

基于正交匹配追踪的云检测算法研究

王 毅^{1*}, 何明元¹, 葛晶晶², 项 杰¹

(1. 国防科技大学 气象海洋学院, 江苏 南京 211101;

2. 中国人民解放军 31110 部队, 江苏 南京 210016)

摘 要: 云的定量识别在卫星数据反演中非常重要。云检测结果的质量直接影响各种反演产品的准确性。云检测实际上是一种目标识别和分类的过程, 检测的目的是为了特征提取, 因此, 大量信号与系统中的新兴算法都被运用于云检测的技术研究中来。匹配追踪算法是近年来发展起来的非常有效的特征提取算法, 而正交匹配追踪算法更能有效提高信噪比。因此文中利用正交匹配追踪算法与多通道阈值法相结合, 进行了卫星云图云检测的相关研究。通过 MODIS 的云检测试验表明, 将正交匹配追踪算法应用到多光谱云图的信息处理中可以有效提高云检测的精度。

关键词: MODIS; 多光谱; 特征提取; 云检测; 正交匹配追踪

中图分类号: P407.8 **文献标志码:** A **DOI:** 10.3788/IRLA201948.1203003

Cloud detection algorithm based on the Orthogonal Matching Pursuit

Wang Yi^{1*}, He Mingyuan¹, Ge Jingjing², Xiang Jie¹

(1. Institute of Meteorology and Oceanography, National University of Defense Technology, Nanjing 211101, China;

2. Unit No. 31110 of PLA, Nanjing 210016, China)

Abstract: Quantitative identification of cloud is very important in meteorological satellite data retrieval, and the result of cloud detection affects the accuracy directly. In fact, the cloud detecting technology is actually a process of distinguishing the objects and background, and the purpose of detection is to extract cloud features. Therefore, much signals processing and system algorithms have been applied to the technology of cloud detection. The matching pursuit algorithm (MP) is a very effective algorithm for feature extraction, which is developed in recent years, and the Orthogonal Matching Pursuit algorithm (OMP) can improve the signal-to-noise ratio more effectively. In this paper, Orthogonal Matching Pursuit algorithm and multi-channel threshold method were combined to carry out relevant research on cloud detection of MODIS data. Based on the MODIS cloud detection results, it could be proved that the integrative algorithm of multi-channel threshold combined with the Orthogonal Matching Pursuit algorithm would be more effective to cloud detection.

Key words: MODIS; multi-spectral; feature extraction; cloud detection; Orthogonal Matching Pursuit

收稿日期: 2019-10-11; 修订日期: 2019-11-21

基金项目: 国家自然科学基金(41576171); 国防科技大学自然科学基金; 江苏省自然科学基金(BK20161465)

作者简介: 王毅(1975-), 女, 副教授, 博士, 主要从事卫星遥感反演等方面的研究。Email: wangyi_rsc@126.com

0 引言

在卫星数据定量反演应用过程中,云检测起着非常关键的作用^[1-3]。关于云检测国内外已经进行了大量的工作,一般说来可以分为统计法和阈值法两大类,另外,还有近年来发展的综合云检测方法^[4]。阈值的选取规则基于有云和晴空特征之间的差异。根据经验选择适当的阈值用以区分有云区域和晴空区域。目前,云检测的算法主要研究集中在多阈值法与其他算法的组合算法研究中,例如:多阈值与数学形态学结合法^[5],多阈值与各种聚类综合方法^[6-14]等。

云检测可以看作作为一个目标识别的过程,依据一定的特征将卫星图像上的有云区域检测出来,最终实现提取有云区域,并将这些信息与背景区域分开,即删除所有受云污染的区域并过滤出晴朗的空气区域。因此,特征提取在云检测过程中起着重要作用。因此,云检测的主要问题就转化为:(1)判定包含最多目标信息量的光谱通道;(2)尽可能去除背景信息的影响;(3)确定最佳信道组合。在这个基础上保证可以最大程度地提取有用信息,同时,将计算中涉及的数据量最小化,并加快云信息提取的速度^[15]。在此基础上,关于信号与系统领域中发展起来的很多新兴算法都被不断地应用到云检测技术研究中,如投影追踪技术、匹配追踪技术、支持矢量机等^[16-18]。

投影追踪算法是 1970 年初开发的用于多元数据分析的一种新的数学方法。这种算法适用于处理和分析非线性、非正态分布的高维数据,不需要进行训练推理,只需要投影诊断即可实现降维。匹配追踪算法是近年来发展起来的新兴算法,具有良好的降维去噪效果。但是该算法必须与其他算法相结合,才能保证迭代计算结果收敛,基于这个原因发展起来的主要相关算法有:匹配跟踪 (Matching Pursuit, MP)、正交匹配跟踪(OOMP)、正则化正交匹配跟踪 (ROMP)、压缩采样匹配跟踪(CSMP)算法等^[19-21]。经高光谱图像相关研究结论证明,匹配追求算法能够有效提高信噪比,针对高光谱图像的非监督目标检测问题取得了较好的检测效果,具有良好的并行性能。

正是基于这种算法本身的优越性,因此文中在针对多光谱数据的云检测处理中采用了匹配追踪算

法,并与业务化多阈值法相结合,改进了云检测算法的精度。

1 MP 算法原理

匹配跟踪(Matching Pursuit, MP)是一种迭代寻优算法,假设该过程在有限空间中进行,将观测到的光谱视为该空间的一个矢量。给定有限的 N 维空间的光谱信号 $x(\omega):x=[x_1, x_2, \dots, x_N]^T$ 。在信号的处理过程中,作用 B 组光学滤波器,那么光谱成分信号 $x(\omega)$ 的变化可用线性变换来描述:

$$y_i = \int w_i(\omega)x(\omega)d\omega \quad i=1, \dots, B \quad (1)$$

由该表达式可见,让信号通过 B 个滤波函数作用,事实上等同于将线性混合光谱信号向 B 维有限空间做映射。也就是说,在信号上应用 B 个滤波器功能等效于将原始频谱信号向 B 维空间做投影。因而上述滤波问题就转化为解决一个 $N \times B$ 的矩阵 $W=[w_1, w_2, \dots, w_B]$ 的问题,由此就实现了从 N 维到 B 维的转化。用矩阵表达形式表示为:

$$y=W^T x$$

也即

$$\begin{pmatrix} y_1 \\ \vdots \\ y_B \end{pmatrix} = \begin{pmatrix} w_{11} & \dots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{B1} & \dots & w_{BN} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \quad (2)$$

此处需要解决的问题是,如何将多光谱图像分解为大量光谱混合信号。在此基础上,将所得混合光谱信号进行降维和重组。假设随机的线性混合信号模型为:

$$x = \sum_{k=1}^M a_k S_k + n = Sa + n \quad (3)$$

式中: S_k 为单位标准谱分量 (Unit-form component spectra); a_k 为相对应的辐值 (Abundance); n 为由于模型不完善或噪声造成的误差矢量。给定的条件是: n 是独立的、正态分布的高斯噪声,并且其平均值为 0、协方差矩阵为 $\sigma^2 I$, 这样光谱就能被划分为目标和背景。假设噪声信号和背景信号彼此独立,原始光谱信号可表示为:

$$x = \sum_{k=1}^P a_k S_k + \sum_{k=P+1}^P a_k S_k + n = S_t a_t + S_b a_b + n \quad (4)$$

式中: M 为 P 、 Q 的集合, 假设 P 和 Q 表示目标和背景的集合中所包含的目标矢量和背景矢量的数目, 并且满足 $Q=M-P$; S_t 为目标矢量; S_b 为背景矢量; n 为噪声矢量。

判断条件为: $H_0 \quad x=S_b a_b+n$, 则目标不存在;

$H_1 \quad x=S_t a_t+S_b a_b+n$, 则目标存在。

为消除背景信号 S_b 的干扰影响, 可以引入去除背景信号算子 $P_b^\perp=I-S_b(S_b^T S_b)^{-1} S_b^T$ 。该算子的作用就是将背景信号最小化, 提高信噪比, 代入公式(4)可得:

$$P_b^\perp x=P_b^\perp S_t a_t+P_b^\perp n \quad (5)$$

公式(5)表明, 去除背景算子 P_b^\perp 作用在信号 x 上, 等于投影到了与空间 B 垂直的空间上, 该空间为 B 空间的补。两个空间中的矢量互相垂直, B 为 S_t 是目标矢量所构成的空间, B^\perp 为干扰背景信号 S_b 构成的空间, 空间 B 与 B^\perp 正交并且互补。干扰背景信号矢量在 B^\perp 上的投影为 0, B^\perp 中不存在背景信号。因此, 将线性混合信号向空间 B^\perp 作映射后, 背景信号被去除, 达到了去除背景信号的目的。当混合信号中噪声信号非常小, 相对于目标和背景信号可忽略, 那么得到投影 $z=P_b^\perp S_t a_t$ 。

假设测得的光谱中有效波长有 N 个, 则理想的带滤光片可以用具有单位范数的矩形函数表示。定义 $D=\{w_\gamma\}$ 是有可能矩形函数的集合, 共有 $N(N+1)/2$ 个函数, 可选择组合方式的数目非常大。因此, 这个问题也可以用上文中的结论来处理, 也即所选择的滤波函数应使目标矢量和背景矢量分别处于目标空间和背景空间, 或者说使目标信号向背景信号做正投影, 也即用 $z=P_b^\perp S_t a_t$ 来描述线性滤波函数的作用。但在实际情况中, 用 B 组滤波器实现信号的降维去噪, 需要采用最小二乘法实现。最简单的过程是只考虑单一类目标的情况, 也就是 $P=1$ 。

$$E[W]=\min \|z-Wa\|^2=\min \left\| z-\sum_{i=1}^B a_i w_i \right\|^2 \quad (6)$$

为解决 B 个滤波函数难以精确描述投影函数的问题, 需要选择最优的滤波函数组合, 从而在最小误差内近似为投影信号 z 。选择一个滤波函数 w_{γ_0} , 表

示从集合 D 中挑选一组滤波函数。当信号 z 在滤波器上投影时, 根据上述理论背景信号将被消除, 而只有最小的剩余量 R , 则可在近似范围内求解 w_i , 即:

$$z=(w_{\gamma_0}^T z) w_{\gamma_0}+R \quad \gamma_0=\operatorname{argmin}_{\gamma \in \Gamma} \left\| z-\left(w_{\gamma}^T z \right) w_{\gamma} \right\| \quad (7)$$

从而问题变成求解 L 阶迭代循环计算, 第 L 次迭代可以表达为 $R_L=w_{\gamma_L} w_{\gamma_L}^T R_L+R_{L+1}$, 因而有:

$$z=\sum_{k=0}^L \left(w_{\gamma_k}^T R_k \right) w_{\gamma_k}+R_{L+1} \quad (8)$$

这里初始值取 $R_0=z$ 。

2 云检测分析

文中将采用中分辨率成像光谱仪(MODIS)的数据进行云检测试验。MODIS 具有 36 个通道, 光谱范围覆盖 $0.4\sim 15 \mu\text{m}$; 空间分辨率分别为 250、500、1 000 m, MODIS 的多波段大大增加了对地球环境的观测力和识别力。文中选择 2018 年 11 月 3 日 (0215UTC) 的 MODIS L1B 数据 (试验区域: 北纬 $20\sim 28^\circ\text{N}$, $115\sim 125^\circ\text{E}$)。MODIS 数据在使用前, 必须经过定标处理。MODIS L1B 数据中包含经过定标的反射波段的辐亮度、反射率、红外波段的辐亮度。

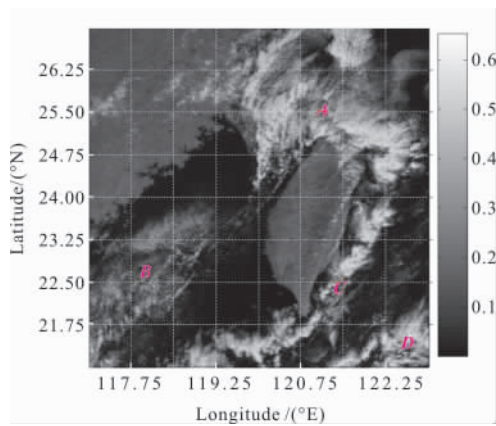
在实际算法处理过程中, 用光谱的线性组合表示一个像素。因此, 逐行读取 MODIS 数据, 并将其作为同时包含有云目标和无云背景的线性混合信号处理; 为简化计算及提高计算效率, 选取比较简单的矩形滤波函数 $\omega(n)=R_N(n)$:

$$\omega(n)=R_N(n)=\begin{cases} 1 & 0 \leq n \leq N-1 \\ 0 & \text{other} \end{cases}$$

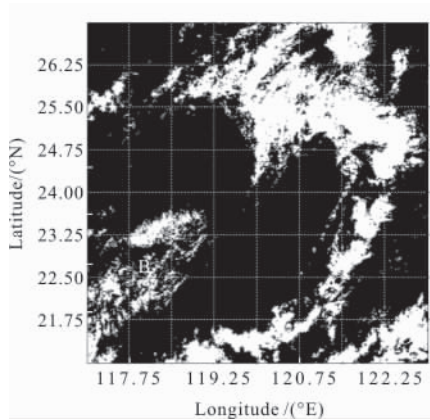
式中: n 为滤波函数的长度。输出的 ω 是由滤波函数所构成的 n 阶矢量。文中经过试验取云检测特征量: (1) 通道 1 反照率; (2) 通道 6 反照率; (3) 通道 26 反照率; (4) 通道 31 亮温; (5) 分裂窗通道亮温差值。在通道 1 和通道 6, 云和下垫面反照率分布有重叠。在 31、32 通道, 云和其他下垫面存在较大的亮温差。结合这些特征值, 最终可以得到映射函数和由匹配追踪计算得到的近似滤镜函数。这种投影函数通过特征量或特征量组合给出在廓线权重函数, 由此达到消除背景的目的。这样, 就将廓线中占有较大比例的部分突出显示在图像上。针对 2018 年 11 月 3 日

的 MODIS L1B 数据进行处理，最终结果如图 1、2 所示。

图 1 为直接采用正交匹配追踪进行云检测的结果。如图 1(a)所示,图中标注的 A、B、C、D 这 4 个部



(a) 原图
(a) Original image



(b) 云检测结果
(b) Cloud detection result

图 1 基于正交匹配追踪算法的云检测

Fig.1 Cloud detection based on Orthogonal Matching Pursuit algorithm

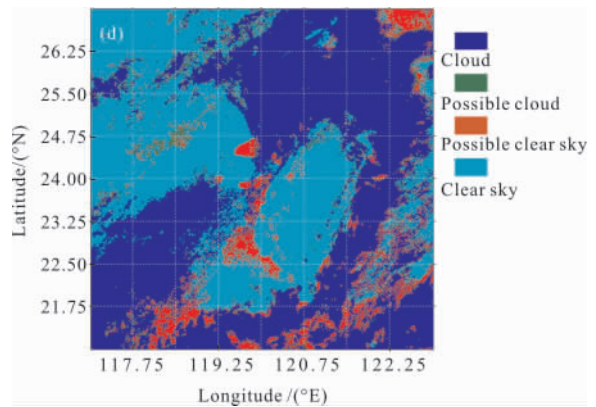
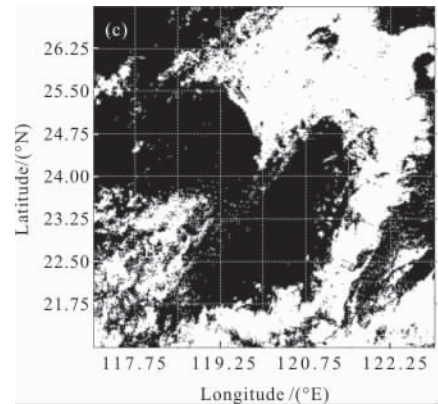
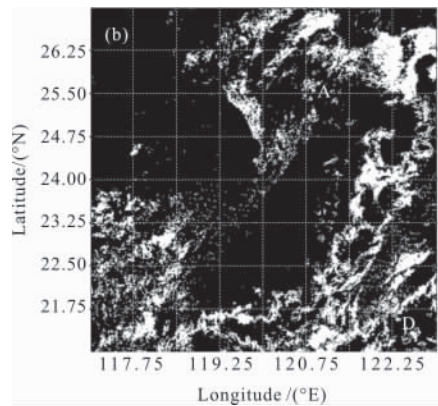
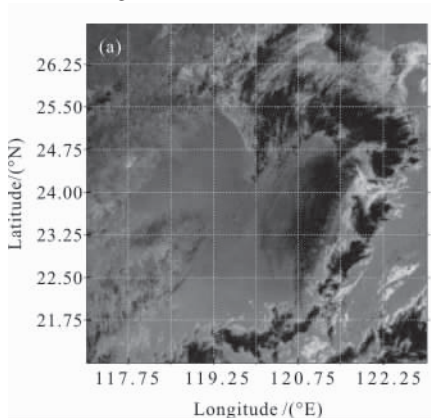


图 2 基于多通道阈值与正交匹配追踪算法的云检测

Fig.2 Cloud detection based on the multi-channel threshold combined with the Orthogonal Matching Pursuit algorithm

分明显有云存在,在 A、C、D 处的云层较厚,在 B 处有薄卷云。经正交匹配追踪多次迭代后的检测结果可以将这些区域的更精细的云特征提取分析出来。如图 1(b)所示,除了 A、B、D 处的厚云,这几处边缘的薄云也被检测出来。特别是 C 处的东南象限的薄卷云也被比较精确地检测出来。

图 2 为结合 MODIS 通道特征与正交匹配追踪的检测结果。图 2(a)为经过处理的亮温差值,图 2(b)

则为亮温差作为特征值薄云检测的结果,图 2(c)为云检测的结果,图 2(d)为最终云检测输出结果。如图所示,结合多阈值与正交匹配追踪算法相结合,可以有效检测出有云与无云区域。

3 结 论

文中利用 MODIS 数据,利用多阈值法和改进的正交匹配追踪算法实现了云检测。正交匹配追踪算法是应用于高光谱图像分析领域中的一种高效优化算法,已经在图像边缘特征分析和提取领域得到较好的结果。运用该算法针对多光谱图像可有降维去噪,提高计算效率和精度。与其他算法比较,可以提取更多信息,不仅可以提高精度,并且能够保证计算效率,适用于目前卫星产品的业务化应用。

参考文献:

- [1] Wang Yi. Development and Application of the New International Earth Observation System [M]. Beijing: China Meteorological Press, 2006: 4. (in Chinese)
王毅. 国际新一代对地观测系统的发展及其主要应用[M]. 北京: 气象出版社, 2006: 4.
- [2] Goodman A H, Henderson-Sellers A. Cloud detection and analysis: A review of recent progress [J]. *Atmospheric Research*, 1988, 21: 203-228.
- [3] William B Rossow, Leonid C Garder. Cloud detection using satellite measurements of infrared and visible radiances for ISCCP [J]. *Journal of Climate*, 1993, 6: 2341-2369.
- [4] Chan J. Comment on changes in tropical cyclone number, duration, and intensity in a warming environment [J]. *Science*, 2006, 311: 1713.
- [5] Ma Fang, Zhang Qiang, Guo Ni, et al. The study of cloud detection with multi-channel data of satellite[J]. *Chinese Journal of Atmosphere Science*, 2007, 31(1): 119-128. (in Chinese)
马芳, 张强, 郭妮, 等. 多通道卫星云图云检测方法的研究[J]. 大气科学, 2007, 31(1): 119-128.
- [6] Li Yanbing, Li Yuanxiang, Zhai Jingqiu. A method of extracting and representing morphological features of satellite cloud image [J]. *Journal of Nanjing Institute of Meteorology*, 2006, 29(5): 682-687. (in Chinese)
- [7] Wu Yongming, Zhang Ren, Jiang Guorong, et al. A fuzzy clustering method for multi-spectral satellite image [J]. *Journal of Tropical Meteorology*, 2004, 20(6): 689-696. (in Chinese)
吴咏明, 张韧, 蒋国荣, 等. 多光谱卫星图像的一种模糊聚类方法[J]. 热带气象学报, 2004, 20(6): 689-696.
- [8] Shi Chunxiang, Qu Jianhua. Cloud classification for NOAA-AVHRR data by using a neural network[J]. *Acta Meteorologica Sinica*, 2002, 2: 123-128. (in Chinese)
师春香, 瞿建华. 用神经网络方法对 NOAA-AVHRR 资料进行云客观分类[J]. 气象学报, 2002, 2: 123-128.
- [9] Hong Mei, Zhang Ren, Sun Zhaobo. A high-dimension feature spaces clustering and corresponding weather classification for multi-spectral satellite images [J]. *Journal of Remote Sensing*, 2006, 2(10): 42-48. (in Chinese)
洪梅, 张韧, 孙照渤. 多光谱卫星云图的高维特征聚类与降水天气判别[J]. 遥感学报, 2006, 2(10): 42-48.
- [10] Chen Gangyi, Ding Xuyi, Zhao Liyan. An automatic pattern recognition techniques of cloud based on fuzzy neural network [J]. *Chinese Journal of Atmosphere Science*, 2005, 29(5): 169-176. (in Chinese)
陈刚毅, 丁旭羲, 赵丽妍. 用模糊神经网络自动识别云的技术研究[J]. 大气科学, 2005, 29(5): 169-176.
- [11] Guo Hongtao, Wang Yi, Liu Xiangpei, et al. Integrated optimal method of cloud detection with meteorological satellite data [J]. *Journal of PLA University of Science and Technology (Nature Science Edition)*, 2010, 11(2): 221-227. (in Chinese)
郭洪涛, 王毅, 刘向培, 等. 卫星云图云检测的一种综合优化方法[J]. 解放军理工大学学报(自然科学版), 2010, 11(2): 221-227.
- [12] 费文龙. 变分方法在 GMS-5 气象卫星云图处理中的应用研究[D]. 南京: 南京理工大学, 2004.
- [13] Shuai Chunxiang, Wu Rongzhang, Xiang Xukang. Automatic segmentation of satellite image using hierarchical threshold and neural network[J]. *Journal of Applied Meteorological Science*, 2001, 12(1): 69-78. (in Chinese)
师春香, 吴蓉璋, 项续康. 多阈值和神经网络卫星云图云系自动分割实验[J]. 应用气象学报, 2001, 12(1): 69-78.

- [14] Xue Juntao, Liu Zhengguang, Liu Huanzhu. Application of wavelet transforms on the boundary processing of the infrared satellite cloud image[J]. *Journal of Tianjin University (Science and Technology)*, 2002, 35 (6): 736–739. (in Chinese)
薛俊滔, 刘正光, 刘还珠. 小波变换在云图边缘处理中的应用[J]. 天津大学学报, 2002, 35(6): 736–739.
- [15] Harsanyi J C, Chang C I. Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach [J]. *IEEE Trans Geosci Remote Sens*, 1994, 32(4): 779–785.
- [16] Mallat S, Zhang Z F. Matching pursuits with time–frequency dictionaries [J]. *IEEE Transactions on Signal Processing*, 1993, 41(12): 3397–3415.
- [17] Wen J M, Zhou Z C, Wang J, et al. A sharp condition for exact support recovery of sparse signals with orthogonal matching pursuit [C]//2016 IEEE International Symposium on Information Theory, IEEE, 2016: 2364–2368.
- [18] Needell D, Vershynin R. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2010, 4(2): 310–316.
- [19] Needell D, Tropp J A. CoSaMP: iterative signal recovery from incomplete and inaccurate samples [J]. *Applied and Computational Harmonic Analysis*, 2009, 26(3): 301–321.
- [20] Manolakis D, Shaw G. Detection algorithms for hyperspectral imaging applications [J]. *IEEE Signal Process Lett*, 2002, 19(1): 29–43.
- [21] Tropp J A, Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit [J]. *IEEE Transactions on Information Theory*, 2007, 53 (12): 4655–4666.