

彩色等高线地形图矢量化的预处理方法^①

余翔, 孙琦, 王永敏

(西北核技术研究所, 西安 710024)

摘要: 地形图矢量化生成 DEM 的应用需求非常广泛, 由于地形扫描图像难免存在模糊及粘连现象, 提取法提取出的等高线效果均不理想. 结合形态学在 RGB 色域空间和 HSV 色域空间分别用剔除法剔除非等高线信息, 对分离出来的等高线图像进行 USM 锐化, 然后再进行灰度剔除和光滑处理, 利用剔除信息提出了一种断点匹配法对断点进行连接, 生成了较高质量的二值等高线. 整个预处理过程算法简单, 易于实现.

关键词: 等高线; 矢量化; USM 锐化; 断点匹配法

Preprocessing Method of Color Contour Map Vectorization

YU Xiang, SUN Qi, WANG Yong-Min

(Northwest Institute of Nuclear Technology, Xi'an 710024, China)

Abstract: DEM generated by topographic map vectorization is wide used in many applications. Because the vague and adhesion phenomena in scanned terrain image will inevitably exist, the results of the contour extracting from blur images are not satisfactory in most cases. Combining with morphological method in both RGB and HSV color space, several elimination methods are used to remove non-contour information, and the remains are color contour with many breakpoints. The UnSharpen Model(USM) is used to sharpen the contour line, then we use models to decrease the line width and to deblur the line in gray modes. By using the removed non-contour information, we propose a new breakpoint-matching method to connect the contour and generate more ideal binary contour with high-quality. The whole preprocessing algorithm is simple, easy to implement.

Key words: contour; vectorization; UnSharpen Model(USM); breakpoint-match method

随着经济、军事和社会需求的不断增长, 数字地理信息的应用越来越广泛和重要, 传统的纸质地形图经过扫描, 形成栅格地形图, 通过图像处理与识别, 采用半自动或者全自动矢量化方式完成地形的数字化. 张耿等^[1]将扫描后的栅格地图使用 Photoshop 淡化背景, 然后将栅格地图转换为 HSI 格式, 利用棕色等高线的 S 分量较高的特征, 提取出等高线, 等高线断裂严重. 钱静^[2]采用本征向量拟合方法分割出棕色等高线, 然后根据断裂方式的不同采用不同的算法进行连接. 侯萍^[3]、曾迎生^[4]等侧重于解决等高线的连接方法研究, 如 A*算法、直接短线法以及曲线重建法等. 上述方法在扫描质量较高、等高线特征明显、断裂不严重的情况下效果较好. 实际工作中扫描图像一般都存

在模糊与粘连, 采用上述方法难以得到质量较高的二值图像, 本文采用形态学方法剔除无用信息, 并采用 USM 锐化等高线、灰度剔除等系列方法, 可以从模糊粘连的扫描地形图中提取出高质量的二值图像, 有效减小矢量化工作量.

1 彩色等高线的识别与提取

姚伟新等^[5]提出了模式识别、模糊技术、样本学习和神经网络等分色方法, 这些方法只在一定程度上解决了彩色地图图像分色问题, 分色后的单版图效果不很理想. 夏春林等^[6]、王建宇等^[7]通过单一阈值方法或者灰度阈值提取其他颜色, 由于扫描图片一般存在颜色混淆和假彩色, 转换为灰度时不存在明显的灰度

^① 收稿时间:2014-07-24;收到修改稿时间:2014-08-25

差异或各物体的灰度值范围有较大重叠,难以有效分离出其他颜色.

在 RGB 色域空间中,视觉上相邻色彩之间的非线性关系使得颜色判别变得非常困难,HSV 是一种直观而且比较符合人体视觉对色彩感知特征的色域空间,其中绿色(含过渡色)植被和蓝色(含过渡色)水系的 H 分量较大,可以通过判别 H 分量的阈值来剔除绿色和蓝色.由于扫描设备、图件质量等原因的差异,扫描结果也不相同,可以依据 HSV 的直方图(见图 1)统计结果,适当修改 H 的阈值剔除植被和水系.通常情况下 H 的上限取 0.18 就可以有效剔除其他颜色,只留下黑色及棕色.确定 H 分量上限的依据是尽可能多地保留棕色(包括过渡棕色),特别是在粘连和模糊区域的棕色信息要尽量保留.

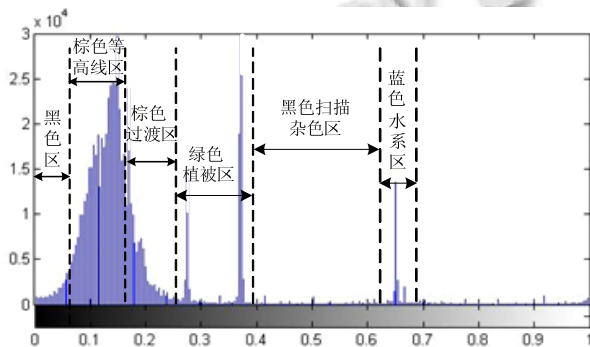
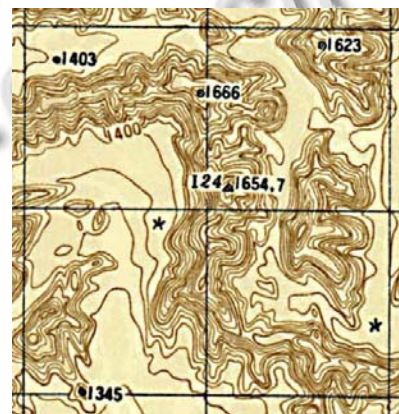


图 1 典型等高线地形扫描图 HSV 色域 H 分量直方图

剔除植被和水系后,剩余图像中的深棕色和浅黑色的 H 分量很相近,通过判别 H 分量阈值的方法很难得到理想的分隔效果.分离方法是在 RGB 空间中剔除 RGB 均大于 230 的背景色,将 RGB 均小于 20 的黑色提取出来,由于提取阈值很小,在边缘处会残留较多的浅黑色及杂色.通过对提取出的 RGB 均小于 20 的黑色图像进行形态学上的处理,可以得到较好的分离结果,步骤如下:(1)对提取结果进行二值化,从形态上分为两类:线性网格线和其他标注.(2)由于提取阈值较小,网格线的宽度不一,并且通常情况下也不连续,可以采用恢复性方法将网格线重新连接完整.方法是采用 3*3 大小的模板扫描图像的边缘,得到网格线水平方向和垂直方向的坐标,然后对坐标点按照大小顺序成对连接并进行修正,线宽为 3~5 像素,可以恢复出网格线,同时可以分离出其他标注,该网格恢复图可作为断裂区的判别基础.(3)对于分离出的地物

标注(如字母、数字、形状符号等),提取结果只是标注的大致轮廓,采用形态学上的腐蚀方法恢复标志的原始形状,腐蚀模板大小的选择依赖于图像扫描质量,一般情况下,扫描质量好的图像可选用 2*2 方形模板,差的可选用 3*3、4*4 或 5*5 方形模板.(4)将恢复的网格线和标注叠加,得到需要剔除的图像.(5)在剔除植被、水系和背景色的图像中继续剔除上述结果,可以得到最终的提取结果(图 2(b)).上述(1)~(5)步骤在 matlab 中很容易实现,不需要复杂的算法.



(a) 原始图像



(b) 剔除后的提取结果

图 2 扫描地形图的剔除、提取结果

2 等高线预处理

提取出的等高线包含大量断点,也存在粘连现象,需要进行锐化和细化预处理,加强等高线的线性特征,减少粘连现象,有利于进行矢量化处理.

2.1 基于彩色图像的 USM 锐化

锐化的目的是增强图像中颜色跳变部分,加强边缘特征. USM(UnSharp Mask)锐化原理是对原始图像进行低通滤波得到一个更加平滑的模糊图像,用原始

图像减去低通滤波后的模糊图像，再将结果(高频部分)叠加回原始图像，可以加强颜色跳变部分的强度。与其他低通滤波器相比，高斯低通滤波器对图像的模糊程度最低，对边缘带来的混叠程度最小，因此本文采用高斯低通滤波器进行滤波，见图 3。高斯函数形如：

$$g(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

其中 σ 用于控制平滑程度， $|x| \leq r, |y| \leq r$ 为影响区域大小。

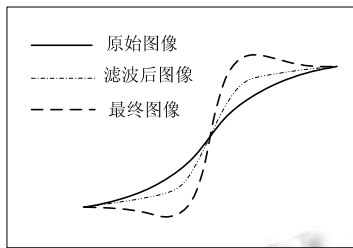


图 3 USM 锐化原理

USM 锐化对处理彩色等高线图的模糊和粘连区域效果较好，如果一遍处理结果不理想，可以采用不同大小的模板多次处理，直到达到理想的效果，如图 4，其中 r 为模板半径， n 为处理次数。USM 锐化是对彩色图像 RGB 三个分量均采用滤波方法，由于相邻色彩的非线性关系，锐化时会造成色彩漂移，因此处理次数不宜过多，处理完成后将彩色锐化图像转换为灰度图像。

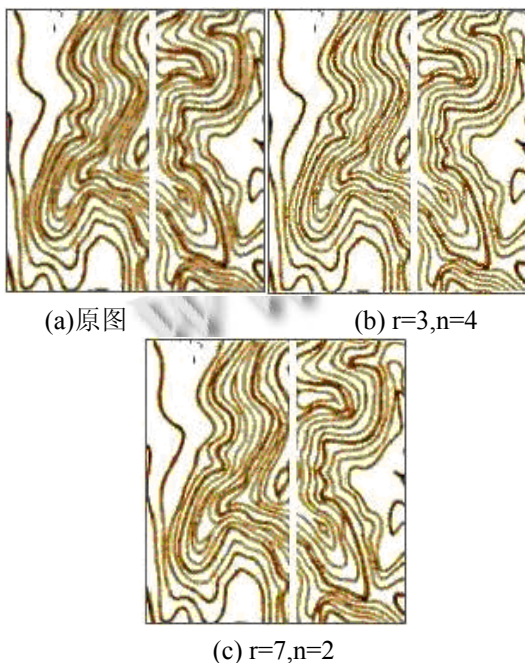


图 4 不同模板半径和处理次数结果对比， r 为模板半径， n 为处理次数

2.2 冗余剔除

灰度图像有 256 个色阶，对于 $m \times n$ 大小的模板有 $256^{m \times n}$ 种形态分布，因此必须选择大小合适的模板和判别方法才能进行剔除。模板越大对目标的空间形态判别越准确，但是判别算法实现会更复杂、计算时间越长。本文采用的方法及步骤如下：

(1)模板大小采用 1×7 像素。典型等高线扫描图像的直线在水平或垂直宽度为 3~4 灰度像素，斜线在水平或垂直方向的最小宽度为 2~5 灰度像素，粘连区域在水平或垂直方向的最小宽度为 3~10 灰度像素，为了减少判别条件，粘连区域单独处理。用 1×7 大小的模板对图像进行水平剔除，转置为 7×1 模板可对垂直方向进行剔除。

(2)模板色阶简化。 1×7 模板的灰度图像形态分布有 256^7 种，将模板中阈值大于 230 的灰度用 0 替代，小于 230 的用 1 替代，这样形态分布减少为 2^7 种，其中阈值可以根据图像的扫描质量进行调整。

(3)剔除原理。模板经过色阶简化后，任意两个 0 之间的距离表示白色之间的灰度等高线宽度。 1×7 模板的可以处理线宽为 1~5 个像素的灰度等高线，其中线宽为 1~2 像素的不处理，只处理 3~5 像素的线宽，这样判别条件简化为 6 个，判别模板 6 个，剔除方法 6 种，见表 1，其中 X 表示任意 0、1。

表 1 判别模板和细化方法

| 判别模板 | 剔除方法 | 判别模板 | 剔除方法 |
|---------|------|---------|------|
| 01110XX | 1 | 011110X | 4 |
| X01110X | 2 | X011110 | 5 |
| XX01110 | 3 | 0111110 | 6 |

(4)剔除方法。将灰度图像的四个边界均扩大 3 个像素。①对原始图像进行水平剔除。取出当前像素点的前 3 个和后 3 个像素，组成 1×7 模板，进行色阶简化，再与判别模板进行比较，选择处理方法进行剔除，具体为：当线宽为 3 像素时剔除颜色最浅的值，线宽为 4 像素时剔除较大的两个值，而线宽为 5 像素时，取出最大值和次大值，如果最大值在正中间时剔除最大值；如果最大值和次大值分别处于两侧端点或位于一侧端点且相邻，则剔除这两个值，见图 6，P 为可剔除像素点；其他情况不处理。

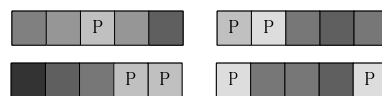


图 6 线宽为 5 像素时的剔除方法

②对原始图像进行垂直剔除. 扫描图像时取出当前像素点的上 3 个和下 3 个像素, 并将 7×1 模板转置为 1×7 模板, 其他过程与水平剔除相同. ③对水平剔除结果和垂直剔除结果进行“逻辑与”操作, 即保留水平剔除和垂直剔除均不为白色的像素值.

(5)光滑处理. 剔除完成后需要剔除毛刺和孤立点, 便于粘连判别和断点连接. 剔除模板见图 7, P 为可剔除点, 将 c、d、e 模板依次旋转 90° 可对上下左右四个方向的空间形态进行剔除.

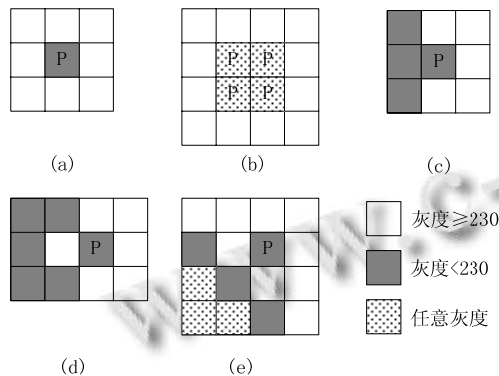


图 7 光滑处理的剔除模板

2.3 粘连分离

对图像中线宽大于 4 像素的区域进行粘连判别. 视以下三种情况为粘连: (1) 4×4 大小区域内像素值均小于 230; (2)5~9 像素大小的“工”型以及垂直方向“工”型的组合形态; (3)5~9 像素大小的“H”型以及以及水平方向“H”型的组合形态. 分离时将第一种粘连区域扩大并剔除, 视为断点, 同时记录剔除区域大小, 便于断点连接. 对于“工”型及其组合型、“H”型及其组合型, 采用 3×11 模板扫描水平方向 5~9 线宽的图像, 粘连部分水平方向的色阶分布如波浪形状, 可以直接剔除波峰处的最大值(最浅色). 由于模板可处理的最大线宽为 9 个像素, 波峰数量为 1~2 个, 实现较为简单. 垂直方向可用 11×3 模板处理“工”型及其组合型粘连, 方法类似. 如果需要处理更大线宽的粘连, 可将模板扩大为 3×13 、 3×15 或者更大, 分离方法相同, 但形态判别会更复杂.

2.4 等高线细化

等高线细化就是在保持等高线连通性的前提下, 逐步消除边缘点, 提取出等高线的中心线. 粘连分离后的等高线灰度图可以采用 USM 锐化再次处理, 由于灰度图像不存在色彩漂移, 高斯低通滤波器模板的大

小可以取 15、17 或更大, 处理次数 10 次以上, 可以有效消除浅色阶像素, 在保持连通性的同时减小线宽. 二值化时将阈值小于 230 的灰度值全部转换为黑色, 再采用经典的 Hilditch 细化算法进行细化.

2.5 断点连接

目前常见的断点连接方法有走向法、费曼编码法、形态膨胀法、拓扑关系法和最大集团图搜索法等^[2,6,8-10]. 本文提出了一种新的点配法进行连接, 方法是用矩形扫描窗口沿着断裂区域进行扫描, 将窗口内的断点分为两个分区, 将两个分区的断点按照一定的规则进行匹配, 然后优先连接匹配成功的断点, 其余断点以此为基础向两侧依次连接.

点配法需要考虑三个方面的因素:

(1)断裂区域的宽度. 断裂区域为前文所述的网格线、粘连剔除区和人工剔除区, 不含边界区域. 扫描窗口的大小与断裂区宽度相关, 例如连接因剔除网格线造成的断裂时, 如果网格线宽度为 n 像素, 矩形窗口的宽度选择为 $3n+2$, 长度可根据网格线的密度来确定, 一般可选为 $10n+1$. 对于断裂较大的区域(如地物标识剔除区、较大的粘连剔除区和人工剔除区)可以将断裂区边界外扩 5 个像素构成扫描窗口.

(2)断点分区与编号. 分区就是把扫描窗口内的断点分成两部分, 对于网格线造成的断裂, 断点分布在两侧, 以网格线中心划分即可. 对于宽大断裂, 断点分布在扫描窗口四周, 通常情况下很难进行分区, 但是如果可将断点分成两个均等的区域, 例如, 根据等高线生成出地性结构线^[11]的延长线将断裂区的断点分成两个均等部分, 点配法仍然适用. 扫描窗口内像素点的 8 邻域联通连接数为 1 表示该像素点是断点, 以断裂中心为 x 轴, 计算每个断点向 x 轴的垂直投影点和距离, 以垂直投影点的大小进行顺序编号, 如果某些断点的投影位置相同, 再以距离远近进行编号.

(3)匹配规则. 计算每个断点的走向, 优先连接走向延长线相交并且编号匹配的两个端点, 如图 8 所示, 虚线箭头为断点走向, 应优先连接 a_3 、 b_3 和 a_6 、 b_6 端点, 以此为基础向两侧依次连接, 此后其他断点的连接不再考虑走向和距离. 如果端点匹配但是走向均不相交, 并且每对匹配端点的水平或垂直位置相差均小于断距或者匹配断点位于地性结构线两侧时, 也可以成对连接. 如果编号不匹配, 表明等高线特征丢失严重, 不能使用该方法连接.

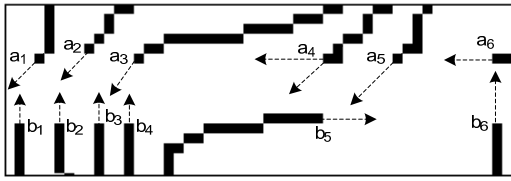


图 8 点配法的匹配规则示意图

断距较小的断点连接采用直线连接, 对于地性结构线两侧的匹配断点, 直线连接会造成交错或者粘连, 改进方法是在构造线上确定一点 c_i , 将匹配断点 a_i, c_i, b_i 用二次贝塞尔曲线连接, 见图 9. c_i 点的位置 (x_i, y_i) 由下式确定:

$$x_i = x_0 \pm [d + \frac{2w - nd \cos \theta}{(n - 2) \cos \theta} (i - 1)] \cdot |\sin(\arctan(k))|$$

$$y_i = y_0 \pm [d + \frac{2w - nd \cos \theta}{(n - 2) \cos \theta} (i - 1)] \cdot |\cos(\arctan(k))|$$

其中 $x_0 > x_i$ 或者 $y_0 > y_i$ 时, “±”取“-”, w 为扫描窗口宽度, n 为断点数目, k 为地性结构线斜率, θ 为地性结构线与扫描窗的夹角, (x_0, y_0) 为 c_0 点的坐标, d 取值 2 或 3.

图 1 所示的彩色等高线地形图经过预处理后可以得到较高质量的二值化等高线(结果见图 10), 可以大大减轻矢量化的工作量.

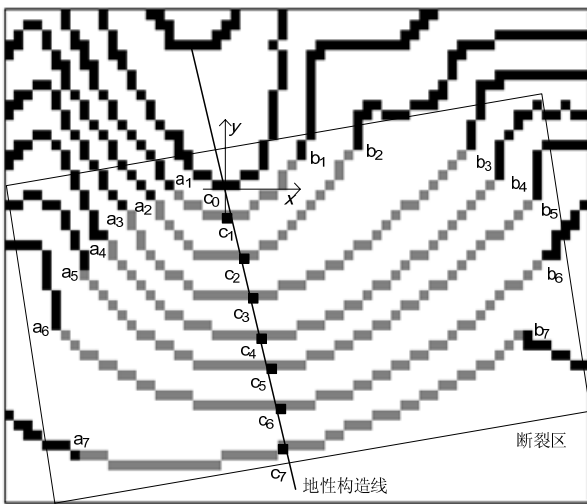


图 9 断点的贝塞尔曲线连接

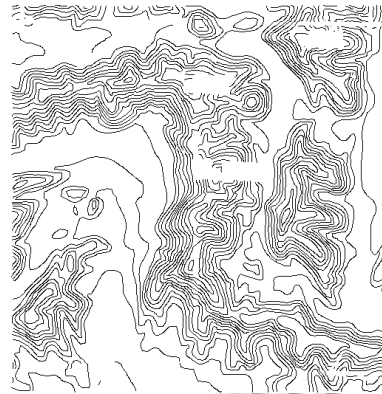


图 10 彩色等高线预处理结果

3 结语

目前常见的等高线处理方法都是采用提取法, 实际应用表明受到扫描图像质量的影响, 提取效果均不是很理想. 本文结合形态学分别在 RGB 和 HSV 空间剔除无用信息, 对剩余图像进行 USM 锐化、灰度剔除和光滑处理, 可以有效处理模糊及粘连现象, 得到高质量的等高线提取结果. 毛刺对点配法的判断影响较大, 光滑处理时尽可能剔除毛刺, 为了提高处理速度, 毛刺剔除模板的形态判别较少, 有可能会剔除断裂处的端点, 加大断裂宽度, 但这不影响对点配法的连接. 整个预处理过程算法简单, 易于实现.

参考文献

- 1 张耿,何翠云.利用 MAPGIS 图像处理实现彩色地图等高线的提取.南方国土资源,2005,4:33-34.
- 2 钱静.彩色地形图中等高线提取方法的研究.中国图形图像学报,2006,11(10):1443-1449.
- 3 侯萍.顾及拓扑关系的彩色扫描等高线的矢量化.武汉:武汉大学,2001.
- 4 曾迎生,冯金光,贺汉根.等高线识别中的一种分类算法.计算机工程与应用,2004,40(3):56-57.
- 5 姚伟新.等高线预处理技术的研究.贵阳:贵州大学,2007.
- 6 夏春林,王雪,叶长青.等高线自动识别方法的研究.矿山测量,2007,(1):32-34.