念应用于岩芯偏振显微图像,得到过度分割的结果,然后对分割结果进行 K-均值聚类 and 区域融合,利用图像中的边缘信息得到了更合理的结果,并大大提高了运算的速度;根据提出的算法,基于 VB.NET 2008 平台构建了一套半自动岩芯图像粒度分析软件,集图像采集、图像处理、粒度参数分析、砾石种类分类以及测量报告输出等功能于一体,大大提高了岩芯粒径分析的工作效率.

**关键词:** K-均值聚类; 岩芯; 偏振显微图像; 图像分割; 粒径分析

**中图分类号:** TP 319

**文献标志码:** A

## New Method for the Segmentation of Polarizing Microscope Image of Rock Core Based on K-Means Cluster Algorithm

CHEN Benting<sup>1</sup>, ZHOU Wu<sup>1</sup>, CAI Xiaoshu<sup>1</sup>, XU Xiqing<sup>2</sup>

(1. School of Energy and Power Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;

2. Exploration and Development Research Institute, Daqing Oil Field Co. Ltd., Daqing 163712, China)

**Abstract:** To conduct the core image particle size analysis, a semi-automatic segmentation method based on K-means clustering algorithms was proposed and a set of grain size image processing programs was formed. In the image processing, a superpixels algorithm was applied to process the polarizing microscopic image of cores, and the excessive segmentation results were achieved. Then the K-means clustering and regional integration were conducted on the segmentation results. In this way, the speed of operation was greatly improved, and the edge information of images was utilized to obtain more reasonable results. Based on the VB.NET 2008 platform, a semi-automatic software, being prove with the functions of image acquisition, image processing, analysis of grain size parameters, types of gravel classification and measurement reporting, was built according to the proposed algorithm. The method greatly improves the efficiency of the core particle size analysis.

收稿日期: 2016-02-29

基金项目: 国家自然科学基金资助项目(51206112); 上海市科委科研计划资助项目(13DZ2260900)

第一作者: 陈本廷(1991-), 男, 硕士研究生. 研究方向: 图像处理与颗粒测量. E-mail: chenbenting@126.com

通信作者: 周 骛(1985-), 女, 讲师. 研究方向: 图像法颗粒与流场测量. E-mail: usst\_wzhou@163.com

**Keywords:** *K-means cluster algorithm; rock core; polarizing microscope image; image segmentation; particle size analysis*

岩芯是在石油钻探过程中用特殊的钻具从地下取出的岩石样品,是油气田勘探开发研究工作中最为重要的基础地质资料之一,通过对岩芯的观察、描述和研究,可以直接地了解地下岩层的岩性、岩相、物性和油气水的分布特征,为石油的勘探开发提供直接依据.对岩芯进行粒度分析是沉积学及石油地质学研究的一种常规手段,因为,岩芯中砾石等颗粒粒度数据是对油气盆地沉积环境分类和识别的主要依据,而且岩芯断面中的颗粒物构造和粒度分布情况蕴含了地底含油构造的信息,对沉积环境的研究具有重要意义.

随着计算机技术在石油地质领域的广泛应用,利用图像分析的方法对岩石薄片的粒度分析已经取得了一定的成果.但由于岩石薄片图像往往存在颗粒特征不明显、轮廓不分明情况,现有的图像处理算法一般通过对灰度化的图像进行二值化而获得分离的颗粒图像<sup>[1-2]</sup>,很难实现较好的颗粒分割.因此,针对典型的彩色岩芯偏振显微图像,需要进一步利用其色彩信息对岩芯图像进行颗粒分割和粒度分析研究.

## 1 针对岩芯偏振显微图像分割的色彩空间的选择

色彩空间描述是用一组数值表示颜色的抽象数学模型<sup>[3]</sup>.色彩空间从提出至今已经有上百种,例如,三原色光模型(*RGB*)和印刷四分色模型等.这些色彩空间各有利弊,因而也有不同的应用.目前最为人们熟知的色彩空间是三原色光模型,该模型将彩色图像用 *R*, *G*, *B* (红,绿,蓝) 这 3 个分量的值来表示.但是,由于这 3 个分量与亮度相关,只要亮度改变,3 个分量都会相应地改变.所以, *RGB* 色彩空间中像素的颜色距离与人眼的感知差距较大,是不均匀的色彩空间,不适合彩色图像的分割<sup>[4]</sup>.

*CIE L\* a\* b\** 是由国际照明委员会为了准确描述人眼可见的所有颜色而提出的一种色彩模型. *CIE L\* a\* b\** 色彩模型是由亮度 *L\** 和有关色彩的 *a\**, *b\** 这 3 个要素组成. *L\** 表示亮度 (luminosity), *a\** 表示从洋红色至绿色的范围, *b\** 表示从黄色至蓝色的范围. *L\** 的值域为 0~100, *L\* = 50* 时,就相当于 50% 的黑; *a\** 和 *b\** 的值域都是为 +127~-128,其中, +127*a\** 就是红色,渐渐过渡到

-128*a\** 的时候就变成绿色;同样地, +127*b\** 是黄色, -128*b\** 是蓝色.所有的颜色以这 3 个值交互变化所组成,其色彩空间模型如图 1 所示.

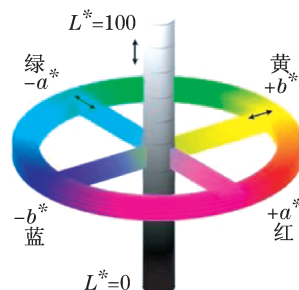


图1 *CIE L\* a\* b\** 色彩空间模型示意图

Fig.1 *CIE L\* a\* b\** color space model

与 *RGB* 和印刷四分色模型等色彩空间相较, *CIE L\* a\* b\** 色彩空间最主要的优点是它最接近于人类视觉,它致力于感知均匀性.因此, *CIE L\* a\* b\** 色彩空间更符合人眼的视觉感知特性,便于应用于图像处理,本文对彩色岩芯偏振显微图像的分割基于 *CIE L\* a\* b\** 色彩空间.

用数码相机采集到的彩色数字图像通常是以 *R*, *G*, *B* 分量的形式保存的,因此,基于 *CIE L\* a\* b\** 色彩空间对彩色数字图像进行图像处理之前需要进行 *RGB* 色彩空间到 *CIE L\* a\* b\** 色彩空间的转换.

*RGB* 和 *CIE L\* a\* b\** 色彩空间之间没有直接的转换公式,本文采用 *XYZ* 色彩空间作为中间层<sup>[5]</sup>, *RGB* 色彩空间到 *XYZ* 色彩空间的转换公式为<sup>[6]</sup>

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

*XYZ* 色彩空间到 *CIE L\* a\* b\** 色彩空间的转换公式为

$$L^* = 116f\left(\frac{Y}{Y_n}\right) - 16 \quad (2)$$

$$a^* = 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right] \quad (3)$$

$$b^* = 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right] \quad (4)$$

$$f(t) = \begin{cases} t^{\frac{1}{3}}, & t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3} \times \left(\frac{29}{3}\right)^2 t + \frac{4}{29}, & \text{其他} \end{cases} \quad (5)$$

式中:  $X, Y, Z$  是 CIE XYZ 色彩模型中的 3 个色彩分量;  $X_n, Y_n, Z_n$  为完全漫反射面的刺激值, 在本文中, 其值均为 1;  $t$  是函数表达式中的自变量。

## 2 岩芯偏振显微图像的分割

常规的岩芯图像分割方法一般先将彩色图像处理为灰度图像, 再采用灰度二值化方法得到分割后的颗粒图像, 但这样的方法在处理本文的岩芯偏振显微图像时往往难以获得理想的结果。其主要原因是对彩色图像灰度化实际上是对图像信息的退化, 丢失了图像的色彩信息, 仅利用了图像的灰度信息进行分割, 而且阈值分割的方法没有关注颗粒的边缘信息, 这也会导致分割误差。因此, 针对彩色的岩芯偏振显微图像, 非常有必要提出一种综合利用颗粒图像的色彩信息和边缘信息的分割方法。

### 2.1 K-均值聚类图像分割算法

聚类是对样本集数据进行分类的过程。K-均值聚类算法<sup>[7]</sup>是聚类技术中非常著名的一个硬聚类算法, 它算法简单、聚类速度快。与其他聚类算法一样, K-均值聚类也是一个迭代寻优的过程。K-均值聚类根据对象的某一特征, 认为每一个对象在这一特征空间中有一个对应的空间位置。通过迭代, 使

得各聚类本身尽可能地紧凑, 而各聚类之间尽可能的分开。这样, 同一聚类内样本的这一特征相似程度较高, 而聚类之间的相似度较低, 从而完成分割。

假设要将样本集数据  $\{x_1, x_2, \dots, x_n\}$  划分到  $k$  个集合中 ( $k \leq n$ ), 其算法步骤如下:

- 首先在特征空间内随机选取  $k$  个聚类中心;
- 然后计算各个样本到这  $k$  个聚类中心的欧式距离, 并根据最小距离将每个样本划分到对应的聚类内;
- 重新计算各个聚类的均值向量, 作为新的聚类中心。重复这一过程直到均值向量  $\mu$  收敛为止。

本文针对图像中各像素点的 CEI  $L^* a^* b^*$  色彩信息对各个像素进行 K-均值聚类。

图 2 是基于 K-均值聚类算法的典型彩色岩芯偏振显微图像结果及其与 Otsu 阈值分割法结果的对比。其中, 图 2(a) 是使用偏振显微镜拍摄得到的典型彩色岩芯显微图像; 图 2(b)~(f) 是应用本文算法对图 2(a) 进行聚类分割得到的各颜色层聚类; 图 2(g) 是将分割得到的各颗粒层合并之后得到的彩色颗粒层; 图 2(h) 是将彩色颗粒层进行二值化所得的分割结果; 图 2(i) 是对图 2(a) 进行灰度化后直接应用 Otsu 阈值分割法得到的二值化分割结果。其中, 对彩色颗粒层进行二值化时, 因为, 已经进行了

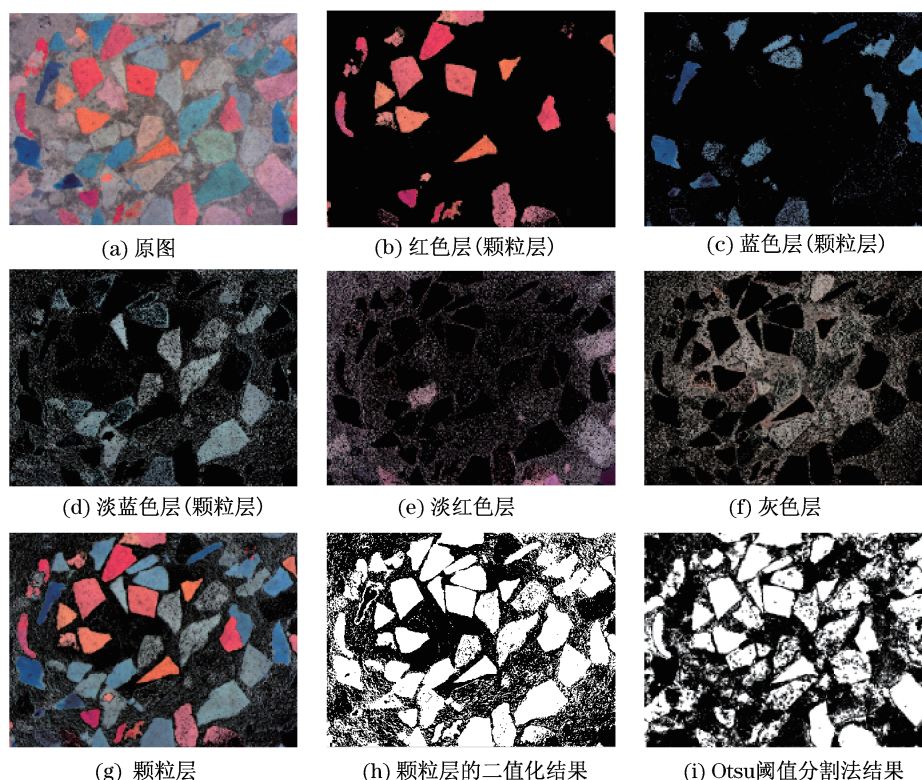


图 2 K-均值聚类图像分割算法典型结果及其与 Otsu 阈值分割法结果的对比

Fig.2 Comparison between image segmentation results by the K-means cluster method and Otus method

颗粒提取和筛选,认为除了黑色的背景之外,其他颜色均为颗粒,直接将其更改为白色,作为颗粒。

从图2可以看出,与经典的Otsu阈值分割法结果相比,颗粒的提取率和分割的完整度得到了很大程度的改善。对于颜色与背景色差异较大的颗粒(红色与蓝色),该算法可以得到良好的分割结果;但是,对于颜色与背景色接近的颗粒(淡蓝色、淡红色和灰色),分割结果中存在误分割的情况。此外,最终的分割结果产生了较多的噪声颗粒。这是由于算法本身过度依赖像素颜色,而且在分割过程中没有利用颗粒图像的边缘信息的缘故。

### 2.2 应用超像素对K-均值聚类图像分割算法的优化

本文提出的基于K-均值聚类的图像分割算法具有良好的分割效果,但是,对于图像尺寸较大的情况,其运算时间较长。另外,由于其分割过程仅仅依

赖于像素的色彩信息,没有利用颗粒图像的边缘和纹理信息,故存在错误分割的情况。针对算法上的这一缺陷,本文应用了超像素分割算法对其进行了优化。

超像素的概念最早在2003年由Ren等<sup>[8]</sup>提出,将图像过度分割成一系列子区域,每个子区域内部的某个特征具有很强的一致性。这样,可以将每一个子区域当成一个像素来处理,这样的子区域被称为超像素,应用超像素可以大量减少数据点数,且每个超像素集合含有特定的语义,利于后续其他应用。

根据超像素分割算法的特点,本文将超像素分割算法应用于岩芯图像,对其分割算法进行了优化。首先对岩芯图像进行超像素分割,将每个超像素区域内像素的色彩均值作为该超像素区域的特征值,最后对这些超像素区域进行K-均值聚类,得到分割结果。图3给出了应用超像素对岩芯图像分割算法进行优化前后的结果对比。

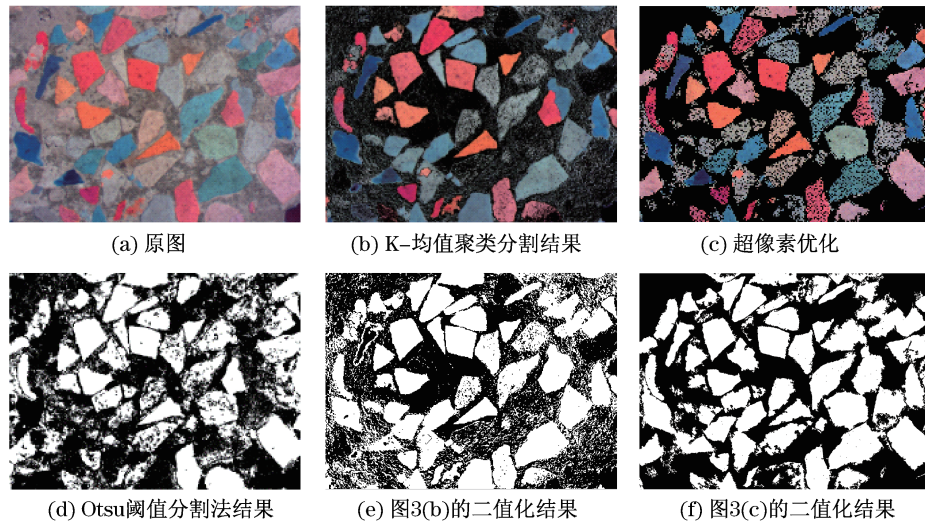


图3 应用超像素对K-均值聚类图像分割算法优化结果比较

Fig.3 Comparison between by the optimized image segmentation results by the K-means cluster method with and without the superpixels algorithm

对比图3(e)与图3(f)可以发现,通过应用超像素过度分割,在分割过程中利用了颗粒图像的边缘信息,超像素的边缘包含了颗粒边缘。因此,分割结果更加可靠,图像中的噪声以及背景对颗粒的影响被大大抑制。此外,超像素分割大大减少了像素个数,减小了计算量,从而很大程度上提高了运算速度。如表1所示,在Matlab R2012a的编程环境下,当聚类数均为6时,未进行超像素分割的图像(1 280×1 024)聚类所需时间为116.57 s,进行超像素分割后的图像(超像素个数为40 000)聚类所需时间仅为3.91 s。

表1 超像素分割对聚类时间影响对比表

Tab.1 Comparison between time cost of clustering with and without superpixels segmentation

| 聚类分割方法   | 图片尺寸        | 超像素个数  | 聚类数 | 聚类所需时间/s |
|----------|-------------|--------|-----|----------|
| 未进行超像素分割 | 1 280×1 024 | 0      | 6   | 116.57   |
| 进行超像素分割  | 1 280×1 024 | 40 000 | 6   | 3.91     |

## 3 岩芯图像粒度分析软件及砾石种类识别

基于上述超像素分割算法,针对石油勘探中的

岩芯粒度分析问题,本研究在 VB.NET2008 平台上构建了集图像采集、半自动化图像处理、砾石种类分类以及检测报告输出等功能于一体的岩芯图像粒度分析软件,可处理包括偏振显微图片、扫描图片和数码相机图片在内的各类岩芯图像,其主要功能和操作步骤如下:

a. 软件控制连接到偏振光显微镜的相机拍摄岩芯切片,得到彩色岩芯偏振显微图像,经过去噪等预处理后,通过图像处理,得出岩芯中颗粒的初始粒径数据,软件图像处理界面如图 4 所示;

b. 对初始粒径数据进行筛析校正和杂基校正,并绘制粒径分布直方图;

c. 分析粒径分布直方图,计算  $C, M$  等粒度分布参数,其中,  $C, M$  分别为累计频率曲线上 1%,

50% 处所对应的颗粒粒径,并分析给出颗粒中砾、砂、粉砂、泥等粒度分类所占百分率。

d. 根据石油天然气行业标准 SY/T 5434 - 2009,软件根据岩芯颗粒中砾、砂、粉砂、泥等的分布情况,自动进行岩石种类定名,图 5 为某种砂砾岩砾石的测量结果显示界面;

e. 输出检测报告。

本软件针对现有的石油勘探岩芯图像粒度分析过程中存在信息采集与图像分割处理自动化程度较低、效率较低、颗粒边缘提取准确性较差等问题,基于本文提出的图像分割算法,形成一套自动化程度较高的软件,可以准确地分割出与背景颜色相近的目标颗粒的边缘,并获得砾石、泥沙颗粒的粒度分布,对于了解含油构造的信息、研究沉积环境具有重要意义。

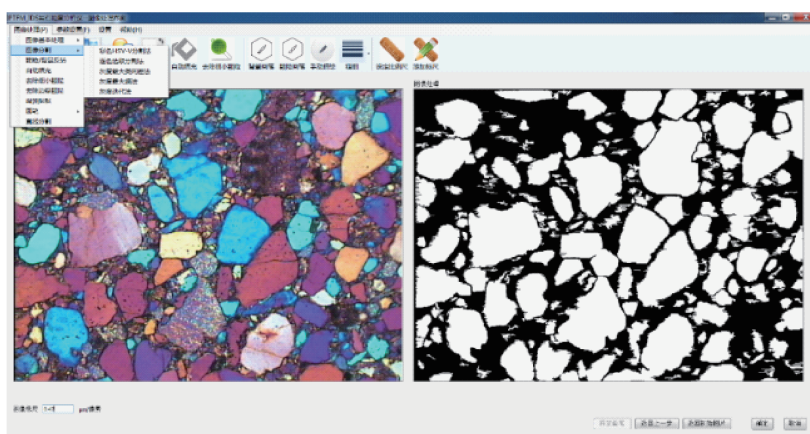


图 4 岩芯粒度分析软件图像处理界面

Fig.4 Image-processing interface of the core image size analysis software

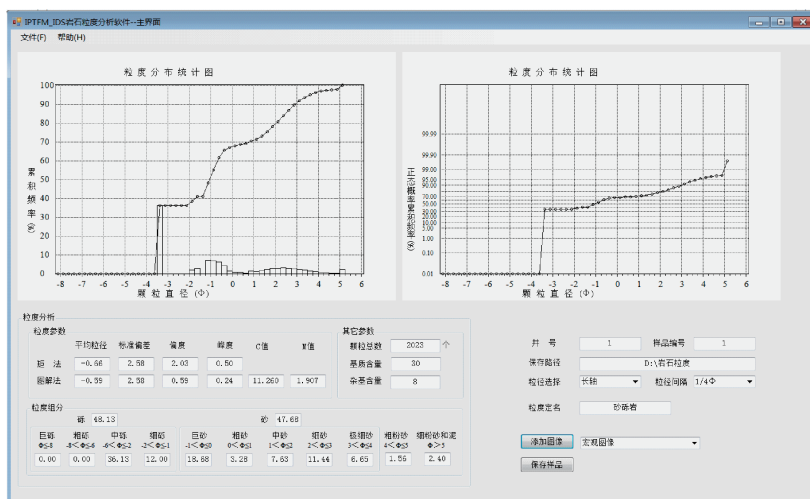


图 5 岩芯粒度分析软件粒度参数计算界面

Fig.5 Parameter interface of the core image size analysis software

- [M]. New York: Springer, 1998.
- [9] 王吉东, 许昌, 王欣, 等. 基于样本修整和支持向量机算法的并网风电机组运行特性研究[J]. 上海理工大学学报, 2014, 36(6): 532 - 537.
- [10] 王斐. 基于多核函数 SVM 的非平衡数据分类研究[D]. 镇江: 江苏大学, 2013.
- [11] 奉国和. SVM 分类核函数及参数选择比较[J]. 计算机工程与应用, 2011, 47(3): 123 - 124.
- [12] 周绍磊, 廖剑, 史贤俊. RBF-SVM 的核参数选择方法及其在故障诊断中的应用[J]. 电子测量与仪器学报, 2014, 28(3): 240 - 246.
- [13] 薛薇, 陈欢歌. 基于 Clementine 的数据挖掘[M]. 北京: 中国人民大学出版社, 2012.
- [14] 林关成, 李亚安. 基于 ANN 与 SVM 的分类和回归比较研究[J]. 声学技术, 2008, 27(4): 226 - 230.
- [15] BOWLES D E, MCPHEE S W J, LI C W, et al. Phase 1 gene therapy for Duchenne muscular dystrophy using a translational optimized AAV vector[J]. Molecular Therapy, 2012, 20(2): 443 - 455.
- [16] FINANGER E L, RUSSMAN B, FORBES S C, et al. Use of skeletal muscle MRI in diagnosis and monitoring disease progression in Duchenne muscular dystrophy[J]. Physical Medicine and Rehabilitation Clinics of North America, 2012, 23(1): 1 - 10.
- [17] ZHU B Y, CHEN H, CHEN B D, et al. Support vector machine model for diagnosing pneumoconiosis based on wavelet texture features of digital chest radiographs[J]. Journal of Digital Imaging, 2014, 27(1): 90 - 97.
- [18] 汪金涛, 高峰, 雷林, 等. 基于主成分和 BP 神经网络的智利竹筴鱼渔场预报模型研究[J]. 海洋学报, 2014, 36(8): 65 - 71.

(编辑: 石 瑛)

(上接第 345 页)

## 4 结束语

根据岩芯图像的特点, 选取  $CEI\ L^* a^* b^*$  色彩空间模型进行图像处理, 利用 K-均值聚类算法对岩心图像进行聚类分割处理, 该算法能够更有效地提取图像中的颗粒信息, 获得比较理想的分割图像, 从而得出更加准确的粒度分布; 针对 K-均值聚类算法耗时长、没有利用颗粒边缘信息等缺点, 应用超像素分割对 K-均值聚类图像分割算法进行了优化, 不仅有效利用了颗粒的边缘信息, 而且在保证分割效果的同时大大提高了算法的运算速度, 提高了该算法的实用性; 基于提出的彩色图像分割算法, 在 VB.NET2008 平台上搭建了半自动岩芯图像粒度分析软件, 集图像采集、图像处理、粒度参数分析、砾石种类分类以及检测报告输出等功能于一体。

### 参考文献:

- [1] 韩思奇, 王蕾. 图像分割的阈值法综述[J]. 系统工程与电子技术, 2002, 24(6): 91 - 94.
- [2] 陈文静, 蔡小舒, 陈孝震, 等. 基于 HSV 空间的岩心图像处理[J]. 中国粉体技术, 2013, 19(2): 1 - 6.
- [3] 李宋, 吴文权, 詹咏. 颜色渐变的算法研究[J]. 上海理工大学学报, 2004, 26(3): 224 - 228.
- [4] 黄飞, 吴敏渊, 曹开田. 基于 HIS 空间的彩色图像分割[J]. 小型微型计算机系统, 2004, 25(3): 471 - 474.
- [5] 胡太银. 面向色彩管理的可视化技术研究[D]. 西安: 西安电子科技大学, 2002.
- [6] TKALCIC M, TASIC J F. Colour spaces: perceptual, historical and applicational background [C] // Proceedings of the IEEE Region 8, 2003 Computer as a Tool. Ljubljana Slovenia: IEEE, 2003.
- [7] MACQUEEN J. Some methods for classification and analysis of multivariate observations[C] // Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability Berkeley, California: University of California Press, 1967, 1: 281 - 297.
- [8] REN X, MALIK J. Learning a classification model for segmentation[C] // Proceedings of the Ninth IEEE International Conference on Computer Vision, 2003, Nice, France: IEEE, 2003: 10 - 17.

(编辑: 石 瑛)