

一种适合高光谱卫星云识别的 Fmask 改进算法

张舒宁^{1,2,3}, 张浩^{4*}, 张兵^{1,3}, 崔珍珍^{4,5}, 肖晨超⁶¹中国科学院空天信息创新研究院数字地球重点实验室, 北京 100094;²可持续发展大数据国际研究中心, 北京 100094;³中国科学院大学资源与环境学院, 北京 100049;⁴中国科学院空天信息创新研究院航空遥感中心, 北京 100094;⁵河南理工大学测绘与国土信息工程学院, 河南 焦作 454003;⁶自然资源部国土卫星遥感应用中心, 北京 100048

摘要 近年来,我国卫星高光谱技术发展迅猛,高分五号、高分五号 02 星、资源一号 02D 星、资源一号 02E 星等相继发射为遥感领域带来了丰富的高光谱数据源。但高光谱卫星在成像过程中不可避免地会受到云及云阴影的影响,如何准确识别成为保障后续应用的关键,Fmask 算法作为国内外诸多算法中的典型代表,已被 Landsat 和 Sentinel 业务化产品生产系统采用。Fmask 算法作为国内外诸多算法中的典型代表,已被 Landsat 和 Sentinel 业务化产品生产系统采用。但该算法对于缺少热红外波段的数据精度偏低,例如对 Sentinel-2 数据的云和云阴影识别精度分别为 84.5% 和 50% 左右。鉴于此,本文通过在原有算法中优化云及云阴影识别算法结构、增加高亮地物识别辅助判据等改进手段,提出了一种适合高光谱卫星的 Fmask 改进算法,并在含有城区、山地、平原等三类不同下垫面场景的 20 景高分五号和资源一号高光谱影像中进行检验,结果表明:云识别的用户精度和生产者精度可达 91.26% 和 99.97%,云阴影识别精度达到 78.66% 和 79.41%,明显优于原始算法。本文算法对于高光谱数据的云及云阴影识别具有精度高、效果稳定和易于工程化实现的特点,可用于支撑国产高光谱卫星数据的业务化处理。

关键词 高光谱遥感; 云识别; 云阴影识别; Fmask; 高分五号; 资源一号

中图分类号 TP751.2

文献标志码 A

DOI: 10.3788/AOS230563

1 引言

云作为大气中水循环的重要部分,在地球的辐射平衡中起着重要作用^[1]。然而,云会对遥感影像中地物的反射产生衰减和遮蔽效果,尤其是在热带地区,云对于遥感数据的影响几乎是不可避免的^[2]。因此,为了更有效地提取地表信息,云和云阴影的准确识别变得尤为重要。

基于云的高反射率、低温度特征和云阴影的低反射率特征,诸多学者针对 Landsat^[3-4]、SPOT^[5] 和 MODIS 等^[6-7] 多光谱卫星开发了诸多云识别算法。其中最常用的是阈值法,方法假设在云区和非云区、云阴影区和非云阴影区得到的卫星观测数据(反射率、亮温或者其他基于基本物理量构造的识别指数)会落入不同的值域空间^[8],并以此进行区分。经典算法有 International Satellite Cloud Climatology Project (ISCCP) 算法、The NOAA Cloud Advanced Very High Resolution Radiometer (CLAVR) 算法和 Fmask

算法等。ISCCP 算法^[9-11] 利用统计手段生成晴空背景场,若像元的实际观测辐射值超出晴空辐射值的最大变化(与季节、气候相关),则确定为云像元。CLAVR 算法^[12] 基于 Advanced Very High Resolution Radiometer (AVHRR) 传感器的 5 个波段,采用决策树的方式将区域图像分类为无云、部分有云和全云三种类型^[13-15]。Fmask 算法最初是由 Zhu 等^[16] 针对 Landsat7 提出的云及云阴影检测算法,以大气顶层反射率和亮度温度作为输入,利用云及云阴影的光谱特征以及二者形状的相似性进行识别,识别精度可达 96.41% 和 70%。阈值法作为传统的云识别方法,原理较为简单,易于实现,广泛应用于多种光学图像。但该类方法仅仅依靠云及云阴影的光谱特性,适用性较差,常常会遗漏一些特征较弱的薄云,或是出现将雪等地物识别为云等错判的情况。此外,机器学习方法近年来也作为一种较普遍的方法被引入多光谱数据的云识别中。该方法利用人工神经网络作为分类器来训练样本数据,提取出影像中云和云阴影的光谱和空间特

收稿日期: 2023-02-16; 修回日期: 2023-03-29; 录用日期: 2023-04-24; 网络首发日期: 2023-05-08

基金项目: 国家自然科学基金(41771397)、海南省重大科技计划项目(ZDKJ2019006)

通信作者: *zhanghao612@radi.ac.cn

征,以达到识别目的。张晨等^[17]、王海涛等^[18]均利用深度学习方法,通过提出新的网络结构,得到了精度较高的云识别结果。机器学习法对噪声的鲁棒性和较强的容错性使得其能够获得较好的结果,但机器学习方法也有一定的缺陷,如数据量大、计算量大、普适性差、过拟合等,对于业务化应用具有一定挑战。

近年来,随着高光谱技术的发展,我国有越来越多的高光谱卫星投入使用,如高分五号(GF-5)、资源一号02D(ZY-1 02D)等。不同于多光谱影像,高光谱影像的波段多、光谱分辨率高,能够得到近乎连续的地物光谱曲线,使得高光谱数据具有图谱合一的特性。精细的光谱分辨率能够反映地物光谱的细微特征,对目标识别、地物精细化分类、作物估产、矿产资源探测等具有重大意义^[19]。高光谱连续细分谱段带来地物连续光谱曲线的同时也带来噪声相对较大、波段间的高相关性和信息冗余度增加的问题^[20]。因此,与多光谱数据相比,高光谱数据的云识别更具有挑战性。目前,针对高光谱数据的云识别方法主要集中在机器学习和利用辅助数据的方法上,任鹏洲等^[21]采用最佳波段指数拟合方法筛选出Hyperion高光谱影像的最佳波段组合,利用随机森林算法进行云和云阴影识别。王蓝星等^[22]利用辅助Landsat影像去除厚云,在GF-5和Hyperion卫星上获得了较好结果,但这些方法需要的数据量大、操作复杂、计算时间长,不适合业务化运行。此外,我国虽然发射了多颗高光谱卫星,但针对这些卫星的系统性云识别算法却并不多,我们迫切需要一种适合国产高光谱卫星的快速、高精度、稳定的云识别算法,而针对多光谱的阈值算法经过不断实验和验证,符合我们对于算法的预期。其中,经典的系统性阈值云识别算法是Fmask算法。

Fmask算法经过不断的改进和功能扩充,已成为一种成熟的云及云阴影识别业务化算法,广泛应用于Landsat系列卫星(Landsat4~8)及哨兵系列卫星中^[16,23-24]。其中,Fmask4.0由于其辅助数据的增加、归一化卷云波段的应用以及新的云概率阈值而备受关注。对于云识别,首先通过多项阈值判据判断出潜在的云像元,随后基于图像阈值从云概率层提取云像元。在识别过程中,热红外波段和卷云波段的使用大大提升了云检测的正确率,热红外波段带来了地物的亮温信息,减少了明亮地物错分为云的情况,对于区别雪/冰和云、城市和云效果显著,而卷云波段可以识别出更高、更薄的卷云,二者都使得云的识别更加完整^[25]。对于云阴影,利用洪水填充筛选出被亮区包围的阴影区域,通过云及云阴影的相似性特征进行匹配。Fmask4.0还能够进行水体、雪/冰、建筑区识别。相比于Fmask3.3,Fmask4.0的识别精度得到了总体提升,对于landsat4-7和Landsat8,总体准确率分别提升了1.67%和1.29%。算法对于Landsat4~7的云识别的平均生产者精度和用户精度分别可达88.04%和

99.06%,云阴影识别的精度也达到了72.58%和50.56%;对于Landsat8,算法的云识别精度超过了95%,云阴影精度最高可达70%^[25]。大量应用及验证表明,Fmask算法具有精度较高、性能稳定、算法步骤简单、运行速度快等优势。但目前研究发现,Fmask算法对于没有热红外波段的数据存在云及云阴影识别精度偏低、水体与云阴影识别混淆、地形阴影识别率低等问题,如Fmask4.0算法对Sentinel-2影像的云分类精度最高为84.5%、云阴影的用户识别精度和生产者精度仅有50.8%和42.8%,与Landsat系列卫星的识别精度有相当大的差距^[26]。

由此,本文结合高光谱数据特点对Fmask4.0算法进行改进,重点解决无热红外波段的高光谱数据带来的光谱性识别问题,从而提高算法的稳定性及适应性,进而实现高精度的云及云阴影识别,满足国产高光谱卫星影像的业务化处理需求。

2 算法原理

利用云和云阴影的光谱特性以及二者间的相似性关系,Fmask算法构建了多个表观反射率及亮温的阈值测试,但对于没有热红外波段的数据识别精度却十分有限。为此,本文引入城区掩模判据,增设云阴影识别条件,主要从云及云阴影两方面对算法进行改进,使其适用于国产高光谱卫星数据。在应用算法时,通常需要利用对比验证的方法检验算法的精度及鲁棒性等,对此,本文将采用精度验证指标的方法,通过比较改进算法与验证算法的识别精度指标大小来探究改进Fmask算法的识别效果。

2.1 Fmask改进算法

基于Landsat8卫星的Fmask4.0算法采用潜在云区检测确定可能的云像元,包括基础测试、白度测试^[16]、HOT^[23]测试以及波段5、6测试。潜在云区测试得到的像元可能是多云像元,也可能是晴空像元。在得到潜在云像元后,再利用亮温和表观反射率数据计算像元为云的概率,若概率大于阈值则确定为云像元。

影像中的云阴影主要由散射光照亮,大气散射在较短波长处(如可见波段)更强,在长波长处(如SWIR波段)相对较小,使得阴影像元比周围环境更暗^[23]。基于该原理,Fmask算法采用洪水填充的方法提取被亮区包围的暗区,得到潜在云阴影像元,进而运用相似性原理将云和潜在云阴影像元以缓冲3个像元的8连接方式进行匹配,匹配成功后得到云阴影识别结果。

2.1.1 预处理改进部分

首先,在计算表观反射率时要使用国产高光谱数据对应的计算公式,

$$\rho = \frac{\pi \times L \times D^2}{E_{\text{SUN}} \times \cos \theta}, \quad (1)$$

式中: L 为辐射亮度,利用卫星提供的增益和偏置参数对原始DN值计算得到; D 为由成像时间得到的日地

距离; E_{SUN} 为大气层顶的平均太阳辐照度; θ 为太阳天顶角。

引言部分提到, 高光谱数据波段较窄, 噪声影响较大。为了降低单波段噪声对算法的影响, 本文对多个高光谱波段的表现反射率利用 Landsat8 的光谱响应曲线进行加权拟合, 使其合并为对应于 Landsat8 卫星的一个宽波段数据。计算公式为

$$X_i = \frac{\int_{\lambda_1}^{\lambda_2} g_i(\lambda) X(\lambda) d\lambda}{\int_{\lambda_1}^{\lambda_2} X(\lambda) d\lambda}, \quad (2)$$

式中: $g_i(\lambda)$ 是 Landsat8 多光谱传感器各波段的光谱响应曲线; $X(\lambda)$ 是高光谱卫星数据的表现反射率影像; λ_1 、 λ_2 分别为该波段的起止波长。加权拟合后的表现反射率可以明显降低噪声对影像质量的影响。

2.1.2 云识别改进部分

高分五号和资源一号 02D 卫星的可见短波红外高光谱相机波段范围为 400~2500 nm, 无法捕捉地面的热发射特性, 因此我们将原始算法中使用亮温数据的部分去掉, 这样的处理也为云识别带来了问题。由于缺少亮温相关的约束, 影像中部分较亮的城区会被误判为云。为此, 本文利用归一化建筑指数 (NDBI) 和归一化植被指数 (NDVI) 以实现城区像元的掩模, 减少误判带来的误差。其中, NDBI 能够利用地物在短波红外和近红外波段反射的差异识别建筑物, 但这一差异不如中红外和近红外波段之间的差异明显, 因此仅仅依靠 NDBI 检测城区可靠性不高^[27], 常常会包含部分植被。因此对于高光谱数据, 本文选用 NDBI 与 NDVI 的差值来进行建筑物识别, 为了验证其有效性, 本文选用 5 景包含城区的影像进行分析, 将 NDBI 与 NDVI 的相关系数及二者的差与 NDVI 的相关系数进行对比, 得到的结果如图 1 所示。图中, 横坐标为影像编号, 纵坐标为相关系数, 黑色实线代表 NDBI 与 NDVI 间的相关系数, 红色虚线代表 NDBI-NDVI 与 NDVI 的相关系数, 折线上的每个点都代表了一景影像的计算结果, 具体数值标在点上方。从图 1 可以看出, 相比于不做处理的 NDBI, 将 NDBI 与 NDVI 相减能够在一定程度上降低与 NDVI 的相关性, 即 NDBI-NDVI 检测出的城区与植被的相关性更小, 更可信。

由此, 本文结合了 NDBI 和 NDVI 两种指数, 利用二者的差值来提取城区高亮地表, 表示为

$$N_{\text{DBI}} - N_{\text{DVI}} > T, \quad (3)$$

$$N_{\text{DVI}} = \frac{\rho_{\text{NIR}} - \rho_{\text{RED}}}{\rho_{\text{NIR}} + \rho_{\text{RED}}}, \quad (4)$$

$$N_{\text{DBI}} = \frac{\rho_{\text{SWIR}} - \rho_{\text{NIR}}}{\rho_{\text{SWIR}} + \rho_{\text{NIR}}}, \quad (5)$$

式中: ρ_{RED} 、 ρ_{NIR} 、 ρ_{SWIR} 分别代表红波段 (0.65 μm 左右)、近红外波段 (0.85 μm 左右) 和短波红外波段 (1.6 μm 左右) 表现反射率。

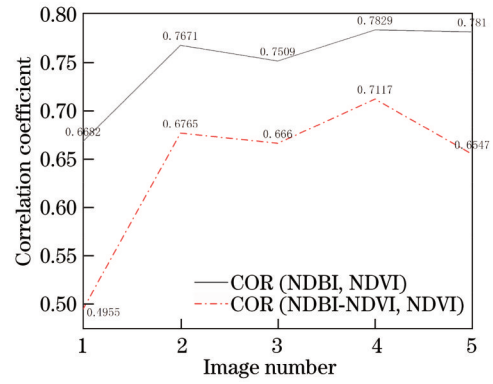


图 1 相关系数比较图

Fig. 1 Correlation coefficient comparison chart

具体而言, 本文运用式 (3) 对城区进行掩模, 使得公式成立的像元被认为是城区像元, 将从潜在云像元中剔除。剔除像元的数量将由阈值 T 决定。值得注意的是, 对于城区误判的情况, Fmask4.0 算法能够利用亮温数据来排除这些云像元。然而对于光谱覆盖范围较小、没有热红外波段的高光谱数据, 仅靠式 (3) 筛选出的像元中可能会存在真正的云像元, 因此在选择阈值 T 的过程中要优先保证云像元正确率, 本文将云像元正确率的期望值定为 80%。此外, 由于城区是影响云识别的主要因素, 因而认为算法识别出的云由正确的云像元及错分为云的城区高亮像元组成。基于这一前提, 构建了阈值确定实验的参量, 如图 2 所示。

定义如图 2 所示的 6 个变量, 分别代表了使用城区阈值判据前后识别出的云像元数量 A_1 和 A_2 、正确识别出的云像元数量 C_1 和 C_2 以及城区像元数量 U_1 和 U_2 。以此为基础, 本文创新性地提出了城区掩模率 ($P_{\text{CT,UM}}$) 及云识别正确率 ($P_{\text{CT,CD}}$) 用于阈值 T 的确定, 如式 (6)、(7) 所示。其中: $P_{\text{CT,UM}}$ 代表了城区错分情况的改善程度, 值越大表示加入判据后错分情况越少; $P_{\text{CT,CD}}$ 反映了正确识别的云像元占有云像元的比例, 值越大代表正确识别的云像元越多。

$$P_{\text{CT,UM}} = \frac{U_1 - U_2}{U_1}, \quad (6)$$

$$P_{\text{CT,CD}} = \frac{C_2}{C_1}. \quad (7)$$

图 3 以一景资源一号 02D 卫星的城区影像为例, 展示了阈值调节时两个参量的变化。其中, 横坐标为阈值 T , 纵坐标代表比例, 方点和三角点分别代表 11 组不同的阈值计算出的城区掩模率和云像元正确率, 黑、红线分别是对参量值的一元二次拟合曲线。由图 3 可知, 云识别正确率和城区掩模率分别呈单调递增和递减的趋势。为保证云识别正确率不小于 80% 且在最大程度上减少误判, 本文选择图中圆点处对应的阈值, 大约为 -0.25。后经过多景影像实验表明, $T = -0.25$ 能在最大程度上保证云像元正确率的同时, 有效对高亮城区进行掩模。

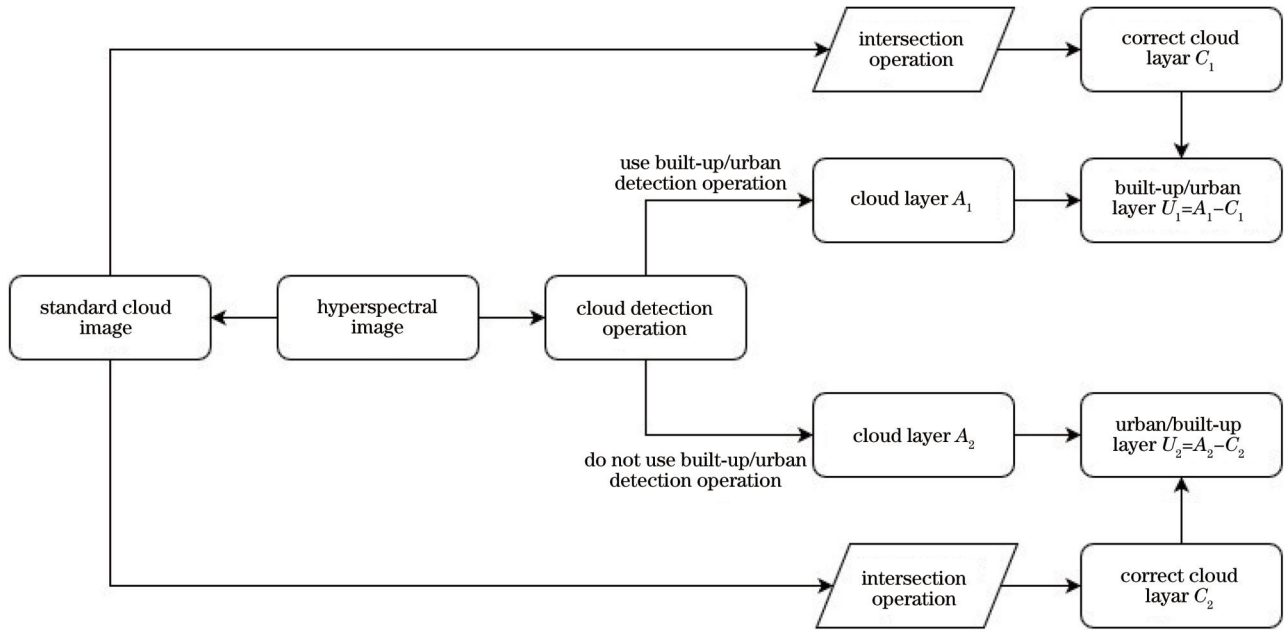


图 2 城区掩模阈值确定流程

Fig. 2 Determination process of urban mask threshold

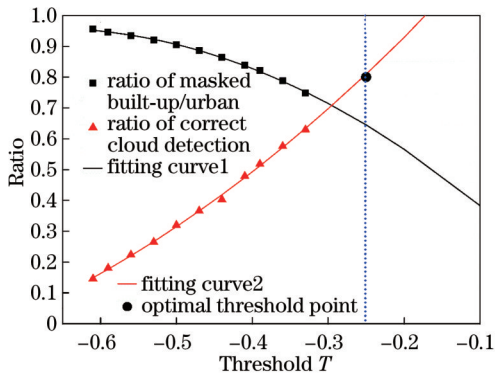


图 3 城区掩模阈值变化图

Fig. 3 Threshold change diagram of urban mask

2.1.1.3 云阴影识别改进部分

Fmask算法在对国产高光谱数据进行云阴影识别时,通常存在以下三个问题:

1) 相似性匹配是得到准确云阴影结果的关键,但对应云在影像外的云阴影无法进行匹配,使得这部分真实的云阴影被标记为地面,进而对结果造成一定

影响;

2) 高光谱数据光谱覆盖较窄,未使用亮温数据会导致部分云阴影误判为水体,影响后续应用;

3) 对于地表起伏较大的地区影像,如丘陵地带,与云阴影形成原理类似的地形阴影会极大地干扰云阴影的识别。

因此,本文在改进算法云阴影识别部分的同时,也考虑到地形的影响,尽可能地减少地形阴影带来的误差。改进算法的云阴影识别流程图如图 4 所示,图中红色部分为改进部分,黑色部分为基于 Landsat8 卫星的 Fmask 算法。

针对前两个问题,本文引入 Atmospheric Correction and Haze Reduction for ERDAS IMAGINE from GEOSYSTEMS(ATCOR)3 大气校正模型^[28]中的云阴影判据,用于在相似性匹配后识别遗漏的云阴影。

$$[0.04 < \rho_{NIR} < 0.12] \cup [\rho_{SWIR1} < 0.20], \quad (8)$$

式中: ρ_{SWIR1} 为短波红外波段(1.57~1.65 μm) 的表现

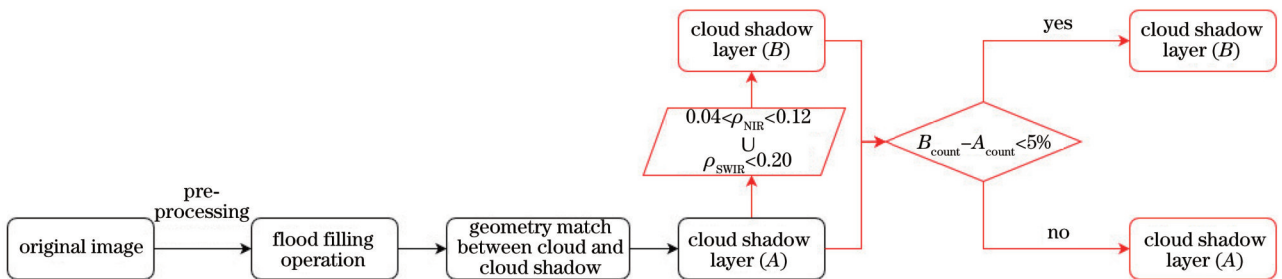


图 4 改进 Fmask 算法云阴影识别流程图

Fig. 4 Flow chart of cloud shadow recognition in modified Fmask algorithm

反射率值。为确保云阴影与水体分离,在执行判据前需利用归一化水体指数 (NDWI) 进行水体识别。若 NDWI 小于 0 则为非水区域,可执行阴影判据。

$$N_{DWI} = \frac{\rho(\text{GREEN}) - \rho(\text{NIR})}{\rho(\text{GREEN}) + \rho(\text{NIR})}, \quad (9)$$

式中: ρ_{GREEN} 代表绿波段的表观反射率。

此外,为了减少地形阴影带来的误差,本文在水体识别后加入变化测试,用于判断改进前后算法识别出的云阴影面积 N_{before} 与 N_{after} 之差占影像总像元数的比

例是否大于 P ,若大于则不执行阴影判据。本文默认算法遗漏的云阴影像元数远小于地形阴影像元数,因此相比于少部分阴影遗漏或被识别为水体的小误差,避免地形阴影误判而导致的巨大误差更为重要。本文中使用的 $P=5\%$ 。

$$\frac{N_{\text{after}} - N_{\text{before}}}{X_{\text{row}} \times X_{\text{column}}} < P. \quad (10)$$

经过云和云阴影两方面的改进,改进 Fmask 算法流程图如图 5 所示。

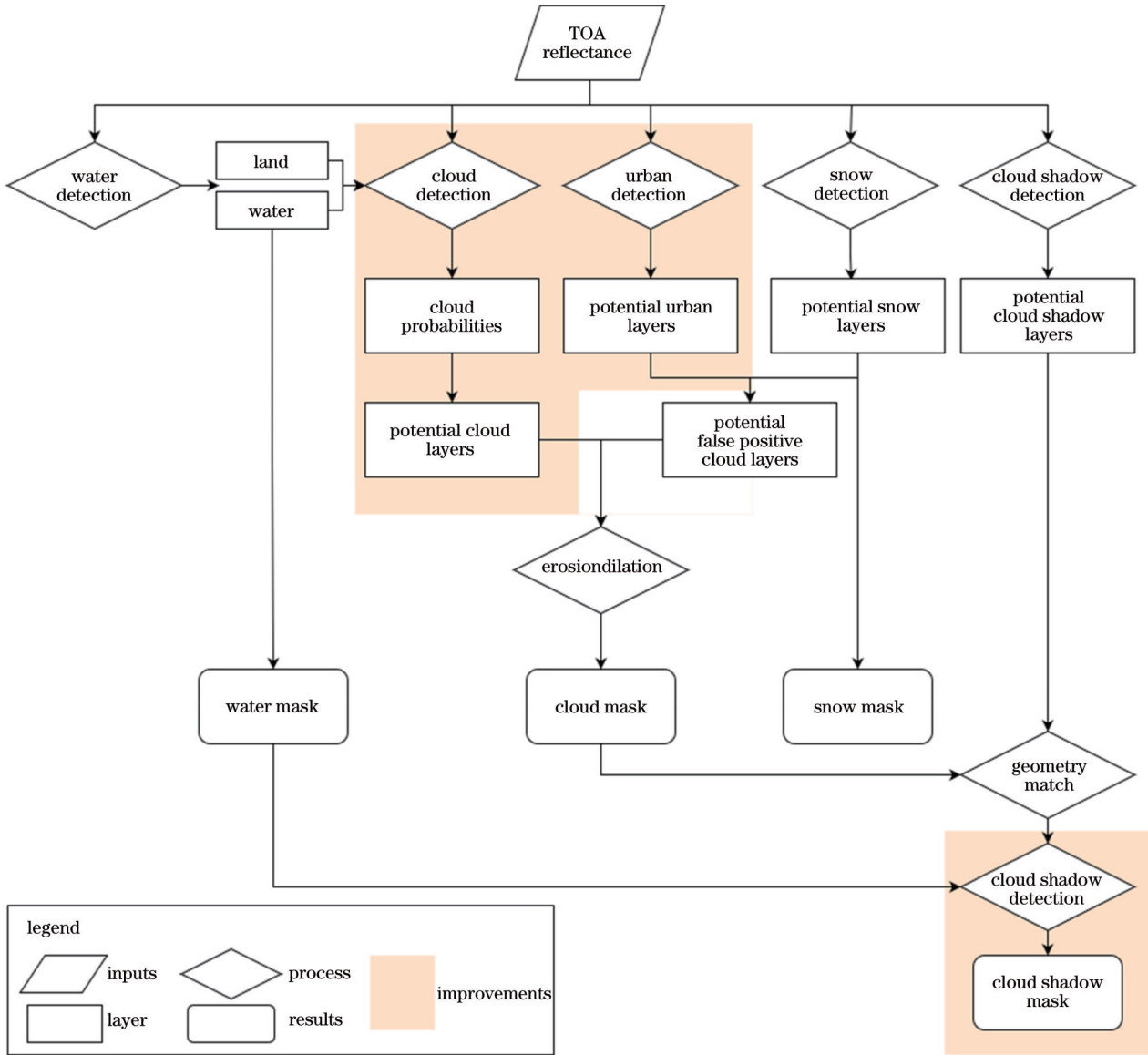


图 5 改进 Fmask 算法流程图
Fig. 5 Flow chart of modified Fmask algorithm

2.2 对比算法

本节描述的对比较算法分为两类:1)云识别算法,将与改进 Fmask 算法的云识别结果进行对比验证,包括原始 Fmask 算法、ATCOR 判据以及 HCC 算法;2)云阴影算法,用来验证云阴影识别结果,包括原始 Fmask 算法、ATCOR 判据。

2.2.1 ATCOR 软件判据

ATCOR 软件能够对影像进行大气校正,在最大程度上去除薄云及雾霾。本文利用了 ATCOR3 中的云及云阴影的识别判据与改进 Fmask 算法结果进行对比验证,分别如下式和式(8)所示。

$$\begin{aligned} & [T_{c1} < \rho_{\text{BLUE}} \leq T_{c2}] \cup [S_{c1} < \rho_{\text{RED}} \leq S_{c2}] \\ & \cup [\rho_{\text{NIR}} \leq 0.8\rho_{\text{RED}}] \\ & \cup [\rho_{\text{NIR}}/\rho_{\text{SWIR1}} > 1] \\ & \cup \{ [N_{\text{DSI}} < 0.7] \cup [DN_{\text{BLUE}} \geq T_{\text{saturation}}] \}, \quad (11) \end{aligned}$$

式中： ρ_{BLUE} 、 ρ_{RED} 分别为蓝、红波段的表观反射率； DN_{BLUE} 为蓝波段的 DN 值。为了探测到最大范围的云，本文中使用的参数为 $T_{c1}=0.15$ ， $T_{c2}=+\infty$ ， $S_{c1}=0.15$ ， $S_{c2}=+\infty$ ， $T_{\text{saturation}}$ 代表饱和像元值，与量化等级有关，一般为 0.95 倍的影像最大像元值，如高分五号量化等级为 12， $T_{\text{saturation}}$ 取值为 3890.25。

2.2.2 针对 Hyperion 的云识别方法

针对 Hyperion 高光谱数据的云识别 (HCC) 算法^[29]，使用 Hyperion 的 6 个波段 (0.55 μm 、0.66 μm 、0.86 μm 、1.25 μm 、1.38 μm 和 1.65 μm) 来识别地物，通过多步阈值判据区分高云、低云、植被、冰雪和沙漠等类型。本文利用该算法的云识别结果对改进 Fmask 算法的精度进行验证。

2.3 精度方法

论文采用如下检验指标评价算法性能：总体精度 (OA)、生产者精度 (PA) 和用户精度 (UA)。这三种精度指标从混淆矩阵简化而来，包含混淆矩阵的绝大部分信息，在云及云阴影识别算法实验中能充分描述结果的准确与精确程度。其中：总体精度指所有正确分类的像元数占总像元数的百分比^[30]；生产者精度表示在此次分类中，该类别的地面真实参考数据被正确分类的概率，反映了漏分情况；用户精度表示分类图中被正确分类为该类别的比率，对像元误判情况有重要帮助^[31]。精度指标的公式^[16,23-24]为

$$P_{\text{OA}} = \frac{\alpha}{\beta}, \quad (12)$$

$$P_{\text{PA}} = \frac{\alpha}{\eta}, \quad (13)$$

$$P_{\text{UA}} = \frac{\alpha}{\delta}. \quad (14)$$

以云检测为例， α 为一种算法相对于标准云图正确分类的像元数， β 为影像全部像元数， η 为标准云图中的云像元数， δ 为算法分类图中的云像元数。在精度验证环节，利用 ENVI5.3 分别勾绘了 20 景影像的标准云图和标准云阴影图，对云和云阴影分别使用了两种验证算法、三个精度指标进行验证。

3 实验数据

本文采用高分五号卫星和资源一号 02D 卫星的高光谱影像数据对改进 Fmask 算法进行验证和评估。

高分五号卫星于 2018 年 5 月 9 日在太原卫星发射中心成功发射，所搭载的可见短波红外高光谱相机 (AHSI) 是国际上首台同时兼顾宽覆盖和宽波段的高光谱相机，空间分辨率为 30 m，幅宽为 60 km，光谱范

围 400~2500 nm，可见光-近红外与短波红外分别有 150 和 180 个波段，光谱分辨率为 5 nm 和 10 nm。与 Hyperion 卫星相比，AHSI 的信噪比更高、幅宽更宽、谱段数更多^[32]，优秀的载荷使得高分五号卫星能够检测到地物近乎连续的光谱曲线，对地物识别及相关研究有重要意义。2020 年 4 月后，高分五号卫星因故障无法使用，作为替代的高分五号 02 星于 2021 年 9 月 7 日发射升空。

2019 年 9 月 12 日，资源一号 02D 卫星在太原卫星发射中心发射升空，其搭载的可见短波红外高光谱相机的幅宽、空间分辨率和光谱范围与高分五号卫星一致，但仅有 166 个波段，分别为 10 nm 光谱分辨率的 76 个可见光-近红外波段和 20 nm 光谱分辨率的 90 个短波红外波段。资源一号 02D 卫星同样可以带来近乎连续的地物光谱曲线，可满足新时代的国土监测和调查等任务需求。

本文所用影像的时间信息如表 1 所示。从空间分布来看，实验影像大部分位于中国云南省和新疆维吾尔自治区，一景位于山东省。影像的地貌类型有平原、丘陵和山地。影像的地物类型主要有陆地、水体、雪区、云区和云阴影。

表 1 实验影像年份表

Table 1 Experimental image year table

Item	2018	2019	2020	2021	All
GF-5	1	9	3	0	13
ZY-1 02D	0	2	2	3	7
All	1	11	5	3	20

4 结果与分析

4.1 定性评估

对 20 景高光谱影像进行了改进 Fmask 云识别算法的实验，得到的部分实验结果如图 6 所示。其中，红色代表云，绿色代表云阴影，蓝色代表水体，黄色代表雪区，灰色代表陆地。

从展示的 10 景影像及识别结果初步来看，改进 Fmask 算法在总体上识别精度较高，符合目视结果。

基于云的高反射率和光谱平缓等特征，改进算法在云识别方面表现稳定，能够有效识别不同厚度的云，如图 6(c) 和图 6(g) 中厚云和薄云的掩模结果都与目视相符。此外，算法对于云的边界有一定的敏感性，即使是薄云边界，其识别结果也符合预期 [图 6(g)]。此外，在云和雪同时存在的影像 [图 6(h)] 中，根据目视结果能够大致判断算法识别出的云雪边界是接近真实状况的。

在云阴影识别方面，对于如图 6(a)、(c) 这类下垫面为平原的影像，算法识别出的绝大部分云阴影均符合目视结果，云和云阴影的相对位置也与实际情况相符。通常情况下，对于下垫面为山地丘陵的影像 [图 6

(d)、(e)、(f)、(i)、(j)],地形阴影的存在会严重干扰云阴影的识别准确度,然而从几景影像的识别结果来看,改进 Fmask 算法能够在很大程度上减少山区阴影带来的误差,云阴影识别结果总体上符合目视结果。

据此,改进 Fmask 算法能够有效识别薄厚不同的云,识别效果和人工目视判断大致相符,算法也能适应不同的下垫面类型,得到较为准确的云阴影结果,如图

6(a)中的平原,图 6(d)、(f)的山地等,有较好的普适性。

此外在识别出云和云阴影的前提下,改进 Fmask 算法对于水体和雪区的识别也比较符合实际情况,且二者对云及云阴影识别的影响并不大,对于水体面积较大[图 6(b)]、云与雪相接[图 6(h)]的影像,识别结果均较为符合目视判别。

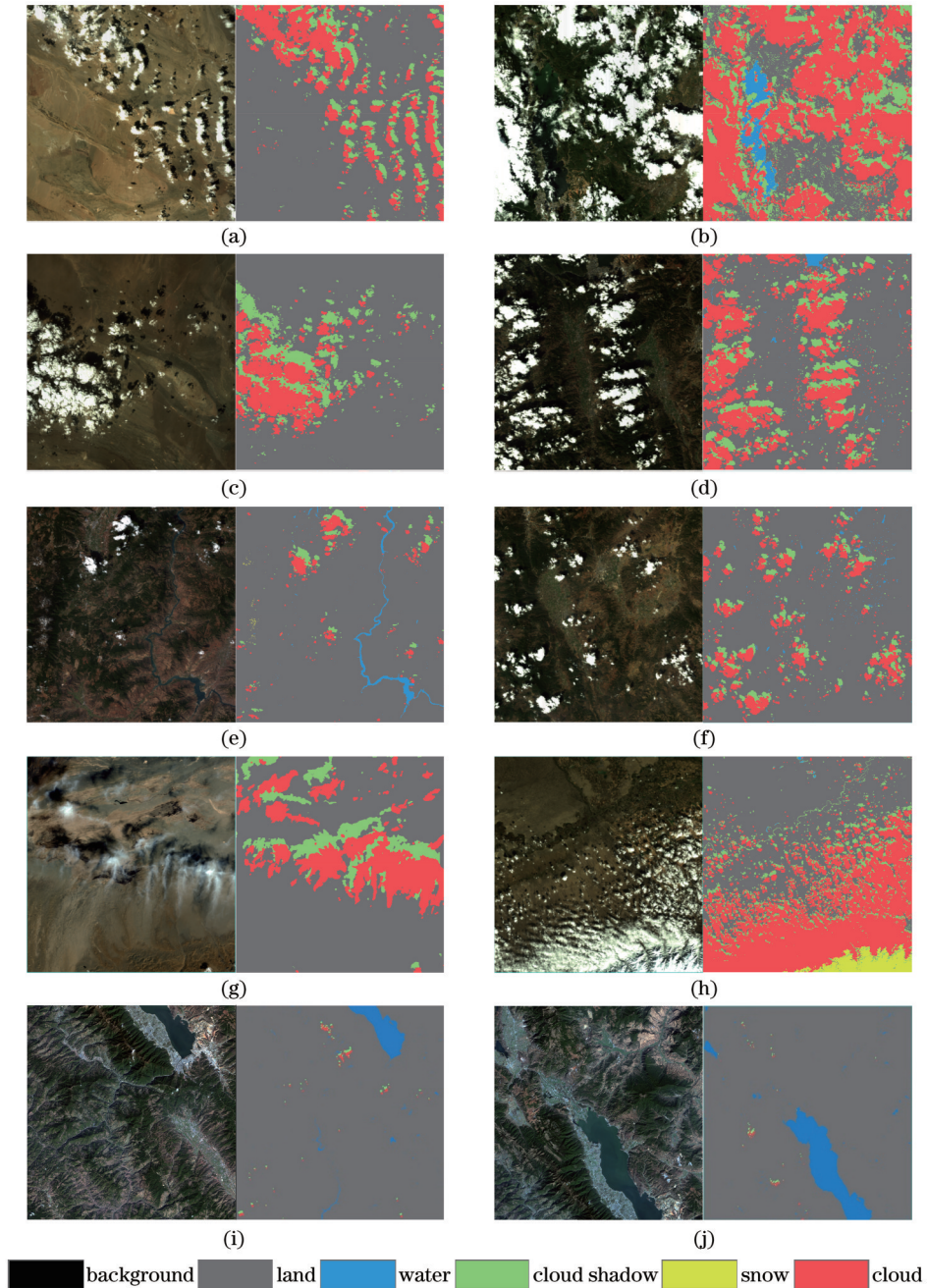


图 6 10 景高光谱卫星原始影像及对应改进 Fmask 算法识别结果。(a)GF5-20190408;(b)GF5-20191003;(c)GF5-20191015;(d)GF5-20200120;(e)GF5-20200311;(f)GF5-20200311(2);(g)ZY02D-20210405;(h)ZY02D-20211003;(i)ZY02D-20191215;(j)ZY02D-20191215(2)

Fig. 6 Recognition results of 10-scene hyperspectral satellite original image and corresponding improved Fmask algorithm. (a) GF5-20190408; (b) GF5-20191003; (c) GF5-20191015; (d) GF5-20200120; (e) GF5-20200311; (f) GF5-20200311(2); (g) ZY02D-20210405; (h) ZY02D-20211003; (i) ZY02D-20191215; (j) ZY02D-20191215(2)

4.2 定量评价

本文基于目视判别的方法获得了 20 景影像的云及云阴影标准图,并将其作为算法效能评估的基准。基于验证算法及改进算法的识别结果及标准图,分别计算云和云阴影的总体精度、生产者精度和用户精度,精度指标的对比热力图如图 7 所示。

图 7 所示的热力图能够利用颜色将精度值以一种直观的方式表达出来:精度值越大,对应色块的颜色越偏向暖色,与之相反,值越小则色调越冷。云和云阴影识别精度的热力图分别由 4×20 、 3×20 个色块组成,每行代表不同的识别算法,每列代表不同的实验影像。6 个热力图反映了云识别及云阴影识别的总体精度、生产者精度和用户精度,能够较全面地反映出 20 景影像利用不同算法进行云和云阴影识别的错分和漏分情况,信息量丰富。

从热力图中的多个评价指标总体来看,本文提及的改进 Fmask 算法在大部分实验影像中明显优于其余验证算法。例如影像 GF5-20191003,在 6 个热力图中改进算法的精度值均大于对应的验证算法。

5 讨 论

5.1 算法的普适性和鲁棒性

云识别方面,通过图 7 看出,对于不同下垫面的高光影像,改进 Fmask 算法与验证算法在山地、城区及平原影像上的云识别精度都保持在较高水平,但相比之下改进及原始 Fmask 算法的结果中正确云像元的数量更多。图 7(b)、(c)中改进及原始 Fmask 算法对应的色块值普遍更大,说明 Fmask 算法相对于验证算法具有更高的云识别精度,可以识别出更多的正确云像元,漏分错分情况更少。

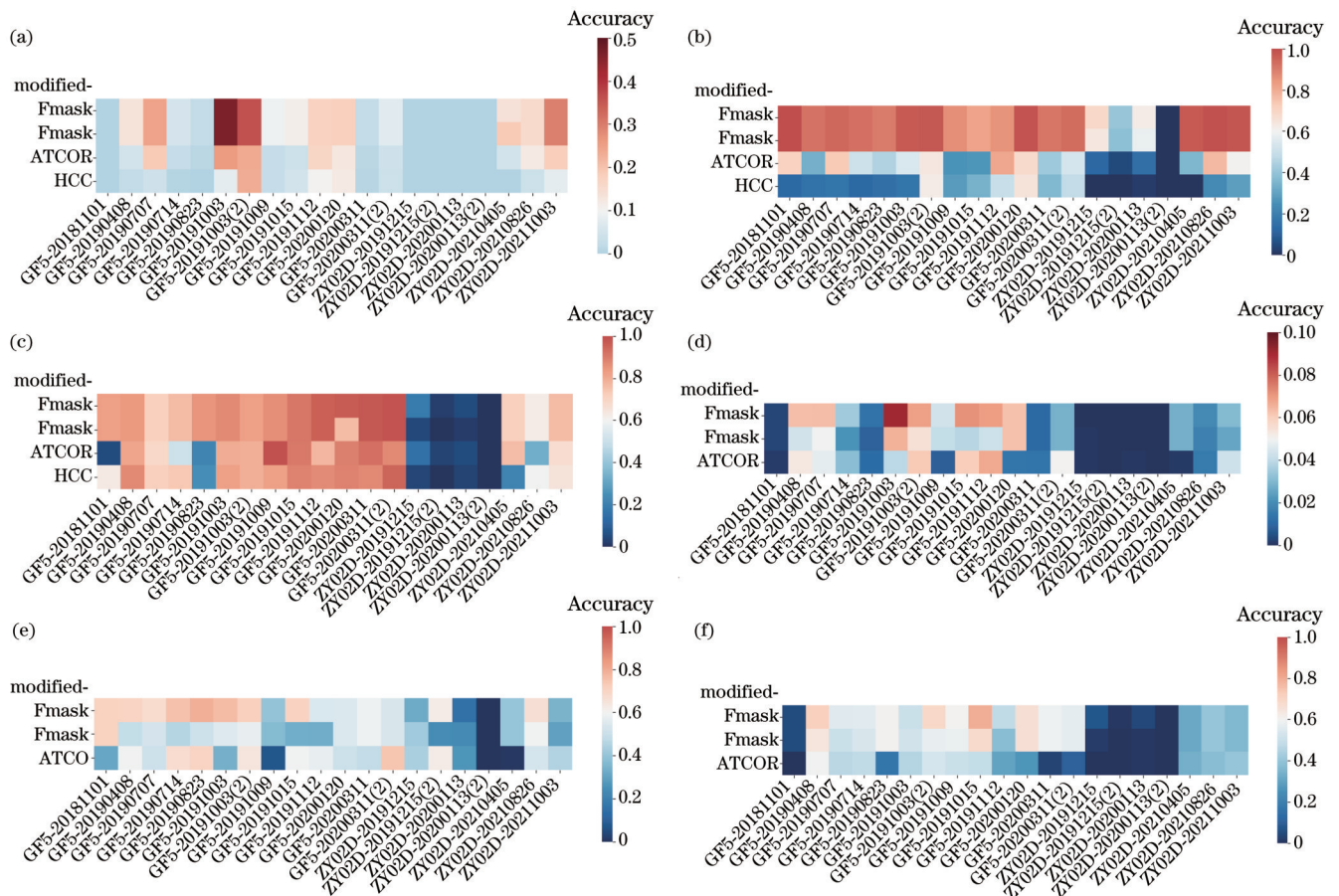


图 7 对比验证的热力图结果。(a)云识别总体精度热力图;(b)云识别生产者精度热力图;(c)云识别用户精度热力图;(d)云阴影识别总体精度热力图;(e)云阴影识别生产者精度热力图;(f)云阴影识别用户精度热力图

Fig. 7 Thermodynamic diagram results for comparative validation. (a) Overall accuracy thermodynamic diagram of cloud identification; (b) cloud identification producer accuracy thermodynamic diagram; (c) cloud identification user accuracy thermodynamic diagram; (d) cloud shadow recognition overall accuracy thermodynamic chart; (e) cloud shadow recognition producer accuracy thermodynamic chart; (f) cloud shadow recognition user accuracy thermodynamic chart

对于覆盖城区的影像,改进 Fmask 算法的结果中出现城区误判为云的情况相对于其他算法更少,这可能是由于改进 Fmask 算法针对该问题进行了具体优化,城区掩模判据的加入使得算法对于城区影像的适

用性显著提升,而其余的验证算法对此并无具体解决方法,因此结果中仍会出现大量误判像元。通过图 6 (i)、(j)中的覆盖城区的影像结果图能够看出,城区误判的情况很少,高亮城区对于云识别的干扰被降低,且二

者对应的改进算法的生产者精度明显高于其余 ATCOR 判据和 HCC 算法,并且相对于原始 Fmask 算法也有提高(图 7),证明在云识别方面本文算法改进效果显著。

云阴影识别方面,三种算法的差异集中在地表有起伏的山地或丘陵影像中。改进 Fmask 算法的云及云阴影匹配以及对于地形阴影的二次处理,使得改进 Fmask 算法在该类影像中的表现相较于两种验证算法更好,例如图 6(e)(GF520200311)、(f)(GF520200311(2))对应的云阴影用户精度均比验证算法高,分别为 58.91% 和 56.87%,说明改进 Fmask 算法对于山地及丘陵影像的云阴影识别更有优势,在保证地形阴影误判少的前提下,补充原先未识别到的云阴影像元,算法适应性更强。因此,对于不同下垫面影像的云及云阴影识别,改进 Fmask 算法具有普适性,识别结果较验证算法更准确。

作为一种针对国产高光谱卫星数据的云识别算法,改进 Fmask 算法的鲁棒性具体体现在:算法对不同云量、分布情况、厚薄程度的云均能识别得到较好的结果。如图 6 中 10 景云量不同的影像的识别结果均与目视判断相符合。不论是图 6(e)中的少云情况,还是图 6(b)中复杂的云覆盖场景,识别结果中红色的云像元均符合目视判断;对于图 6(a)、(c)中的分散式和团聚式云分布,改进算法也能将其很好地标记出来;值得注意的是,改进算法对于图 6(g)这种厚云薄云兼有的影像也有很好的识别效果,能够勾勒出总体符合目视判断的薄云边界。此外,上述影像对应的云识别精度值(图 7)也明显高于验证算法,证明对于云识别,改进 Fmask 算法具有鲁棒性,能够适应多种云量、不同分布类型及高度的云。

5.2 算法精度

精度验证环节,本文认为在云及云阴影识别中总体精度仅作参考,因为不同影像的云及云阴影面积差异明显,不具有可比性,且对于同一影像的不同算法结果,总体精度蕴含信息与生产者精度相似,均能反映算法识别的准确度,因此本节重点关注其余两种精度指标。在这两种指标中,本文认为生产者精度的准确性比用户精度更重要,因为遗漏云或云阴影的代价比识别错误更大。当进行云或云阴影识别时:像元的遗漏会导致无效数据流入遥感应用中,从而影响其结果,如变化检测或图像分类等;而清晰区域被误判为云或云阴影会导致少量数据丢失,对后续遥感应用的结果影响相对较小。因此相比于用户精度,本文在讨论精度指标时将更关注生产者精度。

对比其余验证算法,改进 Fmask 算法在云和云阴影识别方面均有较好的结果,识别精度值总体上更高。

对于云识别,改进 Fmask 算法的云识别总体精度最高为 46.14%;云识别的生产者精度最高可达 99.97%,表明改进 Fmask 算法可以识别出更多的正

确云像元。从热力图中看出,几乎所有影像中改进 Fmask 算法的生产精度都高于 ATCOR 判据和 HCC 验证算法,这体现出改进算法的漏分误差小,真实云像元分类精度高;相对于原始 Fmask 算法,改进算法在对城区影像进行云识别时,生产者精度得到了一定提升,这点也可以从图 7(b)中得到印证,改进后的城区影像(ZY02D-20191215、ZY02D-20191215(2)等)识别结果的生产者精度优于改进前的生产者精度。

同时,改进 Fmask 算法的用户精度最高可达 91.26%,在大部分影像中均优于两种验证算法,表明该算法的错分误差也同样较小。但相比于生产者精度,本文提出算法的用户精度并没有显著高于其他两种验证算法,也存在少数影像中用户精度略低于其余两种验证算法的情况。而与原始 Fmask 算法的对比结论与生产者精度类似,证明改进后算法对于城区影像的云识别能力得到了优化,效果明显。基于实验数据可得,改进 Fmask 算法的云识别精度在总体上优于其余两种验证算法,尤其对解决云像元的漏分情况有一定的优势。

对于云阴影识别,可以得到类似结论。在超过 60% 的影像中改进算法明显优于其余两种验证算法,总体精度最高可达 9.07%,生产者精度可达 78.66%,用户精度可达 79.41%。并且值得注意的是,改进 Fmask 算法的云阴影部分结合了 Fmask 算法和 ATCOR 判据,并表现出更高的生产者精度及用户精度,显示出改进算法对于云阴影识别部分的有效提升。

5.3 算法阈值的敏感性

针对 Fmask 算法的不足,本文增设了两个阈值以改善识别情况,分别为城区掩模判据中的阈值 T 和阴影判据的比例阈值 P ,二者对算法的云及云阴影识别结果有着直接的影响。

在城区掩模判据中,阈值 T 对云识别结果影响显著,这一点可以从图 3 中得到印证。当阈值 T 左调时,云像元正确识别率下降,城区掩模率上升,这意味着云的漏分情况增加,错分情况减少;当阈值 T 右调时情况则相反。本文更关注云的漏分情况,因此在漏分像元数符合期待值的前提下,阈值 T 选用了使得错分像元最少的一 0.25。此外,阈值 T 也可以根据实际情况进行调整,若是期望错分像元数更少,可以左调阈值使之尽量满足要求。同样地,如果对漏分情况更加关注,也可选用更大的阈值来达到期望的结果。

由于没有热红外波段参与识别,本文引入了 ATCOR 软件中的阴影判据作为云阴影识别的补充,但为了尽量减少地形阴影带来的影响,本文构建了比例阈值 P 来进行约束。 P 值的增大会导致地形阴影错分为云阴影的情况增多,但同时云阴影的漏分情况减缓; P 值减小时情况与之相反。从实验结果来看,云阴影的漏分像元数总是远远小于可能被识别为云阴影的地形阴影像元数,因此本文选用 5% 作为比例阈值。阈值 P 也可以根据对漏分错分关注度的不同进行调

整,从而得到更符合预期的云阴影识别结果。

改进 Fmask 算法的云及云阴影识别效果可以从热力图中明显地体现出来。经过对比分析得到,改进 Fmask 算法通过引入判据、增设阈值,较改进前提高了云和云阴影识别精度。识别结果与目视解译更为接近,算法识别性能稳定,可适应于不同下垫面影像、不同类型的云及不同侧重的识别要求。

6 结 论

对于没有热红外波段的高光谱数据,本文针对 Fmask 算法在云和云阴影识别方面精度欠佳的问题,提出了一套适合高光谱卫星的改进算法,重点提升以下两点:1)综合 NDBI 和 NDVI 指数构建城市高亮区域云识别错误像元识别模式,提高城市区域的云识别精度;2)融合 Fmask 算法原有云阴影识别流程和 ATCOR 云阴影识别判据,在保证低山区阴影误识别率的前提下,补充漏分的云阴影像元。该算法在 20 景涵盖城区、山地、平原三种不同下垫面类型的高分五号和资源一号 02D 高光谱卫星图像中开展了应用,主要结论如下:

1)针对没有热红外波段的高光谱数据,提出了高精度的云和云阴影识别改进算法。云识别部分通过改变原有 Fmask 算法结构、加入城区掩模判据减少了城区误判的情况,并增设了自定义的阈值来对结果进行调整,可满足不同侧重(错分、漏分)的云识别实验,可操作性强。对于云阴影识别,引入的可调整的阴影判据能够在弥补没有热红外波段带来的问题的同时,尽可能降低地形阴影的影响。二者均可以通过改变阈值从而在云和云阴影的错分和漏分间达到一个平衡点,灵活地适应不同的地面情况。

2)改进算法对于不同下垫面类型的影像普适性强,在山地、平原及城区影像的识别效果均与目视判断结果较为一致。在 75% 以上的影像中,云及云阴影识别的精度保持在 90% 及 65% 以上,均保持较高水准。此外,改进算法在高精度识别出云及云阴影的前提下,能够有效识别出水体和雪区。

3)以目视判读构建的云图为参考,本文对比了改进 Fmask 算法、Fmask 算法、ATCOR 云判据和 HCC 算法,改进算法的云识别生产者精度和用户精度最高可以达到 99.97% 和 91.26%,在城区影像中的表现优于 Fmask 算法,在超过 80% 的实验影像中优于其余两种阈值算法。

4)云阴影识别部分,同样基于目视判读构建的云阴影图,对比了改进 Fmask 算法、Fmask 算法、ATCOR 云阴影判据,云阴影识别的生产者精度和用户精度分别可达 78.66% 和 79.41%,超过 80% 的影像中明显优于验证算法,其他影像中结果与两种验证算法结果相似,体现出改进 Fmask 算法在云阴影识别部分的优越性。

为了更好地对没有热红外波段的高光谱数据进行云识别,改进 Fmask 算法引入高亮地区识别及控制地形阴影误差的判据,云及云阴影识别精度明显优于传统阈值算法。该算法借用 Fmask 算法已有框架针对高光谱数据进行具体改进,运行速度快,操作步骤简单,有望用于业务化产品生产。但由于山地地区地形复杂,山顶较亮部分及山谷较暗部分与云及云阴影光谱特征相似,且没有亮温数据加以约束,使得在山地影像中的高精度云及云阴影识别仍存在较大困难,如图 6(i)、(j)都还存在错分的云及云阴影。未来也将针对该问题深入研究,或加入 DEM 数据以增强云和云阴影识别算法在多种类型下垫面高光谱影像中的适用性。

参 考 文 献

- [1] Harshvardhan, Randall D A, Corsetti T G. Earth radiation budget and cloudiness simulations with a general circulation model[J]. *Journal of the Atmospheric Sciences*, 1989, 46(13): 1922-1942.
- [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] Goodwin N R, Collett L J, Denham R J, et al. Cloud and cloud shadow screening across Queensland, Australia: an automated method for Landsat TM/ETM + time series[J]. *Remote Sensing of Environment*, 2013, 134: 50-65.
- [4] Hagolle O, Huc M, Pascual D V, et al. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENUS, LANDSAT and SENTINEL-2 images[J]. *Remote Sensing of Environment*, 2010, 114(8): 1747-1755.
- [5] 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.
- [6] Liu R G, Liu Y. Generation of new cloud masks from MODIS land surface reflectance products[J]. *Remote Sensing of Environment*, 2013, 133: 21-37.
- [7] Lyapustin A, Wang Y, Frey R. An automatic cloud mask algorithm based on time series of MODIS measurements[J]. *Journal of Geophysical Research*, 2008, 113(D16): D16207.
- [8] 尚华哲, 胡斯勒图, 李明, 等. 基于被动遥感卫星可见至红外通道观测的云特性遥感[J]. *光学学报*, 2022, 42(6): 0600003.
- [9] Shang H Z, Husi L, Li M, et al. Remote sensing of cloud properties based on visible-to-infrared channel observation from passive remote sensing satellites[J]. *Acta Optica Sinica*, 2022, 42(6): 0600003.
- [10] Rossow W B, Moshier F, Kinsella E, et al. ISCCP cloud algorithm intercomparison[J]. *Journal of Climate and Applied Meteorology*, 1985, 24(9): 877-903.
- [11] Rossow W B, Schiffer R A. ISCCP cloud data products[J]. *Bulletin of the American Meteorological Society*, 1991, 72(1): 2-20.
- [12] Rossow W B, Garder L C. Cloud detection using satellite measurements of infrared and visible radiances for ISCCP[J]. *Journal of Climate*, 1993, 6(12): 2341-2369.
- [13] Stowe L L, McClain E P, Carey R, et al. Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data[J]. *Advances in Space Research*, 1991, 11(3): 51-54.
- [14] Greenhough J, Remedios J J, Sembhi H, et al. Towards cloud detection and cloud frequency distributions from MIPAS infrared observations[J]. *Advances in Space Research*, 2005, 36(5): 800-806.
- [15] Kriebel K T, Gesell G, Ka'stner M, et al. The cloud analysis

- tool APOLLO: improvements and validations[J]. International Journal of Remote Sensing, 2003, 24(12): 2389-2408.
- [15] Lin C H, Lin B Y, Lee K Y, et al. Radiometric normalization and cloud detection of optical satellite images using invariant pixels[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2015, 106: 107-117.
- [16] Zhu Z, Woodcock C E. Object-based cloud and cloud shadow detection in Landsat imagery[J]. Remote Sensing of Environment, 2012, 118: 83-94.
- [17] 张晨, 张秀再, 杨昌军. 基于 RDA-Net 模型的遥感影像云与云阴影检测方法[J]. 激光与光电子学进展, 2021, 58(20): 2028005.
Zhang C, Zhang X Z, Yang C J. Remote sensing image cloud and cloud shadow detection method based on RDA-net model[J]. Laser & Optoelectronics Progress, 2021, 58(20): 2028005.
- [18] 王海涛, 王一琛, 王永强, 等. 基于 MS-UNet 的 Landsat 影像云检测[J]. 激光与光电子学进展, 2021, 58(14): 1401002.
Wang H T, Wang Y C, Wang Y Q, et al. Cloud detection of Landsat image based on MS-UNet[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1401002.
- [19] 童庆禧, 张兵, 张立福. 中国高光谱遥感的前沿进展[J]. 遥感学报, 2016, 20(5): 689-707.
Tong Q X, Zhang B, Zhang L F. Current progress of hyperspectral remote sensing in China[J]. Journal of Remote Sensing, 2016, 20(5): 689-707.
- [20] 王健, 崔天翔, 王一, 等. 高分五号可见短波红外高光谱影像云检测研究[J]. 光学学报, 2021, 41(9): 0928003.
Wang J, Cui T X, Wang Y, et al. Cloud detection for GF-5 visible-shortwave infrared advanced hyperspectral image[J]. Acta Optica Sinica, 2021, 41(9): 0928003.
- [21] 任鹏洲, 岳彩荣. 基于随机森林算法构建 Hyperion 影像云和云阴影掩模[J]. 林业调查规划, 2018, 43(3): 10-15, 39.
Ren P Z, Yue C R. Masking of cloud and cloud shadow for Hyperion image based on random forest algorithm[J]. Forest Inventory and Planning, 2018, 43(3): 10-15, 39.
- [22] 王蓝星, 王群明, 童小华. 融合多光谱影像的高光谱影像厚云去除方法[J]. 测绘学报, 2022, 51(4): 612-621.
Wang L X, Wang Q M, Tong X H. Thick cloud removal of hyperspectral images by fusing with multispectral images[J]. Acta Geodaetica et Cartographica Sinica, 2022, 51(4): 612-621.
- [23] Zhu Z, Wang S X, Woodcock C E. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images[J]. Remote Sensing of Environment, 2015, 159: 269-277.
- [24] Qiu S, He B B, Zhu Z, et al. Improving Fmask cloud and cloud shadow detection in mountainous area for Landsats 4-8 images [J]. Remote Sensing of Environment, 2017, 199: 107-119.
- [25] Qiu S, Zhu Z, He B B. Fmask 4.0: improved cloud and cloud shadow detection in Landsats 4-8 and Sentinel-2 imagery[J]. Remote Sensing of Environment, 2019, 231: 111205.
- [26] Zekoll V, Main-Knorn M, Alonso K, et al. Comparison of masking algorithms for sentinel-2 imagery[J]. Remote Sensing, 2021, 13(1): 137.
- [27] 樊舒迪, 刘振华, 胡月明. 基于光谱特征分析的城市建设用地信息提取[J]. 数字通信世界, 2019(10): 5-9.
Fan S D, Liu Z H, Hu Y M. Remote sensing information extraction of urban built-up land based on spectral signature analysis of normalized difference index[J]. Digital Communication World, 2019(10): 5-9.
- [28] Richter R. A fast atmospheric correction algorithm applied to Landsat TM images[J]. International Journal of Remote Sensing, 1990, 11(1): 159-166.
- [29] Griggin M, Burke H, Mandl D, et al. Cloud cover detection algorithm for EO-1 hyperion imagery[C]//IGARSS 2003.2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477), July 21-25, 2003, Toulouse. New York: IEEE Press, 2004: 86-89.
- [30] Congalton R G. A review of assessing the accuracy of classifications of remotely sensed data[J]. Remote Sensing of Environment, 1991, 37(1): 35-46.
- [31] Story M, Congalton R G. Accuracy assessment: a user's perspective [J]. Photogrammetric Engineering and Remote Sensing, 1986, 52(3): 397-399.
- [32] 梁德印, 安萌, 王啸虎, 等. 资源一号 02D 卫星高光谱相机系统指标论证及总体设计[J]. 航天器工程, 2020, 29(6): 26-34.
Liang D Y, An M, Wang X H, et al. Hyperspectral camera system index argument and system design of ZY-1-02D satellite [J]. Spacecraft Engineering, 2020, 29(6): 26-34.

An Improved Fmask Algorithm for Cloud Detection Applied to Hyperspectral Satellite

Zhang Shuning^{1,2,3}, Zhang Hao^{4*}, Zhang Bing^{1,3}, Cui Zhenzhen^{4,5}, Xiao Chenchao⁶

¹Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;

²International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China;

³College of Resource and Environment, University of Chinese Academy of Sciences, Beijing 100049, China;

⁴Airborne Remote Sensing Center, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;

⁵School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, Henan, China;

⁶Land Satellite Remote Sensing Application Center, MNR, Beijing 100048, China

Abstract

Objective Hyperspectral remote sensing is a new remote sensing technology that emerged in the early 1980s. Hyperspectral data has the advantages of fine spectral resolution, numerous bands, and wide spectral range. It provides

almost continuous spectral curves for each pixel. Its rich spectral information of ground objects can broaden the scope and enhance the depth of remote sensing applications and improve the accuracy and reliability of quantitative analysis. In recent years, hyperspectral technology has developed rapidly in China. The launch of hyperspectral satellites such as GF-5, ZY-1 02D, GF-5B, and ZY-1 02E has enriched abundant hyperspectral data sources and has greatly promoted the development of hyperspectral remote sensing in China. However, hyperspectral satellites, like multispectral satellites, will inevitably be affected by clouds and cloud shadows in the imaging process. Thick clouds in the atmosphere totally cover the reflected surface information, while thin clouds attenuate the reflection of the surface, such as cirrus clouds and haze. Cloud shadow will also degrade the image quality. Therefore, how to accurately identify clouds and cloud shadows has become the key to ensuring the level of further applications. An improved method is proposed to detect clouds and cloud shadows based on domestic hyperspectral satellites.

Methods As a mature cloud detection algorithm, the Fmask algorithm has been widely used and has become the operational algorithm of Landsat and Sentinel product systems. In this algorithm, clouds and cloud shadows are recognized by multiple threshold criteria and flood filling, respectively. Finally, it uses similarity matching to reconfirm cloud shadows, and the detection accuracy of clouds and cloud shadows for Landsat can reach 96.41% and 70%, respectively. However, previous studies have revealed that the detection accuracy of Fmask is relatively low and limited for data without thermal infrared bands. For example, the cloud and cloud shadow detection accuracy of Sentinel-2 data is about 89% and 50%, respectively. It is much lower than the accuracy of Fmask applied to multispectral data. Therefore, an improved Fmask algorithm is proposed specifically for domestic hyperspectral satellites. We optimize the structure of cloud and cloud shadow detection procedures on the basis of the original Fmask algorithm. For urban areas prone to cloud-detected confusion, we add auxiliary judgments to detect bright ground objects. At the same time, the improved algorithm can distinguish the cloud shadow from terrain shadows and improve the accuracy accordingly. 20 hyperspectral images of GF-5 and ZY-1 02D are used to verify the improved algorithm, covering three typical classes, such as urban, mountainous, and flat areas.

Results and Discussions The experimental results indicate that the improved Fmask algorithm performs well in cloud and cloud shadow recognition, highly consistent with the visual recognition results under various underlying surfaces (Fig. 7). The improved Fmask algorithm is compared with the original Fmask algorithm and the other two algorithms, and the cloud and cloud shadow recognition accuracy of all algorithms are calculated, in terms of the overall accuracy, user accuracy, and producer accuracy. The user accuracy and producer accuracy of the improved Fmask algorithm for cloud detection can reach 91.26% and 99.97%, respectively, while the accuracy of cloud shadow detection can reach 78.66% and 79.41%, respectively. The accuracy of all algorithms is illustrated by thermal diagrams (Fig. 8). Evidently, the accuracy of the improved Fmask algorithm is significantly better than the original Fmask algorithm for the scenes containing cities and mountains. Compared with the other two threshold-based algorithms, the improved Fmask algorithm shows significant superiority in aspects of clouds and cloud shadows.

Conclusions This work improves the Fmask algorithm in cloud and cloud shadow recognition to make it suitable for domestic hyperspectral data. The improved algorithm has been tested in 20 hyperspectral images containing typically different underlying surfaces, and the results are highly consistent with the visual recognition. It is also significantly higher than the validation algorithm in terms of user accuracy and producer accuracy in over 60% of images. The improved Fmask algorithm has advantages in terms of robustness, high accuracy, and versatility. The cloud and cloud shadow recognition procedures include an adjustable threshold, which makes the algorithm more flexible to meet different requirements for cloud and cloud shadow recognition. In addition, the improved algorithm does not need extra auxiliary data, running fast and implementing easily. It can be used for high-precision identification of clouds and cloud shadows in hyperspectral data and enable the operational processing of domestic hyperspectral satellite data.

Key words hyperspectral remote sensing; cloud detection; cloud shadow detection; Fmask; GF-5; ZY-1 02D