

高分一号光学遥感数据自适应云区识别

蒙诗栋^{1,2}, 庞勇², 张钟军^{1*}, 李增元²

(1. 北京师范大学 信息科学与技术学院 北京 100875;

2. 中国林业科学研究院 资源信息研究所 北京 100091)

摘要: 光学卫星遥感数据在获取过程中易受云层干扰,云区识别是光学遥感数据应用及分析的一个基础但重要的步骤,高效的云区识别技术对节省数据收集成本和提高数据利用效率具有较强的现实意义。同态滤波算法是经典的基于单幅影像的云区识别方法之一,该算法具有计算快速方便、云区检测精度较高的优点,然而识别的云区范围极大程度取决于同态滤波器截止频率的位置。同态滤波截止频率通常采用经验值,显然经验截止频率无法适应批量遥感数据的自动处理需求。针对以上问题,本文通过建立输入影像频谱能量与截止频率的关系,结合白度指数(Whiteness Index)和形态学算子,实现对国产高分辨率光学卫星高分一号(GF-1)遥感数据的批量云区识别处理。与传统同态滤波方法相比,该算法能根据影像频谱能量自适应判定同态滤波时采用的截止频率,具有更强的适用性。通过对 98 景 GF-1 多光谱数据进行随机点人工目视标记精度检验,精度检验结果表明该算法对云区有较好的检测效果,总体识别精度达 93.81%。该算法对 GF-1 遥感数据能进行批量化云区检测,获得高精度的云区掩膜结果,并有效降低高反射率地物造成的误识率。

关键词: 云区识别; 自适应; 同态滤波; GF-1 遥感数据; 截止频率

中图分类号: O438 **文献标识码:** A

Self-adaptive cloud detection approach for GaoFen-1 optical remote sensing data

MENG Shi-Li^{1,2}, PANG Yong², ZHANG Zhong-Jun^{1*}, LI Zeng-Yuan²

(1. College of Information Science and Technology, Beijing Normal University, Beijing 100875, China;

2. Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China)

Abstract: Cloud detection for remote sensing imageries is a fundamental as well as significant step due to the inevitable existence of large amount of clouds in the optical remote sensing data. A highly efficient cloud detection approach is capable of saving data collection cost and improving data utilization efficiency. Homomorphic filtering algorithm is one of the most commonly methods that based on single-scene image for detecting clouds. This algorithm has the advantage of fast computation and high accuracy in cloud areas detection. However, the detected cloud areas are heavily dependent on the cut-off frequency of the homomorphic filter. The homomorphic filtering progress usually uses cut-off frequency with empirical value which might not be applicable to large amount of intricate input data. Therefore, this paper aims to construct the relationship between the image spectra power and the filter cut-off frequency. Based on the domestic high spatial resolution optical remote sensing data GF-1, this research makes the detection of clouds could be process to achieve a bulk deal. Our approach make the cut-off frequency self-adaptive changes rather than used empirical value when compared with the tradi-

收稿日期: 2018-04-23 修回日期: 2018-06-20

Received date: 2018-04-23 revised date: 2018-06-20

基金项目: “十三五”国家重点研发计划“多尺度落叶松人工林生长预测”(2017YFD0600404); 国家自然科学基金“基于高分辨率遥感数据的森林生物多样性监测”(31570546); 中央高校基本科研业务费专项资金项目“L波段森林的微波辐射与传输特性研究”(2015KJJC12)

Foundation items: Supported by the National Key Research and Development Program of China during the 13th Five-Year Plan Period “Growth and Yield Prediction of Larch Plantation at Multi-scales” project (2017YFD0600404); the National Natural Science Foundation of China “Forest biodiversity monitoring using high resolution remote sensing data” project (31570546) and the fundamental research funds for the central universities “The research of microwave emission and transmission characteristics of forest at L-band” project (2015KJJC12)

作者简介(Biography): 蒙诗栋(1990-),女(瑶族),广西都安人,博士研究生,主要从事遥感影像分析和林业遥感应用研究。E-mail: mengsl@mail.bnu.edu.cn

* 通讯作者(Corresponding author): E-mail: zzj@bnu.edu.cn

tional homomorphic filtering, thus it could be able to meet more complicated scenarios. Further, the post-processing steps including whiteness index, spectral threshold, and morphological opening and closing operators are applied to coarse cloud mask to optimize results. We have tested on 98 GF-1 high resolution multispectral imageries, results indicated that our approach is capable of detecting cloud as well as haze areas with high accuracy of 93.81%. This novel self-adaptive method shows its great application potential for real-time and high efficient cloud detection, meanwhile reduced the error detection rates caused by high reflectance ground objects.

Key words: cloud detection, self-adaptive, Homomorphic filtering, GaoFen-1 remote sensing data, cut-off frequency

PACS: 02.70.Hm, 42.68.Wt, 42.79.Ci

引言

光学遥感卫星传感器通常采用的可见近红外波段(0.38~0.90 μm)无法穿透云层,在实际应用时难以获取云层覆盖区域内的地物信息.国际云气候学计划(International Satellite Cloud Climatology Project, ISCCP)根据 ISCCP-FD 辐射数据集(ISCCP - Flux Data)对遥感数据的云覆盖量进行估测,研究结果显示全球每年获取的光学遥感数据中云覆盖量约占 66%^[1].低纬度热带/亚热带地区的遥感数据中存在的云层遮挡现象相对其他区域更加密集且无法避免,严重影响了遥感数据的判读、解译以及应用^[2].卫星遥感影像的云区识别是对云覆盖区进行判断并使之与无云区域分离的过程,广泛应用于云图分析、图像融合、影像合成以及遥感数据定量反演等应用中.因此,云区识别技术作为运用遥感数据前的预处理步骤,在节省数据收集成本、提高影像利用率的同时也为数据的进一步应用提供保证.

目前,中国高分辨率对地观测系统已成功发射多颗光学遥感卫星.其中,2018年3月31日成功发射的高分一号(GF-1)02、03和04星与2013年4月26日已发射的GF-1号01星共同构成了我国陆地资源调查监测业务星座.我国高分系列光学卫星实现了全天候全天时地对地观测,丰富了高分辨率国产空间影像数据集,满足了快速机动应急变化需求,能更好地服务于我国土地、地质、交通和林业等领域的精细化业务应用^[3].GF-1多光谱数据成像波段包括蓝(0.45~0.52 μm)、绿(0.52~0.59 μm)、红(0.63~0.69 μm)以及近红外(0.77~0.89 μm)波段.国际上普遍使用的中低分辨率星载遥感影像,如AVHRR、MODIS、风云系列、Landsat系列和Sentinel系列等,除了具有可见光和近红外波段之外,通常还包括一个或多个有利于识别云雾的热红外和水汽吸收波段^[4-7].在有限波段的局限下,对高分辨率多光

谱数据进行云区判定时,极易造成识别的云区与其它高反射率地表物体(如道路、房屋、裸地等)发生混淆.因此,对日益增多的高分辨率遥感数据进行高效且准确的云区检测仍是一个十分具有挑战且亟待解决的问题.

云层在可见光和近红外波段相对于其他地物类型具有较高的反射率,且光谱反射率随波长增加而逐渐降低^[8].在遥感影像中,较厚云层与清晰地物的边界明显,而薄云由于边缘过渡缓慢因此难以界定,即云区与背景场的灰度色阶差别越大就越容易识别^[9].基于人工数字化解译的方法的云区识别具有较高的精度,但耗时较长,且多数时候仍难以对云区边界进行明确划分,无法满足现今海量遥感数据的时效性和高效性的需求.高分辨率遥感数据云区检测领域大致包含两类技术方法:基于多时相影像和基于单景影像的云区识别算法.基于多时相影像的云区识别,是以不同时相的无云数据作为参考从而确定云区位置与范围.Han等(2014)^[10]通过简单阈值法分割出目标影像的厚云区,利用尺度不变特征变换(Scale Invariant Feature Transform, SIFT)提取参考影像与目标影像的同名特征点,对云区范围精准定位.该算法的云区识别率较高(约89%),但计算特征匹配点时运算量大,且不适用于薄云区域;Helmer和Ruefenacht(2005)^[11]采用多时相无云数据作为参考,利用回归树(Regression Tree)算法检测并估算目标影像中云和云阴影覆盖下的像元值;Tseng等(2008)^[12]通过混合像元线性分解方法对多时相SPOT影像进行云区判定.基于单景影像的云区识别方法无需参考数据,因此具有更强的适用性和实效性.Lu(2007)^[13]利用基于规则的最大/最小滤波器和统计纹理信息判定IKONOS影像的云/云阴影,识别算法对高分辨率影像有效,但云判结果噪声较大,极易受高亮地物干扰.Marais等^[14]通过光谱阈值法分割有云和无云区,阈值法简单快速,但

最优阈值难以判定,同时不可避免地将高反射率地物与云区相混淆;Fisher(2014)^[15]采用区域生长型分水岭算法,在阈值划分出的“云标记点”(Cloud Markers)上分割出云区;Bai等(2016)^[16]提出了多特征融合的识别云的方法,通过结合GF-1/2数据丰富的纹理信息、形状因子、光谱特征以及植被指数等,利用支持向量机分类算法进行云区和无云区的划分,验证精度优于91.45%。

同态滤波算法仍是最常用的针对单幅遥感影像的云区识别方法之一^[17-20],同态滤波算法通过高通滤波器在频域范围内抑制云区分布的低频区域,使得云区辐射信息被削弱,非云区辐射信息得到增强。影像经过同态滤波处理后,原先亮度较高的云层反射率降低,而相对较暗的非云区域反射率升高。将结果与原始影像对比,即得到云区掩膜的结果。然而,同态滤波的去云效果极大地依赖高通滤波器的截止频率,之前的研究针对截止频率多数采用经验值,显然经验截止频率无法适应云区层次复杂的遥感数据处理,且对厚云和薄云等不同厚度的云区无法达到最佳识别效果。针对上述问题,本文以GF-1遥感图像为例,通过构建影像径向能量谱与高通滤波器的截止频率之间的量化关系,使截止频率根据不同输入影像自适应地调整,从而为实现GF-1数据的云区自动识别奠定了基础。

1 算法介绍

总体思路基于单景影像本身特征,利用同态滤波算法进行云区识别。算法改进之处在于采用云雾厚度图代替可见光波段(常用蓝波段)作为输入数据进行同态滤波,并改进滤波器截止频率判断方式,根据输入数据特征动态调整截止频率,最后利用白度指数和形态学开闭算法做后处理运算。本文以GF-1遥感图像为例,提出的云区自动检测算法主要包括以下三个步骤:(i)计算GF-1多光谱影像的云雾厚度图;(ii)根据图像径向能量谱判定滤波器截止频率;(iii)云识别结果后处理。运算流程如图1所示。

1.1 云雾厚度图

Chavez^[21-22]的方法假设在某场景中总是存在某些反射值极低的像元,其数值(Digital Number, DN)接近于0。由于受到云雾反射的影响,使上述暗像元的像素值并未完全的“纯黑”,即像元DN值远大于0,则认为该类暗像元的DN值客观上反映了场景中云雾的厚度。Makarau^[23]等改进了Chavez的算法,用

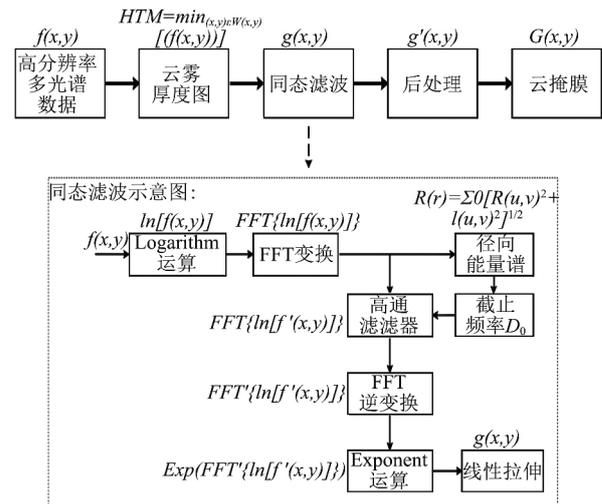


图1 GF-1 卫星数据云区自动识别流程

Fig.1 Workflow of GF-1 satellite images cloud detection approach based on homomorphic filtering

局部窗口寻找多个暗目标(Dark Objects)的算法替代了在全局仅寻找一个暗目标的做法,使得获得的云雾厚度图(Haze Thickness Map, HTM)的细节更明显。HTM表示为:

$$HTM(x_0, y_0) = \min_{(x, y) \in W(x', y')} (f(x, y)) \quad (1)$$

其中, $W(x', y')$ 为中心位于点 (x', y') 的 $w \times w$ 局部窗口,窗口大小 w 的设置与影像分辨率有关^[21],文中设为5倍的像元分辨率,即 5×5 局部窗口; $f(x, y)$ 为输入波段,通常采用极易受云雾影响的蓝波段,但为避免蓝波段算出的HTM造成云雾厚度过检测的现象,本文采用Liu^[24]等建议的线性差值合成波段进行HTM计算。计算得到 $HTM(x_0, y_0)$ 后,再利用 3×3 的中值滤波器进行平滑,并采样到原始图像大小。HTM能更好的抑制地面高反射地物的亮度,缓和过度检测的效果。

1.2 同态滤波云区识别

1.2.1 同态滤波原理

在云雾区域,遥感数据成像模型包括太阳辐射经云层反射部分和太阳辐射经地物反射后再穿透云层的部分。用 $f(x, y)$ 表示传感器接收到的信号, $f_c(x, y)$ 为云层反射率,反映云的分布, $f_g(x, y)$ 表示地物反射率,云的成像模型可简化为式(2)^[25-27]。云在影像的时空域中表现为亮度高,变化平稳,纹理结构相对简单。在频率域中,表示云层分布的 $f_c(x, y)$ 集中于低频区域,影像中变化较快的细节内容的 $f_g(x, y)$ 分布在高频区域。经过对数-傅里叶变换(式(4))将式(3)进行空-频转换,同时对频谱采用高通滤波器(式(5)),即在频域上抑制云层反射率的低频成

分,从而将占据低频成分的云信息从影像中剔除出去。 $Filter(u, v)$ 为高通滤波器的传递函数,本文采用高斯差分(Difference of Gaussian, DoG)高通滤波器(式(6)),其中: γ_H 为高频增益, γ_L 为低频增益,通常设置 $\gamma_H \geq 1$ 且 $\gamma_L \ll 1$; $D(u, v)$ 代表点 (u, v) 到傅里叶变换中心点的距离; D_0 为高通滤波器截止频率.再将滤波后的 $S(u, v)$ 做傅立叶逆变换,完成频-空转换(式(7)).最后进行指数变换(式(8)),得到同态滤波处理后的图像 $f'(x, y)$.同态滤波具体运算公式如下所示:

$$f(x, y) = f_i(x, y) \cdot f_r(x, y) \quad (2)$$

$$\ln(f(x, y)) = \ln(f_i(x, y)) + \ln(f_r(x, y)) \quad (3)$$

$$F\{\ln(f(x, y))\} = F\{\ln(f_i(x, y))\} + F\{\ln(f_r(x, y))\} \quad (4)$$

$$S(u, v) = L(u, v) Filter(u, v) + H(u, v) Filter(u, v) \quad (5)$$

$$其中 Filter(u, v) = (\gamma_H - \gamma_L) [1 - e^{-\frac{D(u, v)^2}{2D_0^2}}] + \gamma_L \quad (6)$$

$$s(x, y) = \ln(f_i(x, y)) + \ln(f_r(x, y)) \quad (7)$$

$$f'(x, y) = \exp\{\ln(f_i(x, y))\} + \exp\{\ln(f_r(x, y))\} \quad (8)$$

1.2.2 云区识别

同态滤波通过将低频成分与高频成分分离,抑制了云分布的低频分量并增强高频信息.经同态滤波处理后的图像,云覆盖像元的 DN 值相对于原始输入受到削弱,而无云覆盖像元的 DN 值则提高^[15].同态滤波处理的结果已改变原有亮度信息,因此先将变换结果 $f'(x, y)$ 拉伸到原来输入数据的幅度:

$$g(x, y) = \min_i + (f'(x, y) - \min_o) \times \frac{\max_i - \min_i}{\max_o - \min_o} \quad (9)$$

其中 \max_i 和 \min_i 分别表示输入数据 $f(x, y)$ 的最大值和最小值, \max_o 和 \min_o 分别表示 $f'(x, y)$ 的最大值和最小值,并由此得到云掩膜结果:

$$Cloud\ mask = \begin{cases} \text{云区 } f(x, y) > g(x, y) \\ \text{无云 } f(x, y) < g(x, y) \end{cases} \quad (10)$$

1.3 高通滤波器截止频率判定

1.3.1 截止频率

同态滤波的关键是要根据输入数据确定合适的高通滤波器的截止频率.在高通滤波器的截止频率点处,滤波器会极大衰减截止频率前的低频信号,使

输出信号强度衰减为输入信号的一半,并允许截止频率后的信号通过.对式(7)中的 DoG 高通滤波器, γ_H 决定了高频信号的幅度, γ_L 表示低频信号衰减的程度.图 2-(b) 展示四种不同截止频率的 DoG 高通滤波器频域特征曲线,其中高频增益 $\gamma_H = 1.0$,低频增益 $\gamma_L = 0.05$.

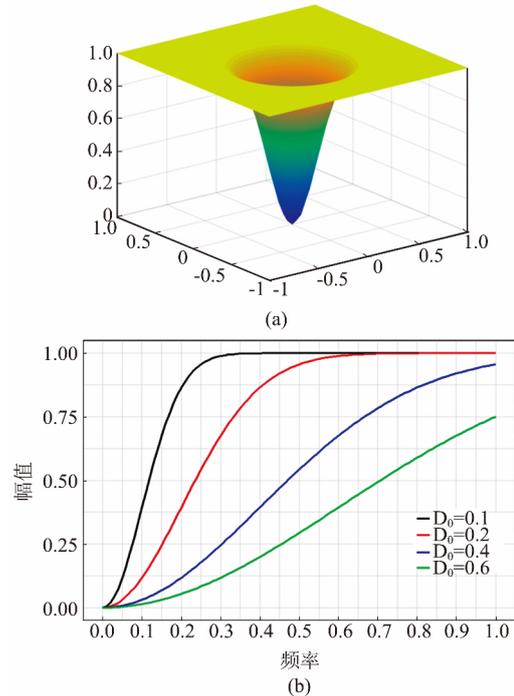


图 2 (a) DoG 高通滤波器频率响应 $Filter(u, v)$ ($D_0 = 0.1$) 及 (b) 四种截止频率 ($D_0 = 0.1, 0.2, 0.4, 0.6$) 频域特征曲线

Fig.2 (a) Frequency response of DoG high-pass filter and (b) characteristic curves of DoG filter with four kinds of cut-off frequencies ($D_0 = 0.1, 0.2, 0.4, 0.6$)

同态滤波运算中,输入影像的傅里叶变换系数在频域内会与高通滤波器相乘.此时,截止频率 D_0 取值会影响滤波结果, D_0 选择不当则会造成图像信号过滤波或者欠滤波.若 D_0 取值较低,高通滤波器无法完全抑制低频的亮度信息,将过多识别云区范围之外的其他高亮地物(图 3-b);若 D_0 取值较高,地表物体的细节信息会被过度削减,云区周围的薄云将无法有效检测(图 3-c). D_0 的选择与照度场(Illumination Field)和反射场(Reflectance Field)的频谱能量有关,需要大量试验进行确定^[18].经验截止频率无法满足大批量云区层次复杂的遥感数据处理需求,实际应用中应根据不同输入影像确定其对应的合适截止频率.本文将从图像的照度场和反射场的频谱能量出发,通过分析照度分量的能量占比,

从而选择合适的 D_0 .

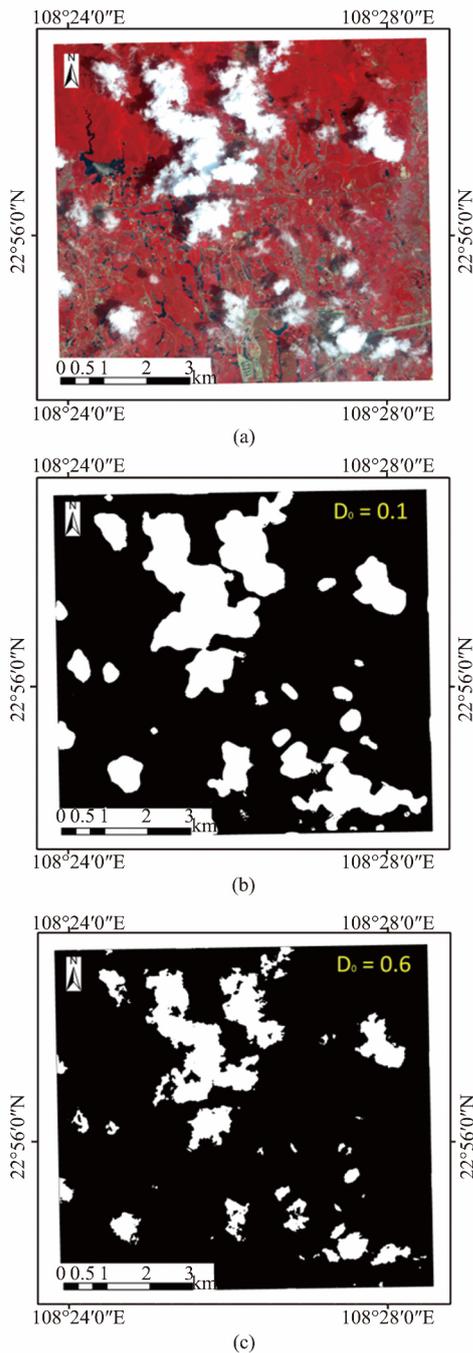


图3 截止频率分别为0.1和0.6的GF-1影像云区识别结果

Fig.3 Cloud detection results of GF-1 image with cutoff frequencies of 0.1 and 0.6, respectively

1.3.2 傅里叶径向频谱

图像经过二维快速傅里叶变换处理(式(11)),将原点平移至频谱中心后,此时频率原点值 $F(u=v=0)$ 表示影像的平均灰度级(即直流分量), $F(u=v=0)$ 值越大表示图像平均灰度越高,反之越低。

$$F(u, v) = \iint_{-\infty}^{\infty} f(x, y) e^{-j2\pi(ux+vy)} dx dy \text{ 或记为 } F(u, v) = |F(u, v)| e^{j\theta} \quad (11)$$

其中,幅值谱表示为 $|F(u, v)| = \sqrt{R^2(u, v) + I^2(u, v)}$, $R(u, v)$ 和 $I(u, v)$ 分别为傅里叶变换复数频谱的实部与虚部.若给定的输入影像尺寸为 $m \times n$,频率 u 和 v 分别代表正余弦波形在影像内的重复次数,根据奈奎斯特频率得到最高正余弦频率为 $u = \frac{m}{2}$ 和 $v = \frac{n}{2}$.假设光照条件是均匀的,则照度场的傅里叶频谱只有直流分量,随着光照不均匀度增加,开始出现反射变量,在傅里叶频谱上表示为除了直流分量,同时增加了谐波分量.此时,用傅里叶变换径向频谱 $R(r)$ (Fourier Radius Spectrum) 来表示影像频域谐波圆环内各个方向的总能量之和(式(12)),反映频域中谐波频率圆环内的能量大小.各谐波圆环能量占总能量的百分比 $R(r)\%$ 如式(13).

$$R(r) = \sum_{\theta} (\sqrt{R^2(u, v) + I^2(u, v)}) \quad (12)$$

$$R(r)\% = \frac{R(r)}{\sum_r R(r)} \quad (13)$$

式中: $r = 0, 1, 2, \dots, \sqrt{u^2 + v^2} - 1$; $\theta = 0^\circ \sim 360^\circ$.图4显示不同的云量覆盖(即照度场强度不同)对应的 $R(r)$ 曲线.频率原点处 $R(r)$ 幅值最高,随着谐波频率逐渐趋于奈奎斯特频率, $R(r)$ 幅值越来越小并衰减趋于0.云量较多时图像径向频谱整体能量水平越高,谐波能量收敛速度越快,此时,体现云的照度分量高度集中于低频部分,此时截止频率设置在低频处较为合理;云量较少或无云覆盖时图像径向频谱能量水平较低,谐波能量收敛趋近于0的速度更慢,此时照度分量较弱,体现反射分量的谐波能量相对较强,因此需提高截止频率以充分抑制地物信息,避免造成欠滤波。

1.4 后处理

同态滤波算法除了正确识别出云区,影像中的其他高反射率地物包括建筑、道路和裸地等,也存在被过度识别的风险.因此采用白度指数(Whiteness Index)和形态学算法对同态滤波云掩膜结果进行优化后处理。

1.4.1 白度指数

白度指数(Gómez-Chova)^[28]等算法基于云区光谱曲线特征(呈现高亮白色且在可见光波段光谱曲线平坦的特性),已应用于MERIS(Medium Resolu-

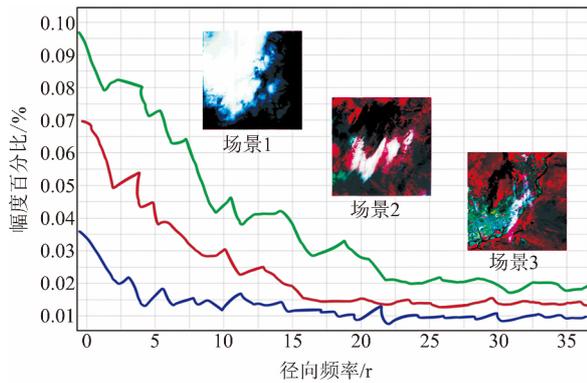


图 4 不同云覆盖场景对应下的傅里叶变换径向频谱曲线
Fig. 4 Radius spectra curves of Fourier transformation for different cloud covered scenes

tion Imaging Spectrometer) 窄波段多光谱卫星数据云区识别,并由 Zhu 与 Woodcock^[6]将该算法拓展到 Landsat 系列卫星的云区识别产品中.云区的光谱曲线在可见光波段的值相对较高,通过对可见光波段进行光谱密度的积分,可得出云亮度指数(Cloud Brightness Index) f_{Br} :

$$f_{Br} = \frac{1}{\lambda_{max} - \lambda_{min}} \sum_{\lambda_i \in (B_v)} \frac{f(\lambda_{i+1}) + f(\lambda_i)}{2} (\lambda_{i+1} - \lambda_i) \quad (14)$$

其中 λ_{max} 表示可见光区域内的最大波长, λ_{min} 表示最小波长,单位为 nm, B_v 为可见光波段范围, $f(\lambda_i)$ 表示波长为 λ_i 波段的 DN 值.从式(14)看出 f_{Br} 并非对波段值进行简单平均,而是采用了考虑波段的波长能量分布的平均值.可见光光谱曲线呈现的“白谱”表示光谱曲线平坦,对应的光谱曲线一阶微分值较低.由于存在噪声和量化误差,因此微分值的精度会受到不同程度的影响.因此将每个像元与 f_{Br} 误差的微分作为白度指数 f_{Wh} :

$$e(\lambda_i) = |f(\lambda_i) - f_{Br}| \quad (15)$$

$$f_{Wh} = \frac{1}{\lambda_{max} - \lambda_{min}} \sum_{\lambda_i \in (B_v)} \frac{e(\lambda_{i+1}) + e(\lambda_i)}{2} (\lambda_{i+1} - \lambda_i) \quad (16)$$

其中 $e(\lambda_i)$ 表示波长为 λ_i 波段的 DN 值与 f_{Br} 的误差.显然,像元越白(DN 值高且各波段值差异较小) $e(\lambda_i)$ 越小,白度指数越低,反之白度指数越高.利用白度指数算法对同态滤波结果进行优化,能将 DN 值较高但非白色的物体剔除,降低错误识别率.对 f_{Wh} 进行阈值判定,将 f_{Wh} 小于 0.1 的部分判定为白色物体(图(5)).

$$White_{Objects} = f_{Wh} < 0.1 \quad (17)$$

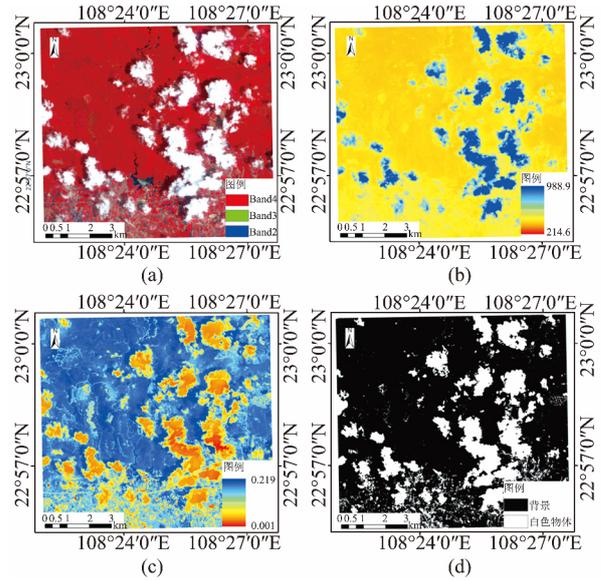


图 5 (a) GF-1 号影像 (b) 云亮度指数 (c) 白度指数及 (d) 对白度指数采用阈值判定 (< 0.1) 的结果
Fig. 5 (a) GF-1 remote sensing image, (b) cloud brightness index, (c) whiteness index and (d) the result of bright object when applied threshold (< 0.1) to the whiteness index

1.4.2 形态学滤波

数学形态学滤波(Mathematical morphology)思想为利用形态学算子(或称结构元素)对二值图像进行滤波.基本形态学滤波算法包括腐蚀、膨胀、开运算与闭运算,其中滤波效果与结构元素有关^[29],结构元素包含线形、矩形或圆形等形状.本文结构元素选用直径为 7 个像元的圆形模板,采用先闭运算再开运算的处理策略对云区掩膜进行后处理优化.闭运算能填充云区内的空隙同时不显著改变其面积,再进行开运算能消除云区边界离散噪声使轮廓平滑.图 6 展示未进行形态学处理的云区掩膜结果(图 6-f)与采用形态学处理后的云区掩膜结果(图 6-g)对比.经形态学处理后(图 6-g),噪声和破碎的小斑块云点被去除,部分误识别的高亮地物也被滤除掉,取得较好的后处理效果.

1.5 精度评价方法

为了验证本文云区识别算法的精度,试验采用人工目视解译对 98 景 GF-1 数据的云掩膜结果进行精度验证.首先根据云区掩膜范围在对应的 GF-1 遥感数据内随机选取若干样本点(5~8 个点);其次在云区掩膜范围外构建 20~40 个像元的云掩膜缓冲区,并在缓冲区内随机选取若干点.最后分别计算云区中心(O)到云区边界距离(D_0)以及云区中心(O)到各随机点的距离(D_1).其中,云区中心点 O 为云区边界最小外接圆的圆心.试验总共采样到 619 个

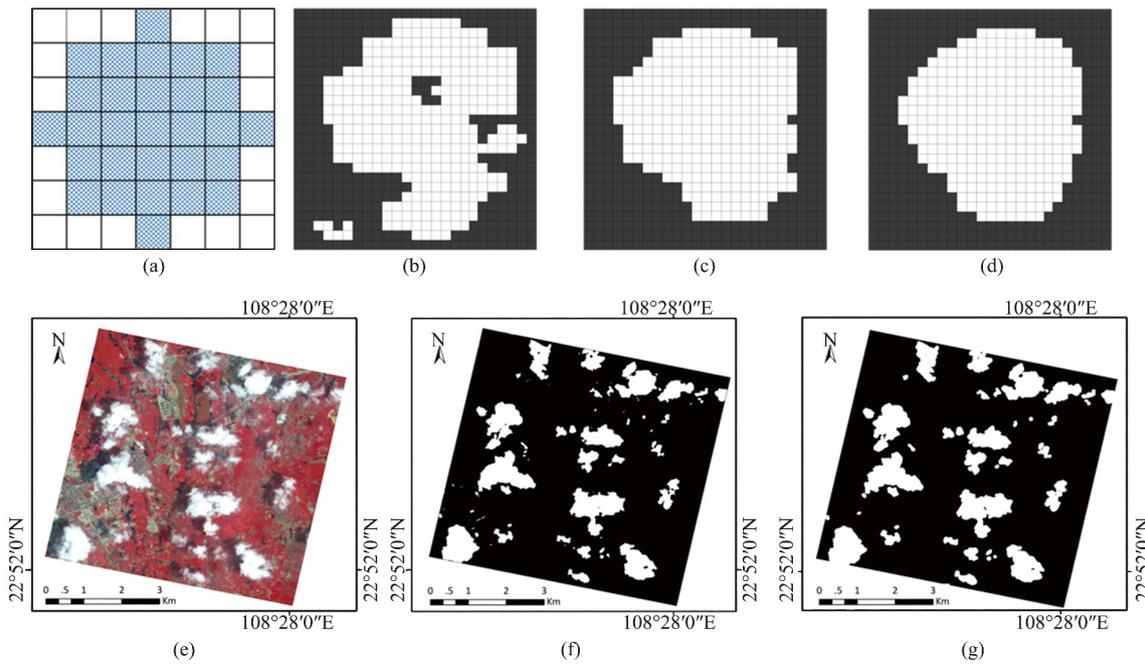


图6 GF-1 影像云区掩膜形态学处理实例 (a) 直径为7个像元的圆形结构元素 (b) 云区模拟图像,白色为云像元,黑色为背景像元 (c) 云区模拟图像经过闭运算处理的结果 (d) 图(c)再经过开运算处理的结果 (e) GF-1 影像云区实例 (f) 基于同态滤波算法识别的云区 (g) 图(f)经形态学优化后的云区

Fig.6 Morphology filtering results of GF-1 cloud mask (a) circular structure element with diameter for 7 pixels (b) simulate image for cloud areas, white color is represent cloud pixels while black is background (c) image after closed operation (d) result then after open operation (e) GF-1 remote sensing image with cloud cover (f) cloud detection result for GF-1 image when applied morphology filtering (g) optimized result after conducting closed and open morphology operation to (f)

样本点,在云掩膜内采样的点若为云区,则标定为“1”,反之标定为“0”;在缓冲区内采样的点若为非云区,则标定为“1”,反之标定为“0”,最后统计随机采样点的准确率.云掩膜精度验证示意图如图7所示:

2 结果分析与讨论

2.1 截止频率判定

对GF-1影像进行裁剪,选取尺寸为 1024×1024 像元的79幅图像作为实例数据,空间分辨率8 m,实例涵盖不同云量覆盖程度的图像.同时计算出实例数据的傅里叶变换径向频谱 $R(r)$,统计径向频谱总能量 $(R[0:r])$ 、图像平均灰度值 $(R[0])$ 以及其占总能量的百分比 $(R[0]\%)$ 、谐波总能量 $(R[1:r])$ 以及其占总能量的百分比 $(R[1:r]\%)$,以及在试验过程中采用的最优截止频率 (D_0) .对于选择最优的同态滤波器的截止频率,首先设置初始值,再以3为步长逐步增加到 r ,目视云判结果得出最优 D_0 区间,取区间均值得到最优 D_0 值.表1展示了部分实例数据的统计结果,图像(1)~(8)显示8幅云量分布从多到少的实例数

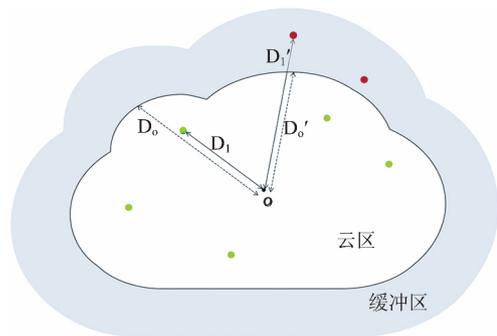


图7 云掩膜精度检验示意图:绿色点为云掩膜内选取的随机点,红色点为在缓冲区内选取的随机点, D_0 为云区中心 O 到云区边界距离, D_1 为点 O 到各随机点的距离

Fig.7 Sketch map of cloud mask accuracy assessment: green points and red points are sample points choose randomly from cloud mask areas and cloud buffer areas, respectively. D_0 shows the distance between cloud center O and cloud boundary, D_1 is the distance between cloud center O and the sample point

据.随着图像(1)~(8)云量覆盖逐渐减少,图像径向频谱总能量 $R[0:r]$ 和平均灰度 $R[0]$ 逐渐递减,而平均灰度对总能量的占比 $R[0]\%$ 却不断增

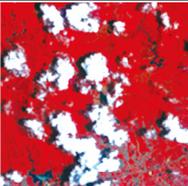
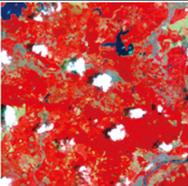
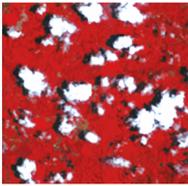
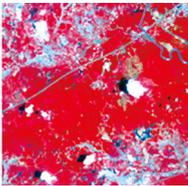
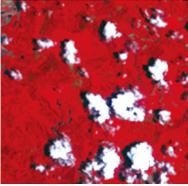
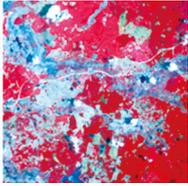
大. 其中云量最多的图像(1)的 $R[0]$ 最高(6.01), 其对应的 $R[0]\%$ 则最低(24.47%); 接近于无云的图像(8)的 $R[0]$ 为 5.38, $R[0]\%$ 为 59.43%. 从实例对应的 D_0 可看出, 云量覆盖最多的图像(1)的 D_0 位于较低频部分; 随着云量减少, D_0 逐渐朝着高频方向移动, 云量最少的图像(8)则位于频率中低频部分.

基于图 4 径向频谱曲线的分析结果, 当影像的照度分量高度集中于低频部分(场景云量多), 截止频率应设在低频处, 当影像反射分量能量较强时(场景云量少), 截止频率应设置较高. 且根据表 1 结果, 图像云覆盖越高时 $R[0]\%$ 越低, 其 D_0 值低; 图像云覆盖越少则 $R[0]\%$ 越高, 其 D_0 值高, 表明 D_0 与 $R[0]\%$ 呈现正相关性. 通过对实例数据的截止频率与图像的平均灰度能量占比进行回归分析, 建立 $R[0]\%$ 与 D_0 处的径向频谱能量 $\sum_0^{D_0} R(r)$ 的定量模型(式 18), 从而可根据不同输入确立的 $R[0]\%$ 得到对应 D_0 的值. 由图 8 可见, $R[0]\%$ 与 $\sum_0^{D_0} R(r)$ 之间相关性较好, 建模精度较优, 决定系数 R^2 达 0.9275, 模型鲁棒性强.

$$\sum_0^{D_0} R(r) = 0.194 * R[0]\% + 0.883 \quad (18)$$

表 1 八幅不同云量实例数据的傅里叶变换径向频谱统计结果

Table 1 Statistic results of Fourier transformation radius spectra for eight sample data with different cloud cover

序号	图像	$R[0:r]$	$R[0]$	$R[0]\%$	$\%R[1:r]\%$	D_0	序号	图像	$R[0:r]$	$R[0]$	$R[0]\%$	$\%R[1:r]\%$	D_0
1		25.56	6.01	24.47%	75.53%	123	5		15.95	5.78	36.21%	63.79%	258
2		25.17	5.93	23.56%	76.44%	165	6		12.24	5.79	47.35%	52.65%	354
3		20.94	5.89	28.11%	71.89%	207	7		12.23	5.76	47.07%	52.93%	339
4		19.61	5.80	29.56%	70.44%	213	8		9.05	5.38	59.43%	40.57%	378

$$* r = \sqrt{\left(\frac{1024}{2}\right)^2 + \left(\frac{1024}{2}\right)^2} - 1 \approx 723$$

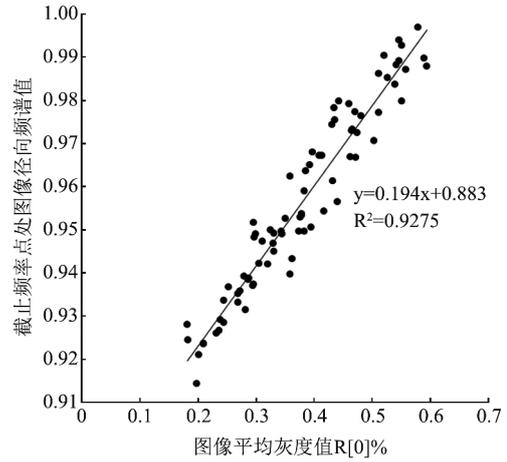


图 8 截止频率模型散点图
Fig. 8 Scatterplot of cutoff frequency model

2.2 实验区与数据

本次试验将广西壮族自治区国有高峰林场地区作为研究区. 广西壮族自治区位于我国西南部, 地处低纬度(20°54' -26°24' N), 北回归线横贯全区中部, 属亚热带季风季候区. 中国气象局气象数据中心 (<http://data.cma.cn/data/weatherBk.html>) 统计结

果显示,广西壮族自治区是我国年平均总云量覆盖较高的地区之一,年均云覆盖总量占七成以上.本研究对 98 景 GF-1 多光谱卫星数据进行试验,包括了 GF-1 高分辨率多光谱数据和宽幅多光谱数据,数据成像时间涵盖 2013 年 10 月至 2016 年 10 月.

2.3 应用实例

为了进一步验证算法的有效性与合理性,利用本文提出的云区识别流程和截止频率判定方法,对 98 景 GF-1 卫星数据进行试验.图 9 展示了两种传感器(高分辨率多光谱 MSS 和宽幅多光谱 WFV)拍摄情况下共 4 景影像的云区识别结果.试验区不同

云量分布的以及的 4 景影像进行算法评估.为 GF1 多光谱遥感数据以及其对应云区检测及除云后的影像结果.

图 9(a)(b)为 GF-1MSS 数据(8 m)云区识别结果,(c)(d)为 GF-1WFV 数据(16 m)云区识别结果.图 9(a)影像被大量云区覆盖,北部为植被覆盖浓密的山区森林区域,南部为城市区域,算法在山区以及城市中均对云区检测获得较好效果.对于无干扰的森林区域,识别的云区区域精度较高;图 9(b)影像无云分布,而城市区域则存在大量高反射亮度的地物,算法将小部分地物虚警检测为云区,总体上仍避免了大量高反射亮度地物的干扰,对复杂环境下的云区检测展现较好的稳定性和鲁棒性.图 9(c)(d)结果表明,对于不同空间分辨率的 GF-1 数据,算法同样展示出较好的适应性,云区检测结果精度较高.

2.4 精度验证

基于本文 1.5 节提出的精度评价方法,对 98 景 GF-1 影像的云区识别结果进行人工解译和精度验证.结果显示,总体云区识别精度为 93.81%,其中,云区内掩膜精度最高,达 97.69%,云区掩膜缓冲区精度略低,为 90.15%.从精度误差分析来看(图 10),误差较多产生于云区边界轮廓处,靠近云中心点以及缓冲区远离点误差较小.从云区中心逐渐向云区边界过渡时,采样点的正确率逐渐下降;从云区边界向缓冲区边缘过渡时,正确率又逐渐升高.

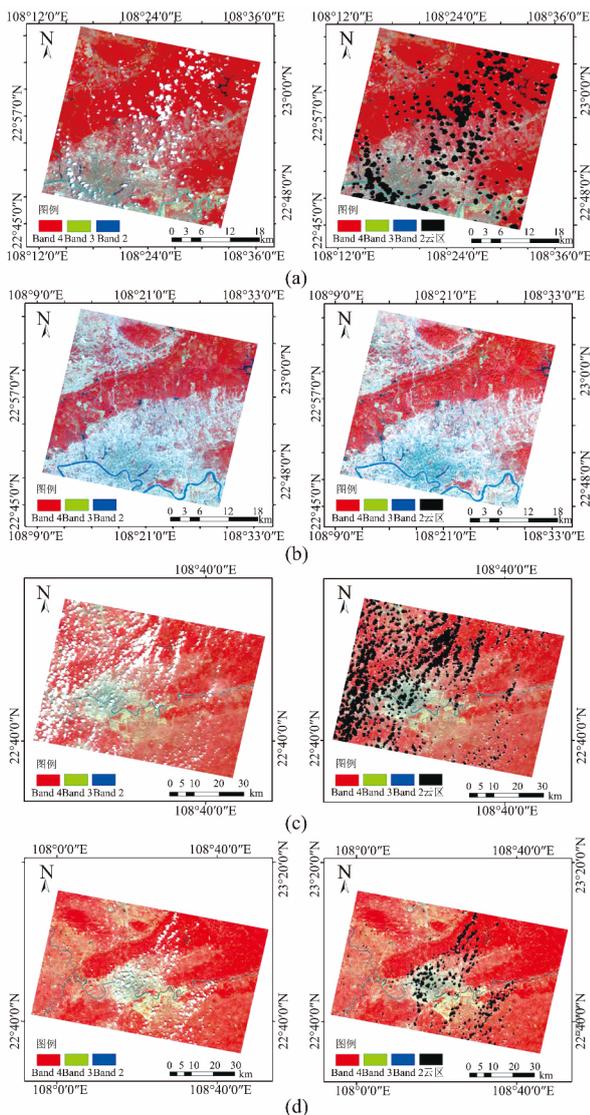


图 9 GF-1 遥感数据云掩膜与云区识别效果,GF-1 高分辨率多光谱数据(8 m) (a) 20150808 (b) 20160229 和宽幅多光谱数据(16 m) (c) 20131022 (d) 20161003

Fig.9 GF-1remote sensing cloud masks and cloud removed results ,GF-1 high resolution multispectral data (8 m) (a) 20150808 ,(b) 20160229 and wide field of view multispectral data (16 m) (c) 20131022 ,(d) 20161003

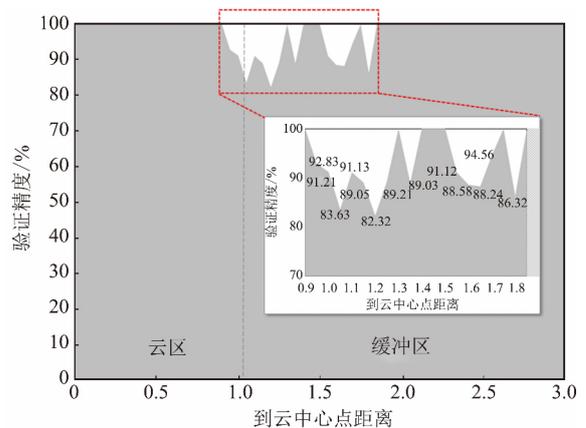


图 10 人工解译云掩膜验证精度

Fig.10 Cloud mask assessment accuracy acquired by artificial interpretation

2.5 时间复杂度

为了测试算法时间复杂度,在 ENVI + IDL 二次开发环境下开发相应程序,进行了两组对比实验.第

一组测试了经典同态滤波算法的运算时间,第二组测试本文的改进同态滤波算法的运算时间.测试时间从读入栅格数据开始直至生成云区掩膜文件.实验采用的硬件为 Intel Xeon 1.70 GHz CPU 和 32G 安装内存.本文的改进算法的时间包括计算云厚度图、判断截止频率、同态滤波以及后处理操作,从表 2 中可以看出,相对于经典的仅计算同态滤波的过程,改进算法的复杂度并不高,显示出该算法在批量处理遥感图像时的优势.

表 2 经典同态滤波算法与本文算法的运算时间统计
Table 2 Statistic results of operation time for classic homomorphic filtering and our develop algorithm

约耗时/s	512 × 512	1024 × 1024	2048 × 2048	4096 × 4096	8192 × 8192
经典同态滤波	0.065	0.255	1.206	6.793	30.4553
改进同态滤波	0.251	1.044	4.252	17.462	74.244

2.6 分析与讨论

通常,会利用蓝波段作为同态滤波处理输入进行频域滤波分解,但蓝波段对云雾信息敏感,极易造成云雾过检测^[24].图 1 算法识别流程提出采用 HTM 代替蓝波段,HTM 能更好地抑制地面高反射地物的亮度进而减少过检测.图 11 选择了具有复杂地物信息的云覆盖影像,分别利用蓝波段和 HTM 作为输入进行同态滤波云识别.图 11(b)的 HTM 显著突出云亮度信息,降低背景亮度,形成云与地物背景的较大反差.在相同截止频率下(并非最优截止频率),基于蓝波段(图 11(c)) 和 HTM(图 11(d)) 的云识别结果不同.图 11(e) 展示蓝波段处理结果相较于 HTM 处理结果的差值图,红色部分为图 11(c) 比图 11(d) 多检测区域,绿色部分为少检测区域.图 11(c) 除了识别出云区外,还将部分人工建筑和噪声点过度判定为云区,造成大量过识别.相对于蓝波段的处理结果,图 11(d) 除了能准确地识别云区,其优势在于能较好地抑制地物亮度的影响,通过配合后续的形态学与白度指数优化,更好地剔除其余高亮干扰.

为了适应云区层次复杂的批量云区识别处理,构建了影像的径向能量谱与滤波器的截止频率的量化关系,使截止频率能根据输入影像自适应地变化,从而为实现自动化批量的云区识别.为此,选取了 79 幅云覆盖情况具有代表性且场景丰富的 GF-1 多光谱实例数据进行模型训练,实际测试了 98 景数据并开展了精度检验,验证结果证实算法可行且有效.在实例数据的选择上,无云场景及有云场景需同时考虑,且对于有云场景尽量涵盖不同云量分布情况

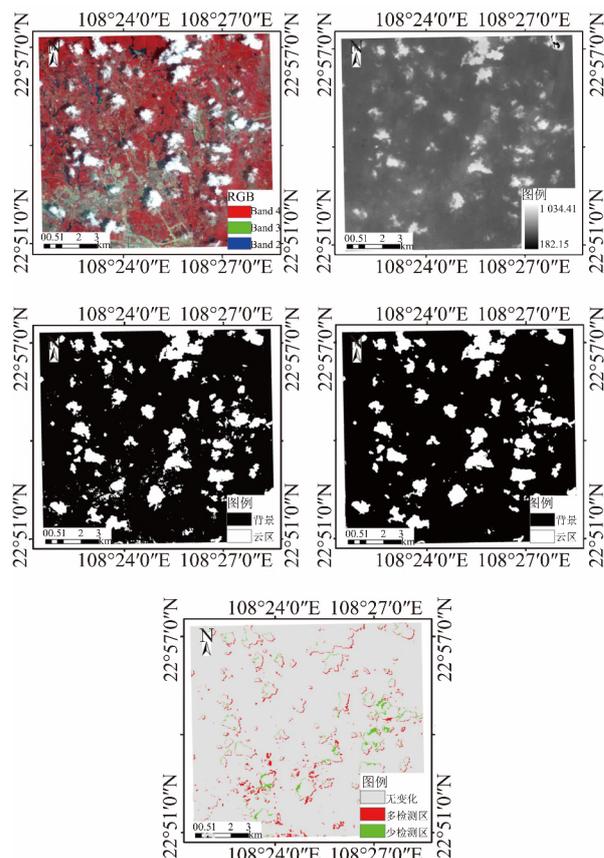


图 11 (a) GF-1 号影像,(b) HTM,(c) 蓝波段同态滤波结果,(d) HTM 同态滤波结果,(e) 蓝波段滤波结果与 HTM 滤波结果差值图

Fig. 11 (a) GF-1 remote sensing image,(b) HTM,(c) homomorphic filtering cloud detection results based on blue band and (d) the result based on HTM,(e) difference map between (c) and (d)

的结果.实例数据越丰富包含场景越广泛对模型的构建和泛化更有效.由于试验区遥感影像覆盖的主要地物类型为植被和人工建筑等,对于推广到其他范围场景如云覆盖区域下为高原、沙漠和海洋时,径向频谱与截止频率关系的模型系数应根据挑选的样本实例重新构建模型,模型的形式应与本研究结果类似,为截止频率正比于图像平均灰度能量占比.从信号学角度来看,场景内的云量覆盖程度直接影响了最优截止频率的位置,场景的平均灰度能量占比与截止频率呈现典型正相关关系.

在精度验证方面,研究中得到的是基于用户精度的精度验证结果,识别误差多产生于云区边界处.由于云层的分布具有一定厚度变化的特性,其厚度从中部到边界区域逐渐减小,多数情况下云区向晴空区过渡的地带并不明显,影像分辨率越高则呈现出更多薄云的细节信息.当云区边界像素点与地物

对比不明显时,算法识别的云区边界与实际目视情况略有出入,使得部分云块周围依然残留少许薄云无法被精准的识别。其次,为了优化云区掩膜,算法采用的形态学滤波处理也会使得云区边界产生变化。最后,对云区边界的人工解译会存在人为判定误差,即对边界的区域判定是否为云区存在主观误差。对于灰度值显著低于高亮地物的部分云区可能存在漏检情况,并且高亮地物的影响并不能彻底去除而造成过识别。因此文中定义的最优截止频率是平衡了正确识别云区范围和过度识别高亮地物的综合取值。同态滤波算法本身无法实现所有局部最优,但相较于多时相的云区识别算法,基于同态滤波算法的单景影像云识别算法还是具有相对便捷高效的处理结果。总体看来,本研究提出的云区识别流程能够准确地检测出 GF-1 遥感数据的云区范围,获得较为精准云区掩膜结果。

3 结论

高分辨率光学遥感影像的云区识别是数据后续应用过程中面临的一个重要问题。以同态滤波算法为主要技术手段,解决了同态滤波过程中滤波器对不同输入图像的适用性需求,采用广西壮族自治区国有高峰林场的 98 景 GF-1 高分辨率多光谱和宽幅多光谱数据为数据源,实现了快速自动地识别 GF-1 图像云区的目标,主要得出以下结论:

(1) 基于同态滤波的算法能对云区进行有效地识别。采用 HTM 代替通常采用的光学波段作为同态滤波输入,相比于单个光学波段,HTM 能更有效刻画云区分布并降低地物亮度。经过同态滤波算法,低频部分的云区被有效地抑制从而达到识别云区的目的。

(2) 影像的径向频谱为自适应确定截止频率提供了思路借鉴。不同输入影像对应不同的径向频谱曲线,云量较多时 $R(r)$ 曲线整体能量水平越高, $R[0]$ 值也越高,谐波能量收敛速度越快;无云覆盖时 $R(r)$ 曲线能量水平较低,谐波能量收敛趋近于 0 的速度越慢。通过建立影像径向频谱与滤波器截止频率之间的定量关系,有效地解决了批量云检测运算时经验截止频率无法适应复杂输入的问题,模型精度较高, R^2 达 0.9275。

(3) 后处理算法有利于优化云区掩膜结果。文中利用白度指数和数学形态学滤波结合的后处理策略对云区掩膜结果进行优化作用,能去除过度识别区域,降低高反射率地物造成的误识率,提高算法

精度。

(4) 提出的基于 GF-1 数据的云区识别算法精度较高,总体精度为 93.81%。通过对 98 景云掩膜数据进行人工目视解译,云掩膜内精度达 97.69%,云区掩膜缓冲区内精度为 90.15%。其中,误差多发生在云区向无云区过渡地带,云边界像元值与地物的对比度降低将会提高云判的不确定性。

(5) 算法运算时间复杂率较低,满足 GF-1 数据业务化批量处理需求。

本文提出的云区识别流程对 GF-1 数据展示出较好的效果。对于国产高分系列遥感数据中具有更高空间分辨率的 GF-2 数据而言,其更加高的空间分辨率所表现出的地物细节将对云区识别带来更多挑战。如何对 GF-2 进行快速有效的云区识别需要进行下一步的研究。

References

- [1] Zhang Y, Rossow W B, Lacis A A, et al. Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data [J]. *Journal of Geophysical Research: Atmospheres*, 2004, **109**(D19).
- [2] Asner G P. Cloud cover in Landsat observations of the Brazilian Amazon [J]. *International Journal of Remote Sensing*, 2001, **22**(18): 3855-3862.
- [3] PAN Teng, GUAN Hui, HE Wei. GF-1 satellite Remote Sensing Technology [J]. *Spacecraft Recovery and Remote Sensing* (潘腾,关晖,贺玮."高分二号"卫星遥感技术.航天返回与遥感), 2015, **36**(4): 16-24.
- [4] Ackerman S A, Strabala K I, Menzel W P, et al. Discriminating clear sky from clouds with MODIS [J]. *Journal of Geophysical Research: Atmospheres*, 1998, **103**(D24): 32141-32157.
- [5] Derrien M, Farki B, Harang L, et al. Automatic cloud detection applied to NOAA-11/AVHRR imagery [J]. *Remote Sensing of Environment*, 1993, **46**(3): 246-267.
- [6] Zhu Z, Woodcock C E. Object-based cloud and cloud shadow detection in Landsat imagery [J]. *Remote Sensing of Environment*, 2012, **118**: 83-94.
- [7] LIU Jian, LI Yu. Cloud phase detection algorithm for geostationary satellite data [J]. *Journal of Infrared and Millimeter Waves* (刘健,李云.风云二号静止气象卫星的云相态识别算法.红外与毫米波学报), 2011, **30**(4): 322-327.
- [8] LI Wei, FANG Sheng-Hui, DIAN Yuan-Yong, et al. Cloud Detection in MODIS Data Based on Spectrum Analysis [J]. *Geomatics and Information Science of Wuhan University* (李微,方圣辉,佃袁勇,等.基于光谱分析的 MODIS 云检测算法研究.武汉大学学报:信息科学版), 2005, **30**(5): 435-438.
- [9] MA Fang, ZHANG Qiang, GUO Ni, et al. The Study of Cloud Detection with Multi-Channel Data of Satellite [J]. *Chinese Journal of Atmospheric Sciences* (马芳,张强,郭妮,等.多通道卫星云图云检测方法的研究.大气研

- 究), 2007, **31**(1): 119-128.
- [10] Han Y, Kim B, Kim Y, *et al.* Automatic cloud detection for high spatial resolution multi-temporal images [J]. *Remote sensing letters*, 2014, **5**(7): 601-608.
- [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): 1079-1089.
- [12] Tseng D C, Tseng H T, Chien C L. Automatic cloud removal from multi-temporal SPOT images [J]. *Applied Mathematics and Computation*, 2008, **205**(2): 584-600.
- [13] Lu D. Detection and substitution of clouds/hazes and their cast shadows on IKONOS images [J]. *International Journal of Remote Sensing*, 2007, **28**(18): 4027-4035.
- [14] Marais I V Z, Du Preez J A, Steyn W H. An optimal image transform for threshold-based cloud detection using heteroscedastic discriminant analysis [J]. *International Journal of Remote Sensing*, 2011, **32**(6): 1713-1729.
- [15] Fisher A. Cloud and cloud-shadow detection in SPOT5 HRG imagery with automated morphological feature extraction [J]. *Remote Sensing*, 2014, **6**(1): 776-800.
- [16] Bai T, Li D, Sun K, *et al.* Cloud detection for high-resolution satellite imagery using machine learning and multi-feature fusion [J]. *Remote Sensing*, 2016, **8**(9): 715.
- [17] LI Hong-li, SHEN Huan-feng, DU Bo, *et al.* A High-fidelity Method of Removing Thin Cloud from Remote Sensing Digital Images Based on Homomorphic Filtering [J]. *Remote Sensing Information* (李洪利, 沈焕锋, 杜博, 等. 一种高保真同态滤波遥感影像薄云去除方法. 遥感信息), 2011, **1**: 41-44.
- [18] Shen H, Li H, Qian Y, *et al.* An effective thin cloud removal procedure for visible remote sensing images [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, **96**: 224-235.
- [19] Jiang H, Lu N, Yao L. A High-Fidelity Haze Removal Method Based on HOT for Visible Remote Sensing Images [J]. *Remote Sensing*, 2016, **8**(10): 844.
- [20] Lv H, Wang Y, Shen Y. An empirical and radiative transfer model based algorithm to remove thin clouds in visible bands [J]. *Remote Sensing of Environment*, 2016, **179**: 183-195.
- [21] Chavez P S. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data [J]. *Remote sensing of environment*, 1988, **24**(3): 459-479.
- [22] Chavez P S. Image-based atmospheric corrections-revisited and improved [J]. *Photogrammetric engineering and remote sensing*, 1996, **62**(9): 1025-1035.
- [23] Makarau A, Richter R, Muller R, *et al.* Haze detection and removal in remotely sensed multispectral imagery [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, **52**(9): 5895-5905.
- [24] Liu Q, Gao X, He L, *et al.* Haze removal for a single visible remote sensing image [J]. *Signal Processing*, 2017, **137**: 33-43.
- [25] Liu Z K, Hunt B R. A new approach to removing cloud cover from satellite imagery [J]. *Computer vision, graphics, and image processing*, 1984, **25**(2): 252-256.
- [26] ZHAO Zhong-Ming, ZHU Chong-Guang. Approach to removing cloud from satellite imagery [J]. *Journal of Remote Sensing* (赵忠明, 朱重光. 遥感图象中薄云的去方法. 环境遥感), 1996, **11**(3): 195-199.
- [27] Adelman H G. Butterworth equations for homomorphic filtering of images [J]. *Computers in Biology and Medicine*, 1998, **28**(2): 169-181.
- [28] Gómez-Chova L, Camps-Valls G, Calpe-Maravilla J, *et al.* Cloud-screening algorithm for ENVISAT/MERIS multispectral images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2007, **45**(12): 4105-4118.
- [29] ZHOU Lin-Na, WANG Dong-Ming, GUO Yun-Biao, *et al.* Exposing Digital Forgeries by Detecting Image Blurred Mathematical Morphology Edge [J]. *Chinese Journal of Electronics* (周琳娜, 王东明, 郭云彪, 等. 基于数字图像边缘特性的形态学滤波取证技术. 电子学报), 2008, **6**.