

基于纹理分析和支持向量机的 极地冰雪覆盖区的云层检测

陈 刚¹ 鄂栋臣¹

(1 武汉大学中国南极测绘研究中心, 武汉市珞喻路 129 号, 430079)

摘 要: 使用 ASTER 卫星影像, 综合运用纹理分析和支持向量机的方法来检测云层, 特别是对薄卷云的检测。试验表明了本文方法的有效性。
关键词: 云层检测; 纹理分析; 支持向量机; 冰雪覆盖区; 极地地区
中图法分类号: P237.3

南极地区上空覆盖有大量的云层, 研究南极云层的分布特性对于极地地区甚至全球的物质平衡和能量交换都有重要的意义^[1]。因此, 首先要准确确定极地地区云量的分布和覆盖情况。南极地区提取云的两个主要难点是: 下垫面为冰雪覆盖, 光谱反射率很强, 在可见光波段与云层很难区分; 冰雪表面温度很低, 在热红外波段容易与云层形成温度反转。目前极地卫星影像云量评估常用算法主要有光谱阈值法^[2-4]、小波变换法^[5]以及神经网络法^[6]等。但以上这些算法或是需要对不同成像条件获取的影像采用不同经验阈值, 或是需要大量的训练样本来提高检测的准确率。本文综合使用纹理分析和支持向量机(support vector machines, SVM)方法用于极地冰雪覆盖区域的云层检测, 它不仅可以避免经验阈值的选取, 而且仅使用有限的训练样本且无需降低维数。

1 纹理特征分析

极地地区的下垫面为冰雪所覆盖, 虽然云与冰雪在可见光和热红外等单一波段上存在难以区分的特性, 但是由于生成云的大气环流、云内气流、水汽含量等差异, 导致云的形态、密度、云顶高度等各不相同, 在不同波段影像上, 云层与地面冰雪呈现不同的纹理特征。因此, 可以通过分别提取两者在多光谱上的纹理特征来获得分类所需的特征向量。纹理特征描述的是影像局部特性, 即

相邻像元之间的关系。描述纹理特征的方法很多, 本文采用灰度共生矩阵法(gray level co-occurrence matrix, GLCM)^[7]。经过试验选取其中 5 种二次统计量: 角度方向二阶矩(angular second moment, ASM), 对比度(contrast, CON), 相关性(correlation, COR), 熵(entropy, ENT)以及局部同性(homogeneity, HOM)。

灰度共生矩阵可以反映方向、相邻间隔和变化幅度的综合信息。数学式表示如下:

$$P(i, j, \delta, \theta) = \{ [f(x, y), (x + \Delta x, y + \Delta y)] \mid f(x, y) = i; f(x + \Delta x, y + \Delta y) = j; x = 0, 1, \dots, N_x - 1; y = 0, 1, \dots, N_y - 1 \} \quad (1)$$

式中, $i, j = 0, 1, \dots, L - 1$; (x, y) 是图像中的像素坐标; L 为灰度级数; N_x, N_y 分别为图像的行列数。因此, 所构成的灰度共生矩阵的第 i 行、第 j 列元素, 表示图像上所有在 θ 方向、相隔为 δ 、灰度分别为 i, j 的像素点对出现的频率。试验中 θ 的取值分别为 $0^\circ, 45^\circ, 90^\circ, 135^\circ$, δ 取 1。计算得到每个图像块的 4 个方向的灰度共生矩阵。为了提取具有旋转不变性的特征, 对 θ 为 $0^\circ, 45^\circ, 90^\circ$ 和 135° 的同一特征取平均值。

2 SVM 分类机理分析

SVM 针对有限样本情况来确定分类超平面, 使之在对未知样本进行估计时, 所得期望风险最小, SVM 从处理两类线性可分数据的分类问题推

广到处理线性不可分数据^[8,9]问题。在极地,由于云层与冰雪的辨识困难,所能获取的准确的训练样本比较有限,此时SVM将发挥它小样本训练的优势。与此同时,SVM提供的是与问题维数无关的复杂性的刻画,从而避免了在处理高维数据时存在的难以收敛、计算复杂和结果难以解释等问题^[10]。而且,试验发现云层与冰雪之间线性不可分。因此,可以在高维特征空间中构造线性决策边界来对应于输入空间的非线性决策。

对于云层与冰雪的分类,假设训练数据由 n 个样本组成 (x_i, y_i) , 其中 $i = 1, \dots, n; x \in R^d; y \in \{+1, -1\}$ 。分类超平面可以定义为:

$$f(x, \omega, b) = \text{sgn}(\omega \cdot \Phi(x) + b) \tag{2}$$

式中, $\Phi(x)$ 为非线性特征映射; ω 和 b 为系数。

最大间隔分类器总是产生一个没有训练误差的一致假设,然而实际数据总有一定的噪声存在,容易导致过拟合现象。因此,这里采用一阶范数软间隔来解决这个问题。引入松弛因子 ξ_i 的优化问题如下:

$$\begin{aligned} \text{minimise}_{w, b} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{subject to} \quad & y_i(\omega \cdot \Phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \dots, n \end{aligned} \tag{3}$$

式中, $C > 0$ 是盒约束的一个上限值。

转化为对偶问题:

$$\begin{aligned} \text{maximise } W(\alpha) = \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j K(x_i, x_j) \\ \text{subject to} \quad & \sum_{i=1}^n y_i \alpha_i = 0, C \geq \alpha_i \geq 0, i = 1, \dots, n \end{aligned} \tag{4}$$

式中,最优解 α_i 为满足 Karush-Kuhn-Tucker (KKT) 条件的 Lagrange 乘子:

$$\alpha_i [y_i(\omega \cdot x_i + b) - 1] = 0, i = 1, \dots, n \tag{5}$$

式中,只有支持向量所对应的 α_i 非零。

最终判决函数为:

$$f(x, w, b) = \text{sgn} \left\{ (w \cdot \Phi(x) + b) \right. \\ \left. \text{sgn} \left\{ \sum_{i=1}^n \alpha_i y_i K(x_i \cdot x) + b \right\} \right\} \tag{6}$$

3 试验数据选取及云层检测过程

本试验采用的是三景 ASTER 南极地区的遥感影像数据。ASTER 具有可见光近红外(VNIR)、短波红外(SWIR)和热红外(TIR)3个子系统的14个光谱通道,波谱范围从0.52 μm

到11.65 μm。试验分析极地云层和冰雪的不同特性,发现波段1(0.52 μm~0.60 μm)中两者反射率相近,波段4(1.60 μm~1.70 μm)中云的反射率高而雪的反射率低,波段6(2.185 μm~2.225 μm)中两者的反射率较波段4又有不同。热红外波段容易产生温度反转的现象,没有选用,而其他波段特性与1、4、6都比较相近,也省略。因此,试验中选取1、4、6这3个波段分别提取纹理特征。

提取纹理特征并利用SVM进行云层检测的过程如下。

1) 去除条带噪声。分析发现,条带噪声对于SVM分类的精度有很大的影响,可通过转换到频率域来消除^[11]。

2) 提取样本的纹理特征值。选取每个样本像素单个波段的5个纹理特征值,加上光谱值,共有18个特征值。采用5×5窗口可以较准确地反映窗口中心像元的特性。需注意的是,训练样本的选取必须准确和全面才能获得较高的分类精度。

3) SVM样本训练。由于云层与地面冰雪是线性不可分的,所以要把训练样本映射到高维特征空间。试验中采用的核函数是径向基函数(RBF): $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ 。 γ 和 C 是SVM分类中的重要参数,它们的选取对最终的最优超平面的获取有直接的影响。本文通过试验发现,选取 γ 为1/18、 C 为1所得效果最佳。

4) 云区检测。根据KKT条件得到Lagrange乘子的最优解 α^* , 训练获得最佳分类超平面。把影像上每个像素的特征向量 x_i 带入判别

函数 $f(x_i) = \text{sgn}(\sum_{j=1}^n y_j \alpha_j K(x_j, x_i) + b)$ 来判别是否是云点。这里使用的是SVM的二值分类算法——序贯最小优化算法。训练样本云区定义为+1,其他区域定义为-1。因此,当 $f(x_i) > 0$ 时,像素点判别为云点; $f(x_i) < 0$ 时,判别为非云点。

4 试验结果分析

本文试验使用ASTER卫星影像的3景南极不同区域的数据进行云层检测对比,且上空的云层既有层积云又有卷云,很有代表性,因此,能够在一定程度上反映出本文算法在不同情况下具有较强的适应性。检测结果如图1所示,图1左侧图像显示的是1、4、6波段合成的假彩色影像;

图 1 右侧图像是云层检测的试验结果影像, 其中, 浅灰色像素代表云层, 黑色像素代表非云地区。图 1(a)~ 图 1(b) 的“条带/序列”编号分别为 032/ 009、032/ 011、032/ 014。

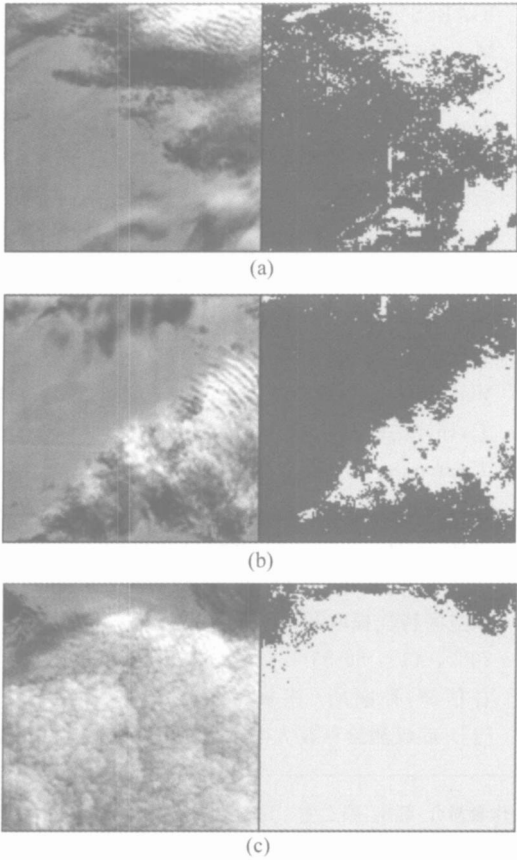


图 1 南极冰雪覆盖地区 ASTER 影像试验结果
Fig. 1 Results of Cloud Detection in Ice-snow Covered Areas in Antarctica

一般情况下, 云大致可分为积云、卷云和层云三大类型。影像 032/ 009 右上角和影像 032/ 011 右下方云层呈现波浪状, 是典型的卷积云; 影像 032/ 014 中部和下部的大部分区域被高积云所覆盖, 呈现出云母片状; 影像 032/ 009 右侧中部和右侧下方以及影像 032/ 014 的最上方分散着细致且很薄的卷云。在云区检测中, 薄的卷云由于其形态多样, 易被穿透, 很难将其与冰雪下垫面很好地区分开, 一直是云层提取的难点。然而, 从图 2 的结果影像可以看到, 本文所提出的算法不仅对于积云具有良好的检测效果, 而且对于薄卷云同样也非常敏感。例如在 032/ 009 和 032/ 014 上需要仔细辨识的薄卷云, 在结果影像上均被很好地反映出来。这里为了简化试验操作, 将所有类型的

云只作为一类进行检测。试验中以 64 像素× 50 像素大小的格网分别在三景影像上各自均匀选取 300 个样本进行测试, 最终获得 032/ 009、032/ 011、032/ 014 三景影像云层的检测精度分别为 91. 3%、94. 7%、93. 0%。

遥感影像每一景通常都有上百万个像素组成, 因此在提取影像纹理特征, 利用 SVM 进行样本训练的时候, 如果对整幅影像进行云区检测, 将是一个非常耗时的过程, 而且对计算机的硬件要求也很高。试验中, 笔者对同一景影像分别进行 2× 2 以及 4× 4 重采样, 即 1 024× 1 050 大小的原始影像通过重采样成为大小为 512× 525 和 256× 263 的影像, 然后使用本文提出的算法进行云层检测。结果表明, 一方面运算时间呈指数级下降(见图 1), 另一方面检测精度却变化不大(见表 1)。

表 1 不同大小影像上云层所占比率/ %

条带/ 序列	Tab. 1 Cloud Percentage of the Different Size of Images/ %			最大偏离值
	1 024× 1 050	512× 525	256× 263	
032/ 009	33. 12	36. 50	34. 00	2. 38
032/ 011	28. 58	27. 62	26. 50	2. 58
032/ 014	89. 42	90. 94	88. 31	2. 63

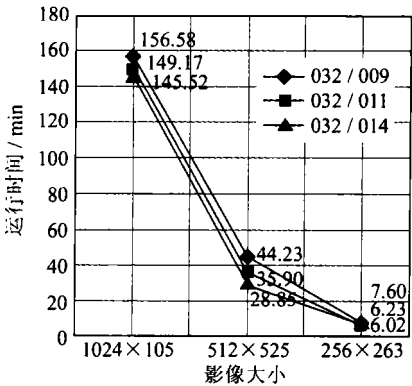


图 2 试验中三景影像在不同大小情况下算法处理所需时间

Fig. 2 Processing Times Depending on the Size of Images

表 1 中的最大偏离值表示了原图像与其他两种采样图像上云层所占比例的最大偏移量, 均不超过 3%。影像大小每缩小一半, 运算时间将近缩小到了原来的 1/ 4, 而云层的检测结果最大变化范围也只为 2% ~ 3%。因此, 相对于较长的运算时间而言, 这种方法在实际操作中将非常有意义。

5 结 语

本文试验结果清楚地表明了综合运用纹理特征分析和SVM方法来进行冰雪覆盖区云层检测具有很明显的效果。算法对薄卷云敏感的优点弥补了很多云检测算法对卷云检测效果较差的不足。由于SVM支持高维特征的输入,因此,今后的研究可以考虑增加其他特征来进一步提高云层检测的准确率,使得在对原始影像进行缩小再检测时,也能产生更加稳定的结果。

参 考 文 献

- [1] Curry J A, Rossow W B, Randall D, et al. Overview of Arctic Cloud and Radiation Characteristics [J]. *Journal of Climate*, 1996, 9(8): 1 734-1 764
- [2] Griffin M, Burke H H, Mandl D, et al. Cover Detection Algorithm for EO-1 Hyperion Imagery [J]. *Geoscience and Remote Sensing Symposium*, 2003, 7(1): 86-89
- [3] Gao Bocai, Yang Ping, Li Rongrong. Detection of High Clouds in Polar Regions During the Daytime Using the MODIS 1.375 μm Channel [J]. *IEEE Trans on Geoscience and Remote Sensing*, 2003, 41(2): 474-481
- [4] Choi H, Bindaschadler R. Cloud Detection in Landsat Imagery of Ice Sheets Using Shadow Matching Technique and Automatic Normalized Difference Snow Index Threshold Value Decision [J]. *Remote Sensing of Environment*, 2004, 91(2): 237-242
- [5] Muramoto K I, Saito H, Matsuura K, et al. Cloud and Ice Detection Using NOAA/AVHRR Data [C]. *IGARSS'97*, Toronto, Canada, 1997
- [6] McIntire T J, Simpson J J. Arctic Sea Ice, Cloud, Water, and Lead Classification Using Neural Networks and 1.6 μm Data [J]. *IEEE Trans on Geoscience and Remote Sensing*, 2002, 40(9): 1 956-1 972
- [7] Haralick R M, Shanmugam K, Dinstein I. Textural Features for Image Classification [J]. *IEEE Trans on Systems Man and Cybernetics*, 1973, 3(6): 610-621
- [8] Vladimir N V. 统计学习理论的本质[M]. 张学工,译. 北京:清华大学出版社,2000
- [9] Cristianini N, Shawe-Taylor J. 支持向量机导论[M]. 李国正,王猛,曾华军,译. 北京:电子工业出版社,2004
- [10] 骆剑承,周成虎,梁怡,等. 支持向量机及其遥感影像空间特征提取和分类的应用研究 [J]. *遥感学报*, 2002, 6(1): 50-55
- [11] 甘信铮,孙家柄. 南极卫星影像条带噪声的消除 [J]. *武汉测绘科技大学学报*, 1994, 19(4): 332-334

第一作者简介:陈刚,硕士生。现从事极地环境遥感研究。
E-mail: g_chen1982@hotmail.com

Cloud Detection Based on Texture Analysis and SVM over Ice-snow Covered Area

CHEN Gang¹ E Dongchen¹

(1 Chinese Antarctic Center of Surveying and Mapping, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: This paper presents one cloud detection method over ice-snow covered area in Antarctica. On account of different texture features of cloud and ice-snow area, five texture features are extracted based on GLCM. Then non-linear SVM is used to obtain the best model from training data. The experimental result shows that this algorithm performs well in cloud detection in Antarctica, especially to the thin cirrus. Furthermore, when images are resampled to the quarter or 1/16 size of the full size, cloud percentage does not change greatly, while the processing time decreases exponentially.

Key words: cloud detection; texture analysis; SVM; ice-snow covered area; polar region

About the first author: CHEN Gang, postgraduate, majors in environmental remote sensing of polar areas.

E-mail: g_chen1982@hotmail.com