

# 基于空间统计学的空间数据窗口大小的确定

马洪超<sup>1</sup> 李德仁<sup>1</sup>

(1 武汉大学测绘遥感信息工程国家重点实验室, 武汉市珞喻路 129 号, 430079)

**摘要:** 提出基于空间统计学的方法来确定空间数据窗口大小, 实例证明是可行的。同时提出了对海底照片成像不均匀光照进行纠正的思路和方法, 该方法简单有效, 效果较理想。

**关键词:** 图像处理; 空间数据窗口; 空间统计学; 不均匀光照的纠正

中图法分类号: TP751; P237.9

在图像处理和其他空间数据处理中, 空间数据窗口的概念至关重要, 尤其在图像处理中更是如此<sup>[1,2]</sup>。这些窗口被定义为由  $m$  行和  $n$  列像素组成的一个二维数组, 行数和列数可以是不相等的, 但通常为方便计算, 取  $m = n$ , 这时窗口就是一个正方形。同时对行列数的取值也有一定的限制, 比如通常取奇数, 且小于或等于 15 个像素。经常使用的空间滤波算子(如 Sobel、Kirch 算子)中, 这个正方形窗口的大小以及每个像素位置上的权值已经确定, 且经检验有效。但正如 Castleman 指出的那样, 这样的窗口大小并不总是可以事先确定的, 有时图像分析人员在理解上有很大的随意性。本文给出的两个应用例子就说明了这一点。笔者以空间统计学作为工具, 探讨自适应地确定空间数据窗口的大小, 并给出这两个应用例子的解决方案。

## 1 问题的提出

### 1.1 Landsat5 TM 第 6 波段分辨率的提高

Landsat5 TM 数据共有 7 个波段: 3 个在可见光范围(第 1、2、3 波段), 3 个在红外或近红外范围(第 4、5、7 波段), 1 个在热红外(第 6 波段)<sup>[3]</sup>。除第 6 波段的地面分辨率是 120m 外, 其他的都是 30m。由于第 6 波段的分辨率远低于其他波段, 且和其他波段光谱特性存在差异, 一般的遥感图像处理中(如分类)通常不考虑这个波段。尽管如此, 它在地热资源的勘察、岩石波谱特征等

地学研究中有其独到的价值。实际上, 这个波段是应美国地质调查局的要求而增加的。

为了能把第 6 波段的分辨率提高到和其他波段一样的水平, 一些学者提出了不同算法, 有的从遥感物理的角度出发<sup>[4]</sup>, 也有的从统计的角度入手<sup>[5]</sup>。文献[5]提出的算法本质上基于多元回归分析, 这个算法的第一步是要将图像分割成一些小的图像块, 然后按照“局部点法”或“局部块法”进行计算。本文探讨的第一个问题是如何确定这些图像块的大小。

### 1.2 海底近景摄影照片不均匀光照的纠正

在大洋洋底表面, 分布着一种以锰为主的多金属结核, 有很大的开采和利用价值。为了在圈定的公海范围内估计这种金属结核的储量, 首先需要拍摄海底照片, 将这些照片二值化, 便可估算结核的面积, 供储量计算用。但由于海底一片漆黑, 要拍摄海底照片, 必须使用人工光源。这是一种非平行光源, 所得到的是非均匀光照下成像的照片<sup>[6]</sup>。这种照片的典型特点是, 在中心或接近中心部分亮度大, 而周围部分亮度小。如果直接将这种照片二值化, 得到的结果很差<sup>[6]</sup>。因此在进一步处理之前, 必须对非均匀光照下得到的照片进行纠正, 使得它和平行光下获得的照片基本一样。对这个问题, 有很多方法, 但多数是针对显微照片的非均匀光照纠正的<sup>[1,7]</sup>, 前提条件是能够获得载玻片上没有样品时的非均匀光照照片(即背景图), 这是不可能的。也有人提出用光度学中严格的数学模型来纠正, 这实际上也是不可

能的。且不说在4 000~5 000m 深的大洋洋底有多少因素影响成像, 即使是理想的情况, 确定模型的参数也是困难的。因此, 笔者提出如下模型。

在微积分中, 经常将曲线化直或曲面化平, 即在一个小范围内, 可以将曲线看成直线, 曲面看成平面。据此, 可以将上述照片分成若干小块, 在这些小块中, 认为是均匀光照成像的。然后, 在这些小块中按照以下的公式进行拉伸:

$$DN = \frac{255}{\max - \min} \times DN_{\text{origin}} \quad (1)$$

式中,  $\max$  和  $\min$  分别代表图像小块中的最大和最小亮度值。

如何确定小块图像的大小是本文探讨的第二个问题。

显然, 不能随意确定图像块的大小(在第一个问题中, 图像块的大小至少应该是 $4 \times 4$ 的倍数<sup>[4]</sup>), 否则在第一个问题中, 结果图像从图像熵的角度不是最优的, 而第二个问题中则会产生方块形的条带。下面以空间统计学作为工具, 自适应地确定这些窗口的大小, 即它们是根据图像数据本身的特点而不是人为确定的, 再用于解决上述问题。

## 2 空间统计学简介

空间统计学是一门起源于矿业领域的应用数学。这种方法的目的是克服经典统计学应用于空间数据处理时存在的一些问题。南非学者 Kriging 最初提出其萌芽思想, 后经法国数学家 G. Matheron(同时又是数学形态学的创始人)从随机过程的理论出发进行彻底的改造和完善, 并于 70 年代末、80 年代初引进到我国<sup>[8~10]</sup>。

在西方有关地理信息处理的文献中, 这种方法经常被用于 GIS 和遥感数据的处理, 并将其作为遥感信息处理的一个数学模型<sup>[11]</sup>。总结起来, 在以下几个领域经常使用: ①GIS 空间分析<sup>[12]</sup>; ②GIS 不确定性研究<sup>[13]</sup>; ③DEM 数据质量评价<sup>[14~17]</sup>; ④遥感影像处理<sup>[18~23]</sup>。可见, 这种方法的应用几乎涵盖了所有地理信息处理领域。空间统计学应用于遥感信息处理开始于 80 年代后期, 由 Curren 和 Woodcock 等人开创先河<sup>[19, 23]</sup>, 此后一直有大量的文献出版, 并将其作为一种常规的处理手段。然而, 在我国将这种方法应用于地理信息处理的研究仅有个别的报道<sup>[23]</sup>。

空间统计学的主要内容包括: 变差函数的性状研究以及理论模型; 各种 Kriging 插值法的理

论和应用研究; 条件模拟; 多变量空间统计学; 非线性非平稳空间统计学等。后面几种属于高等空间统计学的范畴, 涉及到的数学基础比较广泛。

空间统计学和其他描述地球空间数据特性的方法, 如分形理论描述的地球表面的粗糙度、基于小波的空间现象多尺度性等有着密切的联系, 它们是空间统计学研究的内容之一。

## 3 变差函数及其在模板窗口确定中的应用

### 3.1 变差函数的定义<sup>[8~10]</sup>

变差函数是一切应用空间统计学的基础。在空间统计学中, 以三维空间坐标  $(x, y, z)$  为自变量的随机场称为一个区域化变量, 它是空间统计学的研究对象。地形标高、遥感数据、大气污染量、海底深度等都是区域化变量的例子。由于区域化变量的特殊性, 传统的概率统计方法应用面太窄, 于是在空间统计学中引入了变差函数的概念。

设三维空间中位于  $x$  点的区域化变量为  $Z(x)$ , 在  $x + h$  点处的区域化变量是  $Z(x + h)$ , 则变差函数定义为<sup>[10]</sup>:

$$\gamma(x + h) = \frac{1}{2} E[Z(x) - Z(x + h)]^2 - \frac{1}{2} \{E[Z(x)] - E[Z(x + h)]\}^2 \quad (2)$$

式中,  $E[Z(x)]$  表示  $[Z(x)]$  的数学期望。实质上, 上式表示区域化变量  $Z(x)$  和  $Z(x + h)$  差的方差。在遥感影像中, 还可以定义如下。设影像的灰度函数是  $G(x, y)$ , 则灰度的变差函数可以表示成<sup>[22]</sup>:

$$2\gamma(h) = \int_{x'} \int_{y'} [G(x, y) - G(x', y')]^2 dy dx \quad (3)$$

式中,  $h$  是以像素为单位的、像元点  $(x, y)$  和  $(x', y')$  之间的欧氏距离。上面两个定义式本质上是一样的。在实际计算中, 要对上述两个公式离散化, 利用公式<sup>[22]</sup>:

$$\gamma(h) = \frac{1}{2N} \sum_{i=1}^N [G(x, y) - G(x', y')]^2 \quad (4)$$

式中,  $N$  是相隔为  $h$  的像素对的个数。

计算时还要有一个假设<sup>[8~10]</sup>, 即(二阶)平稳性假设: ①区域化变量的数学期望存在且为常数; ②区域化变量的协方差函数存在且平稳(只依赖于  $h$  而与位置无关)。严格的平稳性假设在地理信息处理中是不存在的, 即使是二阶平稳性假设

也未必存在,因此又有比它更弱的本征假设。

由于区域化变量就是以三维空间坐标为变量的随机场,因此,区域化变量  $Z(x)$  的自协方差函数定义为,在两个空间点  $x$  和  $x+h$  处的两个随机变量  $Z(x)$  和  $Z(x+h)$  的二阶中心矩,即

$$\begin{aligned}\text{cov}[Z(x), Z(x+h)] &= C(x, x+h) = \\ C(h) &= E[Z(x) \times Z(x+h)] - \\ E[Z(x)] \times E[Z(x+h)]\end{aligned}$$

上式简称为  $Z(x)$  的协方差函数。当  $h=0$  时,  $C(0)=E[Z(x)]^2-\{E[Z(x)]\}^2$ , 即为  $Z(x)$  的(验前)方差函数。

变差函数和协方差函数之间的关系为:

$$\gamma(h) = C(0) - C(h)$$

式中,  $C(0)$  是验前方差。上式表明,从协方差函数可以推出变差函数,从而表明在刻画区域化变量的统计特征时,变差函数比协方差函数具有更一般的意义。实际上,在有的实际问题中,协方差函数不存在,但变差函数可以存在且是平稳的。

### 3.2 变差函数确定空间数据窗口大小的原理

变差函数的图形如图 1 所示。

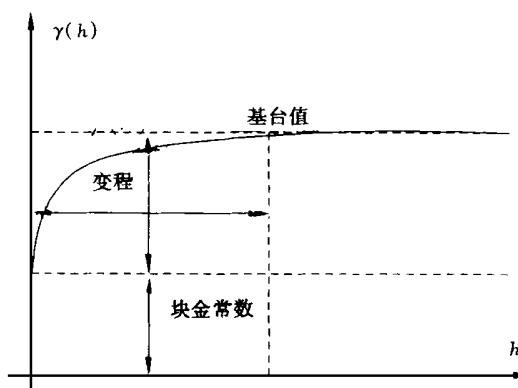


图 1 变差函数图

Fig. 1 Diagram of Semivariogram

通常,变差函数在原点处为零,随  $h$  的增大而增大,反映了随着两个数据点之间的距离增大其相关性变小,当  $h$  超过某一距离以后( $a>0$ ),  $\gamma(h)$  趋向平稳,这时  $a$  称为变程,  $\gamma(\infty)$  称为基台值。但如果  $\gamma(h)$  在原点处间断,这时虽然  $\gamma(0)$  仍然等于零,但  $\lim_{h \rightarrow 0} \gamma(h) = C_0 > 0$ , 这里  $C_0$  称为块金常数。

对二维数据来说,在不同的方向可计算不同的变差函数,即所谓的方向变差函数(directional semi-variogram)。Curren 等<sup>[19]</sup>认为遥感影像满足计算变差函数时所需要的平稳性假设(或本征假设),并首先给出变差函数的物理解释:  $\gamma(h)$  反映的是地面辐射与距离  $h$  的变化性,如果由影像

像元亮度反映的地面辐射与距离  $h$  有空间相关性,则  $\gamma(h)$  的特点如图 2 所示;块金常数是像元内部变化性的度量,同时又反映图像的信噪比;变程就是  $\gamma(h)$  达到稳定时所对应的距离  $h$ ,反映的是区域化变量的空间相关性或均匀性的范围。

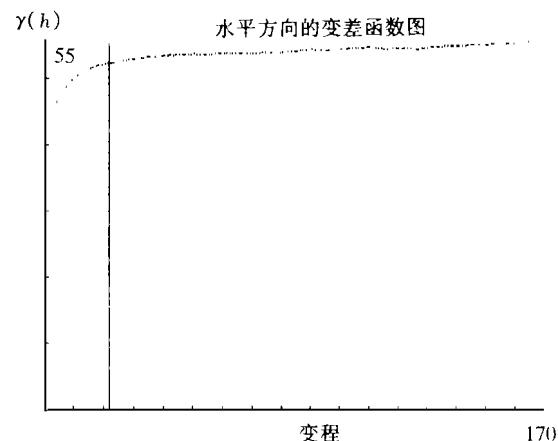


图 2 云南维西某地区 TM3 水平方向上的变差函数图

Fig. 2 Semivariogram of TM3 of Weixi Area, Yunnan Prov.

Franklin 等<sup>[21]</sup>的研究表明,变程是确定空间数据窗口(geographical window)大小很好的工具,即根据地表的实际情况确定遥感图像处理中空间数据窗口的大小,并给出了应用实例。在这里,假定平稳性假设(或本征假设)在海底照片中同样得到满足,并由  $x, y$  两个方向的变程确定第二个问题中矩形图像块的大小。

根据上述原理,笔者给出确定空间数据窗口大小的算法。

1) 计算水平方向上的变差函数,确定水平方向上的变程。对图像来说,以像素为单位,设为  $m$ 。

2) 计算垂直方向上的变差函数,确定垂直方向上的变程。对图像来说,同样以像素为单位,设为  $n$ 。

3) 以  $m \times n$  为大小的窗口即为所求的空间数据窗口大小。

4) 在用式(4)计算变差函数时,  $N$  即为相隔  $h$ (以像素为单位)的像素对的个数,是  $h$  的函数。 $h$  是根据截断距离(truncate section)的大小确定的<sup>[20]</sup>。

## 4 实际应用

以 §1 两个实际问题为例,这里给出一个变差函数的具体变差图(图 2)。在第一个问题中,以云南维西地区的 TM3 影像(512 像素  $\times$  512 像

素)为例,求得平均窗口的大小是 12 像素×12 像素,然后对整幅影像进行分块(分块时,分别以 12 除总像元数再取整),并在小图像块中计算参数,进行回归计算,得到新的 TM6 影像,将新影像以目视和图像熵作为指标和原始影像进行定性和定量对比分析,可知结果图像在目视效果上得到了一定的改善(表现为纹理更加丰富),在图像熵上比原来的影像有很大的提高,说明信息量有很大的提高。具体计算步骤参见文献[26]。

对第二个问题进行同样的处理。在这个问题中,作如下的简化:把二维影像数据看作是一维排列的,即当前行的数据是接着上一行数据排列的,相当于计算机中以行优先的数组组织方式。这时将二维数据简化为一维处理,可以减少运算量。对典型的海底照片<sup>[6]</sup>的处理中,计算得到的变程是 8 个像素,再按式(1)对海底照片进行处理,最后得到的结果图像如图 3 所示,二值化以后的图像如图 4 所示。

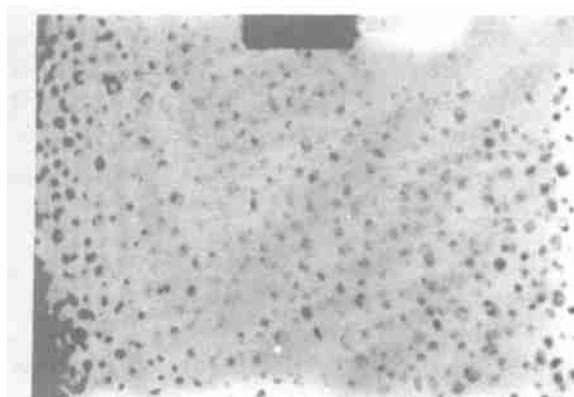


图 3 处理结果

Fig. 3 Result Image

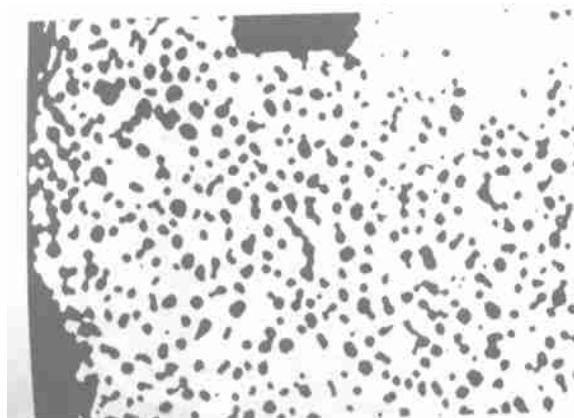


图 4 图 3 的二值化结果

Fig. 4 Binarized Result of Fig. 3

由图 3 和图 4 可知,该方法取得了很好的效果,而且简单易行。相反,利用 Wallis 滤波、 $\hat{\alpha}$  trous 小波变换及根据成像模型的严格纠正方法等取得的结果都不很理想(图 5 是 Wallis 滤波后的结果)。

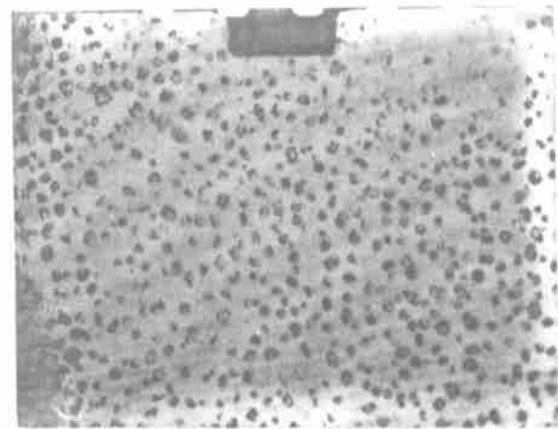


图 5 用 Wallis 滤波的结果

Fig. 5 Result Image Using Wallis Filtering

致谢:文中引用的海底照片由国家海洋局杭州海洋二所提供的,特此感谢。

## 参 考 文 献

- 1 Carstleman K R. Digital Image Processing. N J: Prentice Hall International Inc., 1996
- 2 许殿元, 丁树柏. 遥感图像信息处理. 北京: 宇航出版社, 1990
- 3 Richard J R. Remote Sensing Digital Image Processing: An Introduction. 3rd ed. Berlin: Springer-Verlag 1999
- 4 Liu J G, Moore J McM. Pixel Block Intensity Modulation: Adding Spatial Detail to TM Band6 Thermal Imagery. Int. J. Remote Sensing. 1998, 19 (13): 2 477 ~ 2 491
- 5 罗忠, 朱重光. 基于多元统计的 TM6 影像融合. 航空计算, 1998, 28(3): 40 ~ 42
- 6 金淑英, 包更生, 马洪超, 等. 海底视像图像的灰度不均匀校正方法的研究. 武汉测绘科技大学学报, 2000, 25(6): 531 ~ 536
- 7 Russ J C. The Image Processing Handbook. Raton: CRC Press, 1992
- 8 Isaaks E, Srivastava R. An Introduction to Applied Geostatistics. London: Oxford University Press, 1989
- 9 Journel A G, Huijbregts C J. Mining Geostatistics. London: Academic Press, 1978
- 10 王仁铎, 胡光道. 线性地质统计学. 北京: 地质出版社, 1988
- 11 Schowengerdt R A. Remote Sensing Models and Methods for Image Processing. 2nd ed. San Diego: Academic

- Press 1997
- 12 Cressie N A C. Statistics for Spatial Data. New York: John Wiley & Sons 1991
- 13 史文中. 空间数据误差处理的理论与方法. 北京: 科学出版社, 1998
- 14 Kyriakidis P C, Shortridge A M, Goodchild M F. Geostatistics for Conflation and Accuracy Assessment of Digital Elevation Models. *Int. J. Geographical Information Science*, 1999, 13(7): 677~707
- 15 Arbia G, Griffith D, Haining R. Error Propagation Modeling in Raster GIS: Adding and Ratioing Operations. *Cartography and Geographic Information Science*, 1999, 13(7): 297~315
- 16 Liu H X, Jezeck K C. Investigating DEM Error Pattern by Directional Variogram and Fourier Analysis. *Geographical Analysis*, 1999, 31(3): 249~266
- 17 Chiles J P, Pierre D. Geostatistics: Modeling Spatial Uncertainty. New York: A Wiley Interscience Publication, 1999
- 18 李德仁, 关泽群. 空间信息系统的集成与实现. 武汉: 武汉测绘科技大学出版社, 2000
- 19 Curran P J. The Semivariogram in Remote Sensing: An Introduction. *Remote Sensing of Environment*, 1988 (24): 493~507
- 20 Carr J R. Spectral and Texture Classification of Single and Multiple Band Digital Images. *Computer & Geosciences*, 1996, 22(8): 849~865
- 21 Franklin S E, Wulder M A, Lavigne M B. Automated Derivation of Geographic Window Sizes for Use in Remote Sensing Digital Image Texture Analysis. *Computer & Geosciences*, 1996, 22(6): 665~673
- 22 Miranda F P, Carr J R. Application of the Semivariogram Texture Classifier for Vegetation Discrimination Using SIR-B Data of Borneo. *Int. J. Remote Sensing*, 1992, 13: 2349~2354
- 23 Woodcock C E, Strahler A H, Jupp D L B. The Use of Variograms in Remote Sensing I: Scene Models and Simulated Images, and II: Real Digital Images. *Remote Sensing of Environment*, 1988(25): 323~379
- 24 张景雄, Kirby R P. 模糊类别制图的空间统计学方法. *测绘学报*, 1998, 27(2): 166~176
- 25 Van der Meer F, Wim B, Klaas S, et al. Scaling to the MERIS Resolution: Mapping Accuracy and Spatial Variability. *Geocarto International*, 2000, 15(1): 37~47
- 26 Ma H C. Methodology Study on Spatial Data Fusion and Its Implementation — Taking Geo-data Integration as Example. [Dissertation]. Wuhan: China University of Geosciences, 1999

**作者简介:** 马洪超, 副教授, 在站博士后。主要从事遥感信息处理、空间统计学理论及其在地理信息处理中的应用等理论研究和软件开发工作。代表成果: 资源环境评价决策专家系统; 矿产资源预测评价决策分析系统; 一种自动绘制聚类分析谱系图的新算法等。

E-mail: hema@hp01.wtusm.edu.cn

## Geographic Window Fixing and Its Application Based on Spatial Statistics

MA Hongchao<sup>1</sup> LI Deren<sup>1</sup>

(1 National Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan, China 430079)

**Abstract:** Fixing the size and shape of geographic window is vitally important in geo-spatial data processing, especially in the field of remotely sensed data processing. Conventionally, geographic window (both the size and the shape) will be fixed before the image processing task is carried out and this fixed window will be moved within the whole image while operators such as Sobel filter are being calculated within it. Such is the case commonly encountered in spatial filtering. Though commonly used neighborhood processing operators have their fixed shapes and sizes, and weights in their corresponding position in the window, there is no existing methodology for fixing the geographic window self-adaptedly, that is, determining the shape and size of a window according to the image data themselves, other than arbitrarily chosen by the analysts.

This paper presents the strategy for implementing the objectives mentioned above in the context of two practical examples, by employing theories and approaches from spatial statistics. The

first case is to enhance the ground resolution of TM6 by employing regression model, which requires some statistical parameters before the regression analysis and all these parameters should be calculated within small image blocks. The second case is to correct the non-uniform illumination effect appearing in pictures of deep sea-floor. The non-uniform illumination effect could be easily removed according to the algorithm proposed in this paper, however, the whole image is also needed to be divided into small image blocks before the algorithm can be used. The sizes of all the small image blocks in both cases can not be determined arbitrarily, otherwise, the resultant will not be optimal in the first case or the non-uniform illumination effect will not be removed completely in the second one.

A concise introduction of spatial statistics is presented in the paper in order to help those who are not familiar with this subject. The range, an important parameter from variogram, reflects the area within which the autocorrelation of a regionarized variable between two separated spatial points, say  $x$  and  $x+h$ , is significant or not. This property is actually the embodiment of the homogeneity in the given area. Once this area is sensed by satellite sensors or by other means and digitized to be digital images, this homogeneity will be inherited. So it is obvious that the size of a geographic window within which geo-spatial data analysis will be carried out should be determined according to the homogeneity of the corresponding area. With this conclusion in mind, the horizontal and vertical variogram can be calculated and the corresponding ranges can be obtained. Assign the values of the horizontal and vertical range to be the width and height of a geographic window respectively, the window is then determined. The main steps implementing the suggested strategy are listed and some attentions that should be paid when using this method are also expounded in the paper. Finally, the two problems presented at the beginning of the paper are successfully solved using the method described here.

Besides, the method proposed in this paper of non-uniform illumination correction for deep sea floor pictures is a simple but successful one. Borrowing the idea of linearization and planarization from calculus, a stretching formula proposed in this paper removes the non-uniform illumination effect completely. This method has been widely adopted in practice.

**Key words:** image processing; geographic window; spatial statistics; non-uniform illumination effective removing

---

**About the author:** MA Hongchao, associate professor, post doc. fellow. His research interests include remotely sensed image processing and its application, spatial statistics and its applications to geo-information processing. His typical achievements are assessment and decision expert system in natural resources and environment; prediction, evaluation and decision analysis system for mineral resources; a new algorithm for automatically drawing dendrogram of clustering analysis etc.

E-mail: hama@hp01.wtusm.edu.cn