

浅析聚类算法在宝石颜色分级中的应用

中国地质大学(武汉)珠宝学院 赵天豪

摘要: 在绿松石和翡翠的分级中,颜色分级是非常重要的一个考虑因素。目前,对于绿松石和翡翠的颜色分级主要采取人工方法,十分费力且容易受主观因素影响。笔者尝试使用聚类的方法对绿松石和翡翠进行自动颜色分级,从而实现以一种客观的量化标准对二者颜色分级,同时减少人工时间消耗。实验结果表明,使用基于聚类的绿松石和翡翠自动颜色分类的方法可以在一定程度上实现对二者颜色的自动分类,但是不能完全取代人工分类,本方法可以作为一种前置的自动分类方法,为后续的人工分级提供一定的依据。

关键词: 绿松石;翡翠;聚类;图像处理;主颜色

DOI: 10.12433/zgkjtz.20232438

随着人工智能领域的发展,使用聚类算法对图像进行处理是聚类算法的一大应用方向。

基于绿松石和翡翠的颜色分级时存在的受主观因素影响、大批量样品分类费时费力等问题,根据现有的颜色分级标准通过计算机实现对于二者颜色的自动分类,使用客观的量化指标消除分类时的主观因素的影响以及对于分类环境的要求,也可以节约大量的绿松石或翡翠进行颜色分类的时间。

一、图像收集与预处理

(一)图像收集

本次实验所收集绿松石图像共计209张,颜色包括绿色、绿蓝色、蓝绿色、蓝色、黄色、橙色;翡翠图像共计133张,颜色包括绿色、紫色及红黄色翡翠。

(二)图像分割

在处理图像前,为了消除图像背景对于结果的影响,使用Grabcut的图像分割程序对绿松石和翡翠图像进行图像分割。

二、绿松石与翡翠颜色聚类分析

(一)量化标准的选取

在对绿松石和翡翠图像使用聚类方法进行颜色自动分类之前,首先要有自动颜色分类的量化依据。在《绿松石 分级》(GB/T 36169—2018)中给出了绿松石

的色相、饱和度和明度不同分类的HSV值的参考值,而在《翡翠 分级》(GB 23885—2019)中在对翡翠分级时,色相给出的供参考的量化指标是光谱色主波长参考值,彩度给出的供参考的量化指标是由CIE XYZ颜色空间得到的色纯度值,明度给出的供参考的量化指标是色卡灰度标尺参考值。考虑到翡翠分类国标中使用的量化指标不容易计算,笔者采用绿松石分级国家标准中使用的HSV值作为二者颜色分类的量化指标。

对于绿松石,直接使用绿松石国家标准的不同HSV值对应的分级类别进行分类。对于明度和饱和度参考值,笔者对原标准中的百分比进行相应的转换。对于两个类别的边缘值均认为属于上一类别。

对于翡翠,翡翠分类在国家标准中未给出色相、饱和度和明度不同级别对应的参考值。笔者依据绿松石国家标准中对于饱和度和明度不同级别的参考值对翡翠的饱和度和明度进行分类。翡翠色相类别不同且存在不能用色相分类的无色翡翠,因此,笔者暂时不对无色翡翠分类,对色相的翡翠依照opencv中的不同颜色对应H值划分色相类别与色相参考值,见表1。

表1 翡翠色相类别与对应色相参考值

色相类别	色相参考值 $H/^{\circ}$
紫	125 ~ 155
绿	35 ~ 100
黄(橙)	11 ~ 35
红	0 ~ 11, 156 ~ 180

(二)基于聚类的绿松石和翡翠颜色自动分类算法

笔者共设计了两个基于聚类算法的算法来实现绿松石和翡翠的自动颜色分类。两个算法的基本步骤如下:

1. 算法一

(1) 提取一张图像所有像素点的HSV值。

(2) 为了去除杂色影响根据H值,对像素点进行筛选,对筛选后的像素点的HSV值进行聚类。

(3) 定义含有像素点最多的簇为主颜色簇,提取主颜色簇的簇中心点作为该张图像的颜色特征值,对所

有经过图像进行如上操作提取出各种的颜色特征值。

(4) 对所有图像的颜色特征值再进行聚类, 提取这一次聚类的每个簇的簇中心点对应的HSV值。

(5) 根据HSV值所对应的颜色分类结果认为属于该簇内的所有特征值对应的图像均属于该颜色分类结果。

算法一流程示意图1。

2. 算法二

(1) 用户输入一张图像或者一个包含图像的文件夹, 提取读取到的图像的像素的HSV值。

(2) 在提取之前先对像素进行筛选, 筛选出 H 值不符合的像素, 排除背景颜色或无关颜色对后续处理的影响。

(3) 之后对筛选出的像素的HSV值进行聚类, 选出包含像素点最多的簇, 若该簇包含的像素点数量小于筛选后总像素点数量的40%就认为该绿松石或翡翠为杂色绿松石或翡翠, 若该簇包含的像素点数量大于筛选后总像素点数量的40%则进行后续处理。

(4) 提取主颜色簇簇中心点的HSV值及主颜色簇中随机99个像素点的HSV值。

(5) 根据《绿松石 分级》(GB/T 36169—2018) 对像素点的 H 、 S 和 V 值进行分成不同的类别(翡翠 H 值的分类依据不同颜色对于的 H 值范围), 然后将前面选取的100个点的HSV值进行分类。

(6) 统计每类中分到的像素点的数量, 认为该绿松石/翡翠的颜色分类即为包含像素点最多的类别对于的颜色分类。

算法二流程见图2。

(三) 基于算法一的绿松石和翡翠颜色自动分类

1. 基于算法一的绿松石和翡翠颜色自动分类结果展示

对所有绿松石图像的特征值进行聚类, 选用K-means++算法且聚类数为8时每个簇中心点的HSV值(表示为 H, S, V 三元组的形式)及HSV值对应的分类见表2。

从表2中可以看出, 当聚类数选为8时, 实际只将绿松石分为了(绿, 浓, 明)(黄, 艳, 暗)(绿, 浓, 亮)(黄, 艳, 明)(黄, 中, 明)和(绿, 艳, 明)共计6类。色相只分出了两个类别。出现了分类明显错误的情况, 绿松石被分到(黄, 艳, 暗)这一类中。聚类的DB指数为0.98, CH指数为133。

对所有翡翠图像的特征值进行聚类, 选用K-means++算法且聚类数为8时每个簇中心点的HSV值(表示为 H, S, V 三元组的形式)及HSV值对应的分类, 见表3。

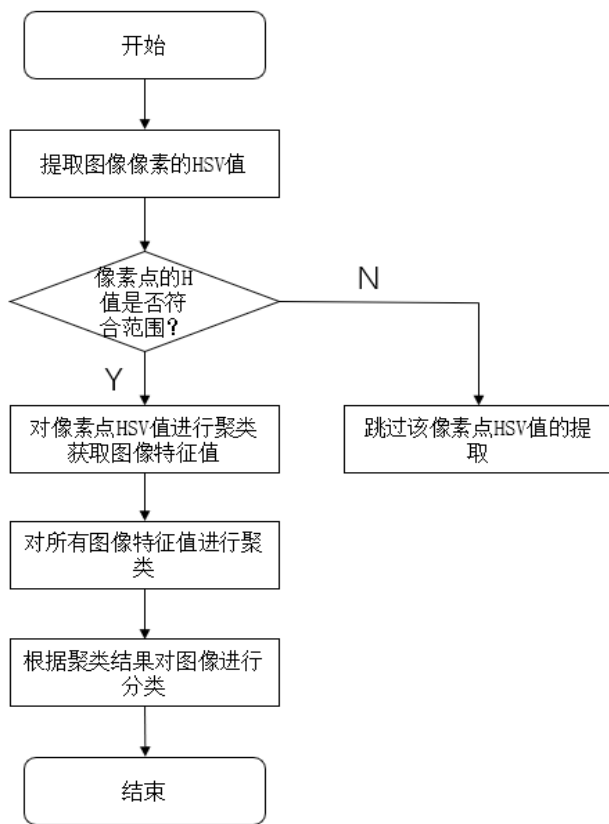


图1 算法一流程示意

从表3中可以看出这次将翡翠分为8类。色相分出了绿、紫、红三个类别, 饱和度分出了全部三个级别, 而明度分出了明和暗两个级别。聚类的DB指数为1.03, CH指数为85。

2. 算法一结果分析

使用算法一对绿松石和翡翠进行颜色自动分类, 结果表明, 该算法对于翡翠的分类效果较好, 能够分出更多级别的翡翠, 这可能是因为绿松石的色相类别较接近, 将翡翠分为4类。而翡翠的色相类别范围较大, 因此, 相对效果较好。另外, 使用DB指数和CH指数对二者的聚类结果进行评价, 虽然对绿松石进行聚类的DB指数和CH指数都优于对翡翠进行聚类的情况。实际是翡翠的分类效果更好, 这也说明在评价聚类效果时, 不能仅从聚类的评价指标对聚类效果进行评价。

(四) 基于算法二的绿松石和翡翠颜色自动分类

对所有处理后的绿松石和翡翠图像使用算法二进行颜色自动分类。算法二的分类效果较好, 可以将不同色相、明度及饱和度的绿松石和翡翠图像进行分类。

但是, 对于两类绿松石分类效果较差: 一类是绿松石铁线的存在导致其将该绿松石分为杂色; 另一类

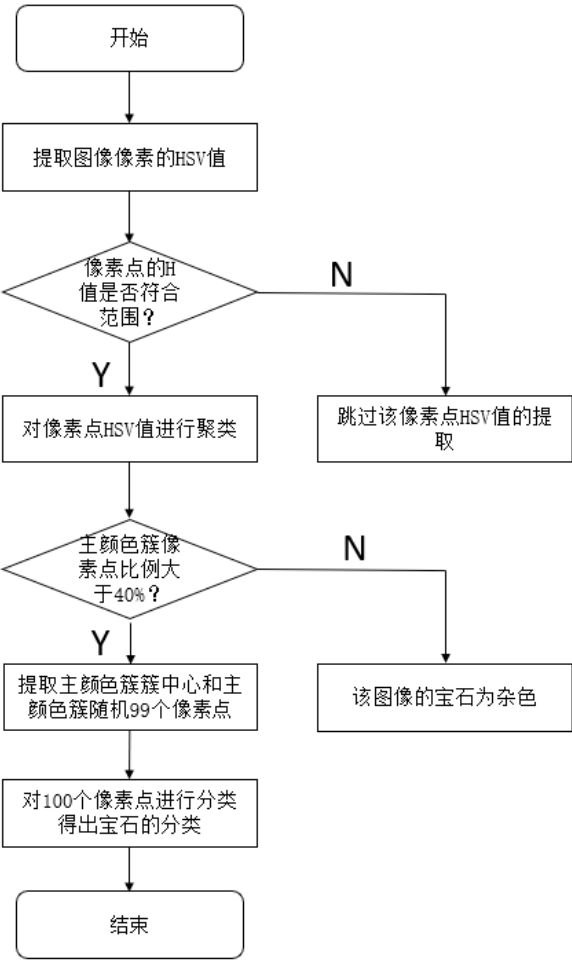


图2 算法二流程示意

表2 K-means++算法聚类数为8时每个簇中心点HSV值与对应颜色分类级别

簇中心点 HSV 值	对应颜色分类级别
(90,225,135)	(绿, 浓, 明)
(92,187,161)	(绿, 浓, 明)
(70,109,77)	(黄, 艳, 暗)
(92,219,183)	(绿, 浓, 亮)
(82,149,147)	(黄, 艳, 明)
(68,87,131)	(黄, 中, 明)
(92,144,196)	(绿, 艳, 明)
(75,90,182)	(黄, 中, 明)

是杂色绿松石可能会被分为纯色绿松石。原因在于算法中进行杂色的判断时阈值设置过低,因此,需要调

表3 K-means++算法聚类数为8时每个簇中心点HSV值与对应颜色分类级别

簇中心点 HSV 值	对应颜色分类级别
(77,141,94)	(绿, 中, 暗)
(68,185,157)	(绿, 浓, 明)
(56,232,111)	(绿, 浓, 暗)
(129,99,159)	(紫, 中, 明)
(61,92,165)	(绿, 中, 明)
(102,40,199)	(紫, 浅, 明)
(86,39,90)	(绿, 浅, 暗)
(7,214,174)	(红, 浓, 明)

高算法中对杂色的判断时的阈值,对被分为纯色的杂色绿松石进行再次分类。

结果表明,当阈值达到80时,所有杂色绿松石识别为杂色。说明在对绿松石自动颜色分类时,对于绿松石是否为杂色需要较高的阈值才能判断准确,但是此时可能将纯色绿松石识别为杂色。因此,对于杂色绿松石的判断可以先设置一定的阈值梯度逐步筛选出纯色绿松石后剩下的即为杂色绿松石。

而算法二对于翡翠图像的分类效果好于绿松石图像的分类效果。

三、结论

综合根据上述两种基于聚类算法的绿松石和翡翠颜色自动分类算法的实现思想、实现步骤和颜色分类结果,可以得出以下结论:

(1) 算法一思想及实现较为简单,但是对绿松石和翡翠的颜色自动分类效果较差。不过,因为使用算法一实际上不用预先得知最后聚类中的每个簇中心点HSV值对应的颜色分类级别,仅仅是根据宝石图像的HSV值的相似性将它们分类,所以当对一种没有根据HSV值颜色空间分类的宝石进行分类时,可以使用该算法粗略分类。

(2) 算法二实现相对复杂,但是对绿松石和翡翠的颜色自动分类效果较好,虽然仍存在一定程度的误差,但是可以作为人工分类前的前置分类使用。

参考文献:

[1]中华人民共和国自然资源部.绿松石 分级:GB/T 36169—2018[S]. 北京: 中国标准出版社, 2018.
[2]中华人民共和国国土资源部.翡翠 分级:GB/T 23885—2009[S]. 北京: 中国标准出版社, 2009.