

分类线性回归的 Landsat 影像去云方法

吴 炜^{1,2} 骆剑承² 沈占锋² 王卫红¹

(1 浙江工业大学计算机科学与技术学院, 杭州市留和路 288 号, 310023)

(2 中国科学院遥感与数字地球研究所, 北京市大屯路甲 3 号, 100101)

摘 要:首先,对参考影像和待去云影像上的云覆盖区域进行检测和掩膜;然后,对掩膜后的参考影像进行 ISODATA 聚类,并建立各个类别参考影像到待去云影像灰度值的线性回归方程;再对待去云影像上的云覆盖区域,依据参考影像上的灰度值进行最小距离方法分类,划分到聚类形成的各个类别之中;最后,依据各个类别回归方程进行灰度值预测。实验结果表明,所提方法能够进行云区的检测和去除,预测精度相比传统方法有较大提高。

关键词:影像镶嵌;去云;云检测;分类线性回归

中图分类号:P237.4; TP753

由于天气原因,获取较大区域无云覆盖遥感影像的概率较小。云或云阴影对地球表面的遮挡造成遥感影像上地物信息的损失,限制了大区域遥感调查与监测的应用水平与能力。文献[1]对 2000~2002 时段内所有 Landsat ETM+ 的历史存档数据进行研究,发现获得同一季度或者更短时间内全球或部分区域的无云影像特别困难。因此,云检测与去除是遥感影像处理过程中的一个关键问题。

云和阴影去除包括云检测和去除两个过程。常用的云检测方法主要采用云和阴影的光谱特征进行阈值分割,如光谱反射率阈值法^[2]、云顶温度阈值法^[3-4]等,而由于遥感成像条件的差异,阈值的选择存在较多的不确定性,较难区分云层与高反射率人工地物、阴影与低反射率水体之间的光谱差异^[5]。针对 Landsat ETM+ 数据提出的 ACCA 云监测算法较充分地利用了影像各个波段的光谱特征,采用了多组阈值分割算子估算云量百分比,但此算法复杂度高、计算量大,且不能获取云覆盖的区域^[6]。云去除的方法包括滤波修复法和镶嵌去云法。前者主要利用小波分析法、同态滤波法等去除影像中代表云层的低频成分^[7-9];后者利用多时相遥感影像进行镶嵌,获得

无云影像^[10-14],并采用辐射归一化等线性回归模型方法改正不同时相影像的辐射差异^[11,15],目前,Landsat、MODIS 等大区域遥感影像去云大多都采用这种方法^[10]。然而,受大气、光照及物候等因素的影响,不同时期遥感影像的光谱差异并不是线性的,尤其是容易受季相变化影响的植被等地物类型。

本文针对遥感影像镶嵌去云中多时相辐射特征的非线性变化,提出了一种利用多期同区域的影像进行云检测和去除的方法,其主要思路是通过为每种地类建立线性回归方程,将全局影像辐射一致性处理中的非线性问题转换为线性问题,从而提高预测精度。

1 方法原理

本文采用的云影检测及去除方法的流程为:

- ① 对参考影像和待去云影像差值进行直方图分割,将影像的所在区域划分为云区 A' (待去云影像上有云)、 B' (参考影像上有云) 和无云区 C' ;
- ② 对参考影像的 C' 区域进行 ISODATA 聚类,建立各个类别从参考影像到待去云影像灰度值的回归方程;
- ③ 将参考影像的 A' 区域采用最小距离方

收稿日期:2013-05-13。

项目来源:国家自然科学基金资助项目(41101398);浙江省自然科学基金资助项目(R1090569,LY12F02039);国家 863 计划资助项目(2013AA12A401);国家科技支撑计划资助项目(2011BAH06B02,2012BAH33B01)。

法划分到聚类形成的类别中,并根据对应类别的回归方程预测待去云影像上对应位置的像素值。

1.1 差值影像直方图云检测方法

本文的云检测方法基于两个假设:①在地表覆被类型未发生变化的情况下,引起两景时相相近的影像变化的因素主要是云或云阴影;②云覆盖区的影像亮度值较高。根据以上假设,选取两景影像上亮度值变化“较大”且亮度值“较大”的区域,采用阈值法检测出影像上云覆盖的范围。

参考影像与待处理影像差值影像的直方图在理想状态下呈三峰分布。左边的峰对应参考影像的云阴影区域以及待去云影像的云覆盖区域;中间的峰对应未变化的区域;右边的峰对应待去云影像的云阴影区域以及参考影像上的云覆盖区域。取3个波峰间的两个谷底作为差值影像的阈值分割点,将影像划分为云覆盖区和无云区。对于云量分布较小的影像而言,云污染区域占整个影像面积的比例较小,有云覆盖区的波峰可能淹没在未变化区域波峰峰旁的缓坡里面(见图1)。由图1可知,直方图上远离波峰处的坡度变化缓慢,包含了较多亮度值变化较大的云层及其云阴影区域的信息。

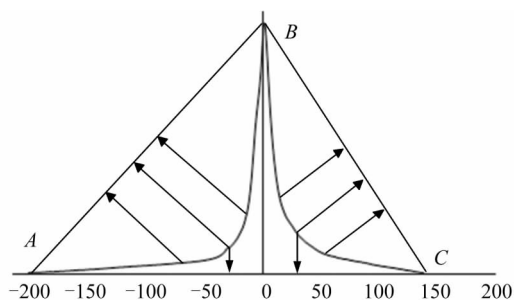


图1 直方图分割方法

Fig. 1 Method of Histogram Segmentation

为了解决这个问题,本文根据直方图凹凸性分析^[16]提出了一种针对云检测的改进阈值确定方法(见图1)。直方图凹凸性分析将图像的直方图看成一个平面区域,计算该区域的凸包,并分别统计各个亮度值对应的凸残差。由于凸残差的最大值对应于直方图高峰的肩处,可以使用对应于最大凸残差的灰度值来进行图像上多个类别的分割。具体算法如下:①连接灰度值最小点、波峰点及灰度值最大点组成一个三角形,即图1中A、B、C三点;②计算直方图上AB弧段各整数点到直线AB的距离,以及弧段BC到直线BC的距离;③分别取弧段AB和弧段BC上到两直线距离最大处对应的差值 T_1 和 T_2 ,作为阴影区和云覆盖区的分割阈值,并按照式(1)将影像划分为

A' 、 B' 和 C' 三个区域。

$$I_{(x,y)} = \begin{cases} A', f(x,y) < T_1 \\ C', T_1 \leq f(x,y) \leq T_2 \\ B', f(x,y) > T_2 \end{cases} \quad (1)$$

式中, $I_{(x,y)}$ 为待处理影像的云覆盖类型; $f(x,y)$ 为差值影像的灰度值; T_1 和 T_2 分别为阴影区和云覆盖区的分割阈值。

利用云在可见光、近红外和短波红外波段的反射率较高、云阴影反射率较低的特点,分别计算两景影像的波段均值 μ 及其方差 σ ,取小于 $\mu - \sigma$ 为云阴影以及大于 $\mu + \sigma$ 为云覆盖的可能区域;然后,取同时满足上述两个条件的区域作为云覆盖区。

1.2 分类线性回归云去除方法

在不考虑地物变化的情况下,不同时相影像中不同地类的亮度值变化并非呈一致的变化趋势,建立精确的回归模型较为困难。从同一地区两景影像之间的线性相关系数来看(见表1, B_i 表示波段*i*),不同波段的相关性差异较大,其中,波段1、2与各个波段的相关性较差,3、4(红/近红外)波段其次,5、7(短波红外)波段最强。因此,在地表类型复杂的情况下,采用单一的线性回归方程进行云去除难以取得较好的效果。

表1 两景不同时相的TM影像波段之间的相关关系

Tab. 1 Correlation of Two Temporal TM Images of the Same Area

	B_1	B_2	B_3	B_4	B_5	B_7
B_1	0.381	0.493	0.620	0.508	0.670	0.707
B_2	0.493	0.559	0.697	0.609	0.753	0.781
B_3	0.62	0.697	0.762	0.703	0.860	0.876
B_4	0.507	0.609	0.703	0.925	0.860	0.791
B_5	0.670	0.753	0.860	0.860	0.959	0.94
B_7	0.707	0.781	0.876	0.791	0.944	0.948

考虑到不同地物随时相变化在不同影像上表现出的不同变化规律,可以对不同类型的地物分别采用不同的线性变换方程,从而减少影像镶嵌后的辐射差异。对相同的图像按照地表类型分为6类(C_1, \dots, C_6),并按照地表覆盖类型分别计算波段1(B_1)和波段2(B_2)的各个类别之间的相关系数,如表2所示。从表2中可以看出,按类别统计的相关系数明显比不分类(N)的相关系数高。

表2 分类之后各个类别之间的相关系数

Tab. 2 Correlation of Different Class after Classification

	N	C_1	C_2	C_3	C_4	C_5	C_6
B_1	0.381	0.577	0.742	0.646	0.693	0.765	0.636
B_2	0.559	0.646	0.741	0.740	0.766	0.666	0.791

在上述两景影像上的第一波段上随机选择了 32 个点,并求出线性回归方程(见图 2),可以看出,分段线性回归方程明显比单一回归方程的相关系数高,从而回归预测的误差更小。

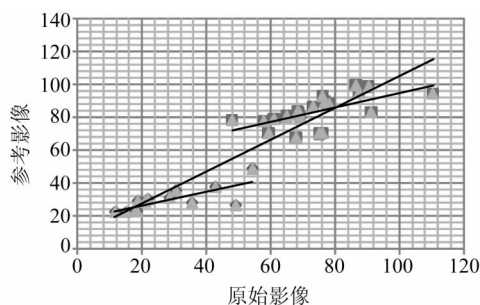


图 2 分类线性回归对比

Fig. 2 Comparison of Classified Linear Regression

分类线性回归去云方法可以总结为:

1) 按照影像差值的直方图阈值分割方法,提取出两幅图像中的云/阴影覆盖范围,并将影像划分为待去云影像有云区 A' 、参考影像有云区 B' 、无云区 C' 三个区域。

2) 对参考影像上的无云区 C' 进行 ISODATA 聚类,将影像划分为 6 个类别。聚类数目设置为 6,可以有效区分植被、水体、裸地等光谱不同变化模式的地物,同时还可以减少聚类和镶嵌填充的计算量。

3) 将参考影像作为自变量,待处理影像作为因变量,分别建立各个类别之间的回归方程。本文采用多元线性回归方程,如式(2)所示。

$$Y = AX \quad (2)$$

式中, Y 表示 $[y_1 \ y_2 \ \cdots \ y_7]^T$; X 表示 $[x_1 \ x_2 \ \cdots \ x_7]^T$; x_i 表示参考影像上第 i 个波段的像素值; y_i 表示第 i 个波段的回归值,即对待去云影像上的填充值; A 表示变换系数矩阵。

4) 将参考影像上待去云的区域(B' 区域),根据最小距离的分类方法划分到上述聚类形成的各个类别中。

5) 根据各个类别所对应的变换方程,预测待去云影像上被云/阴影遮挡区域的像素值,得到无云的镶嵌影像。

2 实验与分析

根据云遮挡区域地表覆被类型是否发生变化,本文选取了两组影像进行实验和分析。

2.1 实验 1

实验 1 选取 2007-05-30~2007-06-15 获取的中亚咸海地区 Landsat 5 的 TM 影像进行镶嵌去

云实验,标准假彩色影像(波段 4、3、2 合成)如图 3(a)、3(b)所示。图 3(a)中,云覆盖区的地表覆被类型有居民地、裸地、耕地、林地以及水面;图 3(b)中,云覆盖区的地表覆被类型有城市、裸地与少量植被。由于时相仅相差 16 d,季相变化较小,亮度值变化的主要因素是光照、系统畸变等。

按照本文提出的去云方法进行实验,云检测的效果如图 3(c)、3(d)所示,图 3(e)为去云之后的影像。图 3(b)、3(e)方框所示区域的放大叠加卷帘显示效果如图 3(f)、3(g)所示。可以看出,云/阴影覆盖区域经过去云处理后,细节清晰,“粘贴”区域与原始影像未出现明显的接缝线,这说明采用参考影像进行回归得到的预测值与待去云影像的影像本身值的差异较小,本文方法有效。

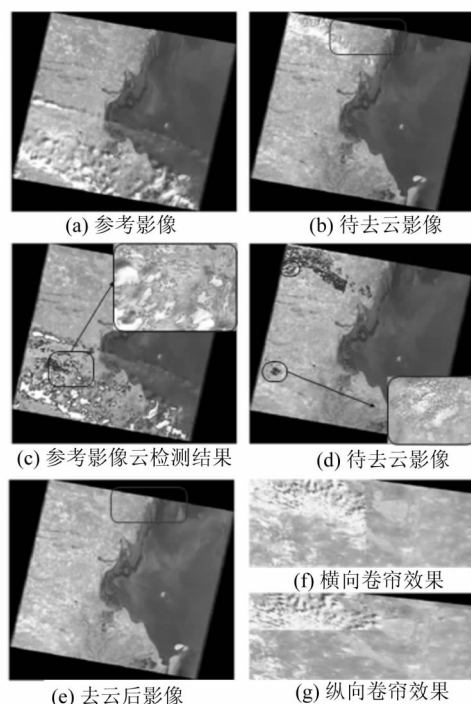


图 3 实验结果 1

Fig. 3 The First Experiment Results

2.2 实验 2

为了分析时相不一致带来的地表类型变化对去云方法的影响,实验 2 选取 2009-07-21 与 2010-07-08 获取的哈萨克斯坦西北部 Landsat 5 的 TM 影像进行实验,如图 4(a)、4(b)所示。图 4(a)左上和图 4(b)右上部分分别被云遮挡,其中被云遮挡的地区主要为耕地,农作物类型变化较大。采用本文的去云镶嵌方法得到的效果图如图 4(c)~4(f)所示。由图 4(f)放大卷帘的效果可知,去云后被云遮挡的信息得到了有效的恢复,方框内部区域在两景影像上的土地覆被类型发生了变化,用参考影像的回归预测值进行替换;而方框外部的部分

地物类型变化较小,回归预测值与原始影像的色彩差异较小,得到无缝镶嵌的结果。因而,本文方法也能适应于地表类型发生变化的区域。

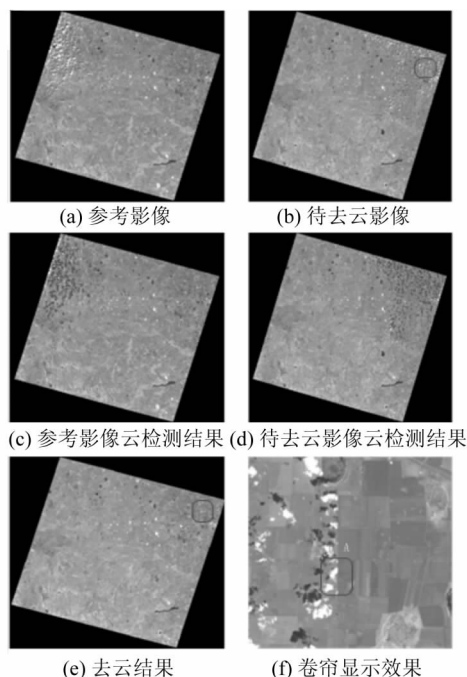


图4 实验结果2

Fig. 4 The Second Experiment Results

2.3 精度评价

云检测的精度采用正确检出率,即正确检测为云的像素数目与全部云像素数目的百分比来衡量。将人工目视解译并手动标记出的云范围作为标准,与本文方法得到的云范围进行比较,分别得到参考影像与待去云影像的云检测率。均方根误差(RMSE)是一种常用的预测精度度量方法:

$$R = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (3)$$

假设将影像分为 M 类,分别计算 M 个类别的加权残差,其方法如式(4)所示:

$$S_R = \sum_{j=1}^M W_j R_j \quad (4)$$

式中, W_j 表示各个类别像素占整个像素的比例,其和为 1。

将参考影像和待去云影像上均无云区域(C' 区域)各个类别 $1/3$ 的像素作为样本,另外 $2/3$ 像素用于测试,根据建立的回归方程,利用参考影像上的灰度值预测待去云影像上的灰度值,并计算回归预测得到的均方根误差。分别选取 5 组影像,聚类数目设为 6,统计结果如表 3 所示。其中参照方法为直接进行回归。从表 3 中可以看出:① 云的检出率较高,基本能够达到 85% 以上;② 分类所得出的残差均小于不分类所得出的残

差,使用分类线性回归的方法能够有效减小算法的残差,提高预测的精度。

表3 均方根误差统计表

Tab. 3 Statics of RMSE

编号	影像	检出率/%	本文方法 RMSE	参照方法 RMSE
1	I	85.9	4.984 086	8.680 42
	R	87.3		
2	I	89.0	5.566 784	10.805 083
	R	92.5		
3	I	83.4	5.953 788	7.457 551
	R	87.8		
4	I	88.3	3.950 764	6.996 949
	R	86.5		
5	I	92.1	2.548 45	7.844 638
	R	93.5		

3 结 语

本文提出了一种通过多景被云层覆盖的影像镶嵌一景无云影像的方法。此方法通过多特征自动确定多阈值,进行云覆盖区域检测;通过聚类形成不同地物类别,并分别建立回归方程;根据聚类形成的类别,对参考影像上的有云区域进行最小距离分类,并根据相应类别进行值预测。

进一步需要研究的问题有:① 聚类数目对于预测精度的影响;② 本文方法对 CBERS、HJ-1、SPOT 等多种影像的适应性。

参 考 文 献

- [1] Ju J, Roy D P. The Availability of Cloud-Free Landsat ETM+ Data over the Conterminous United States and Globally[J]. Remote Sensing of Environment, 2008, 112(3): 1 196-1 211
- [2] Song M, Civco D. A Knowledge-based Approach for Reducing Cloud and Shadow[C]. ASPRS-ACSM Annual Conference and FIG XXII Congress, New York, 2002
- [3] 王伟,宋卫国,刘士兴,等. K-Means 聚类与多光谱阈值相结合的 MODIS 云检测算法[J]. 光谱学与光谱分析, 2011, 31(4): 1 061-1 064
- [4] Gabarda S. Cloud Covering Denoising Through Image Fusion[J]. Image and Vision Computing, 2007, 25(5): 523-530
- [5] 吴传庆,王桥,杨志峰. 基于混合像元分解的水体遥感图像去云法[J]. 遥感学报, 2006, (10)2: 176-183
- [6] Irish RR, Barker J L, Goward S N, et al. Characterization of the Landsat-7 ETM + Automated Cloud-Cover Assessment (ACCA) Algorithm[J].

- Photogrammetric Engineering & Remote Sensing, 2006, 72(10): 1 179-1 188
- [7] 陈奋, 闫冬梅, 赵忠明. 基于无抽样小波的遥感影像薄云检测与去除[J]. 武汉大学学报·信息科学版, 2007, 32(1): 71-74
- [8] Maalouf A, Carré P, Augereau B, et al. A Band-let-Based Inpainting Technique for Clouds Removal from Remotely Sensed Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2009, 47(7): 2 363-2 371
- [9] 马建文, 顾行发, 冯春, 等. CBERS-02 卫星图像薄云的去除方法研究[J]. 中国科学 E 辑: 信息科学, 2005, 35(增刊 I): 89-96
- [10] Tseng D, Tseng H, Chien C. Automatic Cloud Removal from Multi-temporal SPOT Images[J]. Applied Mathematics and Computation, 2008, 205(2): 584-600
- [11] Helmer E H, Ruefenacht B. Cloud-Free Satellite Image Mosaics with Regression Trees and Histogram Matching [J]. Photogrammetric Engineering & Remote Sensing, 2005, 71(9): 1 079-1 089
- [12] Melesse A. A Comparison of Fuzzy vs. Augmented-ISODATA Classification Algorithms for Cloud-Shadow Discrimination from Landsat Images [J]. Photogrammetric Engineering & Remote Sensing, 2002, 68(9): 905-911
- [13] Roy D, Ju J, Lewis P, et al. Multi-Temporal MODIS-Landsat Data Fusion for Relative Radiometric Normalization, Gap Filling, and Prediction of Landsat Data[J]. Remote Sensing of Environment, 2008, 112: 3 112-3 130
- [14] Hagolle O, Huc M, Pascual D V, Et Al. A Multi-temporal Method for Cloud Detection, Applied to FORMOSAT-2, VEN μ S, Landsat and SENTINEL-2 Images [J]. Remote Sensing of Environment, 2010, 114(8): 1 747-1 755
- [15] Helmer E H, Ruefenacht B. A Comparison of Radiometric Normalization Methods when Filling Cloud Gaps in Landsat Imagery[J]. Canadian Journal of Remote Sensing, 2007, 33(4): 325-340
- [16] Rosefeld A, Delatorre P. Histogram Concavity Analysis as an Aid in Threshold Selection in Image Processing [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1983, 13(2): 231-235

第一作者简介: 吴炜, 博士生, 现从事遥感影像信息提取与分析研究。

E-mail: wuwei@irsa.ac.cn

Classified Linear Regression Based Landsat Image Cloud Removal Method

WU Wei^{1,2} LUO Jiancheng² SHEN Zhanfeng² WANG Weihong¹

(1 College of Computer Science and Technology, Zhejiang University of Technology, 288 Liuhe Road, Hangzhou 310023, China)

(2 Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, 3A Datun Road, Beijing 100101, China)

Abstract: An approach for cloud removal based on linear regression after image classification is proposed in this article. First of all, the clouds in a remote sensing image and its referenced data to be processed are detected, from which two cloud masks are built. Then, an ISODATA classification is applied to the referenced image with the cloud mask. Next, the masked part of the contaminated image is classified with the existing clusters of the referenced data using the minimum distance method. Last, the digital numbers of the cloudy areas of the contaminated image are replaced with by the prediction value of the referenced data calculated by the linear relationships determined between clusters on the referenced image and the corresponding contaminates done according to the pixel location. This algorithm is programmed to automatically detect and remove the clouds areas in Landsat images. The accuracy of cloud detection and the prediction of original values of the cloud cover are evaluated. Results show that the proposed method is effective.

Key words: image mosaic; cloud removal; cloud detection; classified linear regression

About the first author: WU Wei, Ph. D candidate, majors in remote sensing information processing and analysis.

E-mail: wuwei@irsa.ac.cn