

用双通道动态阈值对 GMS-5 图像进行自动云检测

刘 希¹⁾²⁾ 许健民³⁾ 杜秉玉¹⁾

¹⁾(南京气象学院,南京 210044)

²⁾(国家海洋局第三海洋研究所,厦门 361005)

³⁾(国家卫星气象中心,北京 100081)

摘 要

该文用双通道动态阈值法对 GMS-5 图像进行自动的云检测。在红外通道和可见光通道,分别对每 32×32 个像元组成的像元阵进行直方图统计,求出区别云和地物值域的阈值,然后对每个像元阵内的逐个像元进行云判识。讨论了在进行直方图分析以求取阈值的过程中,像元阵大小的选取和动态平滑间距的选取对云检测结果的影响。分析结果表明:像元阵大小取 32×32 时,像元阵所占的地域空间尺度足够小,像元阵内的观测像元样本数足够多,保证了在直方图聚类分析时,每一类含有足够多的样本;GMS-5 观测图像的红外通道对原始直方图进行二次平滑时小平滑间距取 1.6 K,可见光通道小平滑间距取 1.2%,使得确定动态阈值时步长相对小,保证了分析的精度。用目视图像对分析结果进行真实性检验,在中低纬度地区,可见光和红外两个通道都有资料时,该算法的云判识精度较好。在高纬度地区,由于地表温度低,积雪覆盖多,太阳光照角低,该算法的云判识精度较差。

关键词:云检测 直方图 阈值

引 言

云是由悬浮在大气中的水滴和(或)冰晶微粒组成的可见聚合物,是降水等许多重要天气现象的产生者。云系在全球大气环流和各种尺度的天气气候系统中起着重要的作用。因此对云的观测和研究,一直是气象工作的一项重要内容。

自从发射了气象卫星,卫星资料就成为进行云分析的一种重要研究手段。为了从卫星数据中取得比较准确的云参数,首先要进行云检测,把有云的像元判识出来。只有识别出无云区,各种地表和大气遥感产品(陆面和海表温度,描述雪/冰覆盖,气溶胶等)才得以计算,从而可以用卫星遥感资料服务于地表环境监测和评估。目前云检测的技术主要分为 3 类:阈值法、聚类分析法和人工神经网络法。

阈值法是一种易于实现,相对成熟的方法。它的思想是将被分析像元不同通道(组合)的亮温、亮温差、反射率与设定的阈值比较,来判识该像元是否被云污染。ISCCP 法^[1]就是一种典型的阈值法,所用的阈值是晴空辐射与观测辐射之间的差。另一个比较成熟的阈值法是 CLAVR 法^[2]。当观测范围很大时,不同地域的阈值设定是不同的,因此阈值的确定就成了阈值法的关键。

聚类分析法是 20 世纪 70 年代由 Diday 提出,80 年代起被用于云分类研究^[3]。这种方法把图像分割成一系列像元阵,通过聚类分析获得像元阵内各像元所代表的云或地表类型。当像元阵内不同地物类型的观测值域客观上存在明显的差别时,可能得到合理的结果,但该方法的结果较依赖于像元阵内观测对象的值域特征,且像元阵与像元阵之间分析结果往往不连续。80 年代末以来,阈值法与聚类分析法趋向于互相融合,其成功的例子就是 APOLLO 法^[4]。

人工神经网络法^[5]模拟人脑神经系统的工作原理,具有很强的自组织自适应的学习能力,它通过对多个样本的学习,获得样本的知识并将知识分布存贮于网络中,从而达到对云的理解和识别。但该方法的结果对于训练资料的选取非常敏感^[6]。

近年来,Alan 等人提出了一种自动化的动态阈值法(DTCM 法)^[7],适用于白天陆地的 AVHRR 图像。该方法认为像元阵的直方图曲线中,地表峰值往云一侧的直方图曲线二阶差分的极大值点,即直方图曲线的最大变率处,更适合作为区分云和地面的阈值。文献^[7]对 3 种不同的数据组合分别进行了云检测,认为用通道 1、通道 3 和 4 的差值进行基于像元的云检测,即 DTCM2 法是最好的方法。

本文的研究工作是国家卫星气象中心风云二号气象卫星业务云检测研究工作的一个组成部分。风云二号气象卫星业务云检测方法倾向于采用目前国际上主流的阈值法。APOLLO 法和 CLAVR 法在确定阈值时都用到多天的历史晴空合成数据。因为风云二号气象卫星目前尚未积累长时间的历史资料,本文的工作目标是:试验当不用历史资料、不知道区域先验气候值和辐射值时,有没有可能获得合理的结果。本文区分云和地表的阈值在 32×32 像元阵范围内用直方图统计的方法求出,采用了 DTCM 的做法^[7]。为了解决好阈值法中动态阈值的确定和像元阵列之间的不连续问题,本文还专门对动态分辨率和像元阵列大小的选取问题进行了讨论。本文首先简述了该项研究工作所采用的双通道动态阈值法的工作流程,并以 1996 年 6 月 11 日 03:00(世界时,下同) GMS-5 图像上我国东南沿海连续的 6×6 个 32×32 像元阵的云检测结果为例说明算法应用的效果,讨论了寻找动态阈值的步长和像元阵大小两种参数的选取问题,最后对双通道动态阈值法进行了小结。

1 双通道动态阈值法的工作流程

与云相比,地表具有较低的反射率和较高的温度。可见光图像定标后,得到的是表观反射率。红外图像定标后,得到的是入瞳等效黑体亮温。为了得到区别云和地表的反射率和温度阈值,把图像分割成一系列像元阵,每个像元阵由 32×32 个像元组成。在每个像元阵内部做直方图统计,在峰的底部确定动态阈值,找出像元阵中反射率最低、温度最高的群落。若像元阵的观测范围内既有海洋又有陆地,应对海洋和陆地像元分别求阈值。

红外图像中很难识别出低云或雾,这是因为它们的温度特征与其下面的地表背景太相似,以至于分辨不出来,而在可见光图像中则相对容易检测到这类云或雾。可见光图像中卷云通常模糊不清,但在红外图像上它却清晰可见,尤其是当它位于比它暖得多的地面之上时,更容易识别。所以结合使用两个通道,比用单个通道更易得到准确的云检测结果。

该算法对于红外通道适用于整张圆盘图,而对于可见光通道,太阳天顶角很大的情况下,太阳反射光被大气介质强烈地散射和吸收,对地物的代表性较差。夜间缺少可见光通道数据,也无法用可见光数据进行云检测。因此人为地规定,对于太阳天顶角小于 70° 的像元,只要在红外和可见光两个通道中的某一通道被判识为云像元,则该像元为云像元,而对于太阳天顶角大于 70° 的像元,则采用红外通道的判识结果。

红外或可见光通道求动态阈值的算法流程基本相似,图 1 为其算法流程示意图。算法由两个部分组成,互相嵌在一起。第 1 部分是用直方图的斜率求阈值,第 2 部分是用二分法求阈值。Alan 等人^[7]认为,直方图上地物峰向云一侧的斜率突变点,比直方图的谷点更能代表区分云和地物的阈值。若观测值域在像元阵内只呈现一个峰,或者对于较宽的峰,分析不出斜率突变点,为了避免值域靠得非常近的两种地物混在一起分不开,试图用二分法把像元阵强制分割为两个群落。

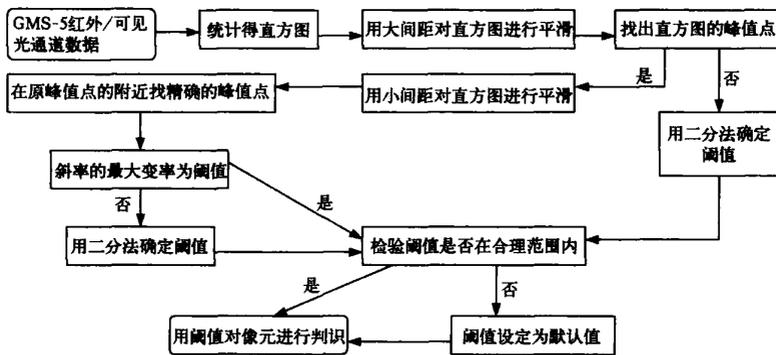


图 1 红外/可见光通道求动态阈值的算法流程

下面以 GMS-5 卫星 1996 年 6 月 11 日 03:00 可见光图像上某一 32×32 像元阵为例,说明图 1 所示的求动态阈值步骤。

(1) 定标后的可见光图像(见图 2)反射率除以太阳天顶角的余弦,归一化到天顶方向。所得归一化后的可见光通道反射率乘以 1000,标在横轴上,使横轴数据范围落在 $0 \sim 1000$ 之间,扩大了数据的动态范围,在以后的计算中减少截断误差。可见光图像原始灰阶范围为 $0 \sim 63$,扩大动态范围后,每隔 16 个等级相当于跨越了一个灰阶。对像元阵内的 1024 个像元统计,得直方图(见图 3a)。

(2) 对直方图进行平滑。设平滑间距为 18,使得统计在大致相当于一个灰度等级的范围内进行。每个平滑间距内像元的

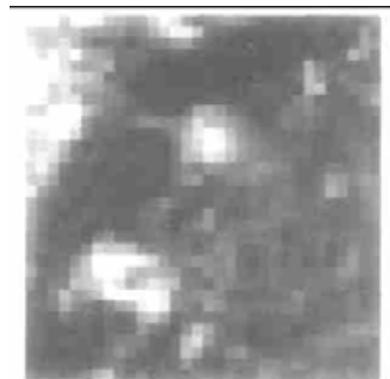


图 2 GMS-5 卫星 1996 年 6 月 11 日 03:00 可见光图上某一 32×32 像元阵图像

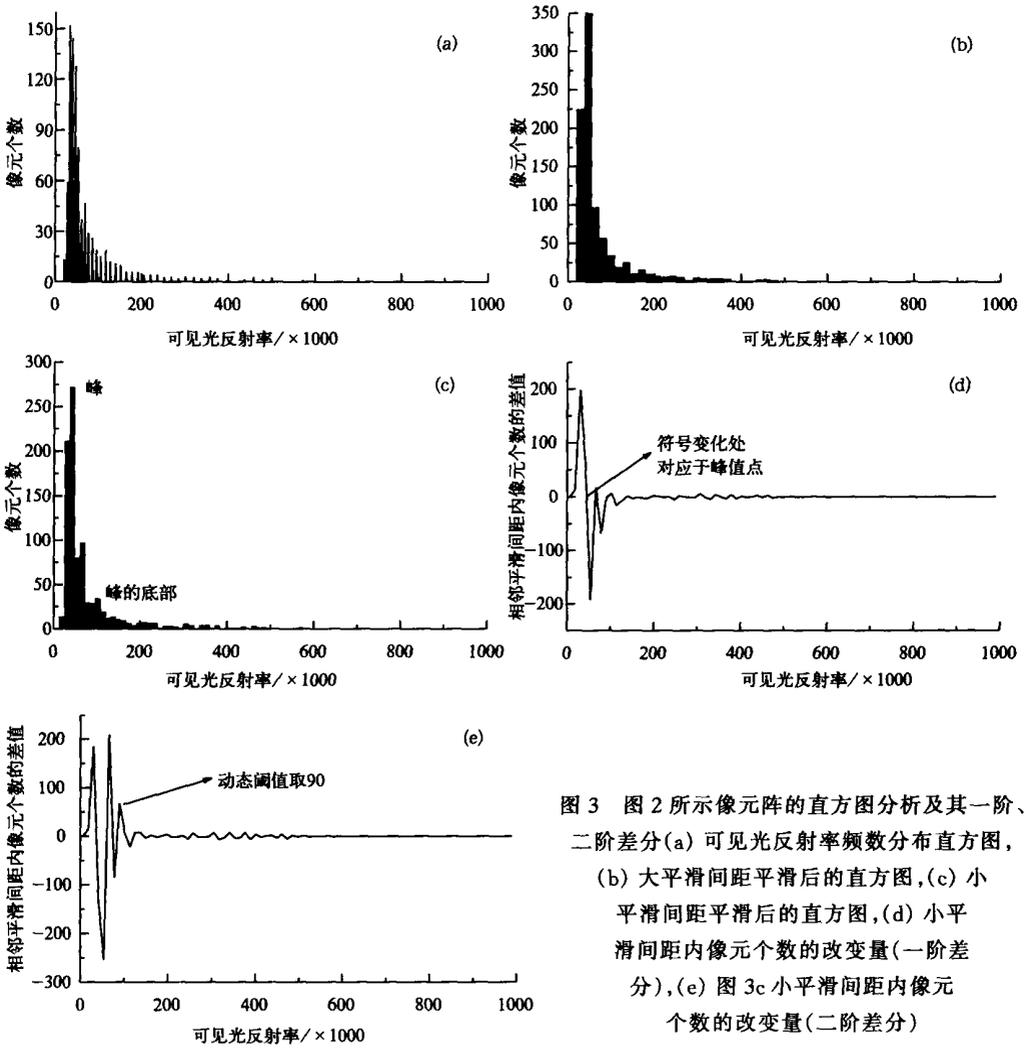


图3 图2所示像元阵的直方图分析及其一阶、二阶差分 (a) 可见光反射率频数分布直方图, (b) 大平滑间距平滑后的直方图, (c) 小平滑间距平滑后的直方图, (d) 小平滑间距内像元个数的改变量(一阶差分), (e) 图3c小平滑间距内像元个数的改变量(二阶差分)

个数为纵坐标值,每个平滑间距的中心,如9,27等,为横坐标。横坐标取9时,对应的纵坐标为落在0~18内的像元数。以此类推,即得一次平滑后的直方图(见图3b),最右边剩下的不能被18整除的高值点未做统计。

(3) 用相邻平滑间距内像元个数的改变量除以平滑间距,得直方图的变化趋势,即一阶差分。对于同一像元阵而言,由于平滑间距相同,只需求相邻平滑间距内像元个数的差值,就得到曲线的一阶差分,从而避免了浮点计算。地物群落的峰值点就落在斜率从正值变成负值的地方。

(4) 为了得到更精确的峰值点,在较小的平滑尺度(平滑间距取12)(图3c)上求一阶差分。即在第(3)步找到的峰值点附近,对一阶差分从正变负的值进行再确认(图3d)。

(5) 当像元阵内云和地表的物理特性(值域)相差较大时,直方图呈多峰形态。但是当像元阵内云和地表的物理特性差别不大时,两种地物特征可能落在直方图的同一个群

落里。由两个群落组成的直方图的形状,可能与单一群落直方图的形状略有不同,表现为斜率(一阶差分)有显著的变化。为了分析斜率的显著变化,在用小平滑间距对原始直方图进行平滑,求一阶差分的基础上,求相邻平滑间距内像元个数的改变量,除以平滑间距,就得到直方图的二阶差分(图 3e)。二阶差分反应了直方图斜率的变化率。从精确的峰值点加上两个小平滑间距开始,继续找二阶差分的极大值。二阶差分极大值所在的地方,就是斜率变化最多的地方,可以用它来区分群落。

(6) 为了使分析结果有代表性,设分析出的任何一种群落(包括地表群落)含有的像元数,应大于 32 个像元。

(7) 直方图上分析出的动态阈值和峰的差值不应小于两倍小平滑间距。在图 3d 所示的例子中,精确的峰值点为 60,而图 3e 中 60 以后在 66 和 90 分别出现极大值。但 66 因太接近峰值而舍弃,所以动态阈值取 90。

(8) 若像元阵内地表特性非常均匀,所有像元的值可能都落在精确峰值左右两个小平滑间距以内,而在离精确峰值点两个小平滑间距以外,直方图曲线的二阶差分均为 0,找不到斜率的最大变率。在这种情况下取第一个二阶差分为 0 的点为动态阈值。

(9) 当以上步骤找不到动态阈值时,为了避免值域靠得非常近的两种地物混在一起分不开,用二分法把像元阵强制分割为两个群落。二分法简单地假设像元阵中既有云像元,又有地表像元。首先用这 1024 个像元的平均值 m 将像元阵中的像元分成两组,分别计算这两组像元的平均值 m_1 和 m_2 ,取 $m = (m_1 + m_2)/2$,再将这 1024 个像元分成两组。重复以上步骤,直至持续两次分割的 m 值不发生大的改变。阈值取两个群落均值的中间值。

(10) 考虑到太阳耀斑的影响,海洋像元阈值的上限是 20%。考虑到雪的可见光反射率较高,陆地像元阈值的上限是 40%。当动态阈值或二分法确定的阈值超过上限时,阈值取默认值。海洋上阈值的默认值为 15%,陆地上阈值的默认值为 30%。

(11) 表观反射率大于动态阈值的像元为云像元,反之为晴空像元。

与上面分析的可见光像元阵相对应的红外原始图像为图 4。像元阵内 1024 个像元统计所得的直方图见图 5。红外通道大平滑间距取 12,小平滑间距取 8。找峰值及谷值从高值往低值方向进行,和可见光通道正好相反,其他步骤和可见光通道相同。



图 4 图 2 所对应区域的红外像元阵图像

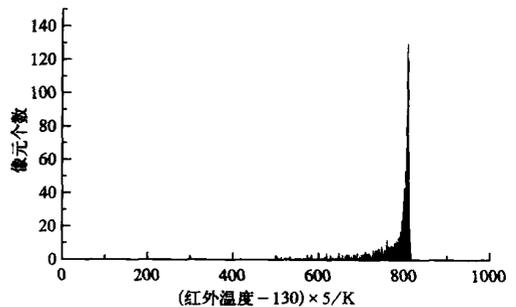


图 5 红外亮温频数分布直方图

红外数据定标后温度取值范围大致为 130 ~ 330 K,观测亮温高于 330 K 时,统一设定为 330 K。将所有数据减去 130 K,乘以 5,使数据范围落在 0 ~ 1000 之间,便于计算。海洋像元阈值的下限是 263 K,陆地像元阈值的下限是 253 K。当动态阈值法和二分法找到的阈值超出合理范围时,采用默认值。海洋上阈值的默认值为 269 K,陆地上阈值的默认值为 263 K。

2 用 GMS-5 双通道数据进行云检测的实例

用动态阈值法对 1996 年 6 月 11 日 03 :00 至 20 日 03 :00 ,2003 年 2 月 11 日 04 :00 至 20 日 04 :00 的 GMS-5 整张圆盘图进行了云检测和目视观测。可见光通道的云检测结果优于红外通道,能判识出绝大部分的云。因此在中低纬度地区,可见光和红外两个通道都有资料时,该算法的云判识精度较好。在高纬度地区,由于地表温度低,积雪覆盖多,太阳光照角低,该算法的云判识精度较差。

以 1996 年 6 月 11 日 03 :00 我国东南沿海连续的 6 × 6 个 32 × 32 像元阵为例,图 6a 为 1996 年 6 月 11 日 03 :00 我国东南沿海的红外图像,图 6b 为相应的可见光图像,图 7a 为用红外单通道进行云检测的结果,图 7b 为用可见光单通道进行云检测的结果,图 7c 为同时用两个通道进行云检测的结果。从图 7a 可看出只用红外通道进行云检测时,漏判了海面上和陆地上空的部分低云、碎云。这是因为这些地方云顶温度和地表温度十分接近。而这些云在可见光图像上清晰可见,因此在图 7b 中它们大部分被可见光图像判识出来了。从表 1 可看出仍有少部分云在可见光通道未被判识出来,而在红外通道被判识为云。而从图 7c 可见,同时用两个通道的图像,绝大多数云被正确地判识出来了。

表 1 像元个数统计

总的像元个数	红外通道判识出的云像元个数	可见光通道判识出的云像元个数	双通道判识出的云像元个数	标识海陆边界所占用的像元个数
36864	5918	9328	10440	1292

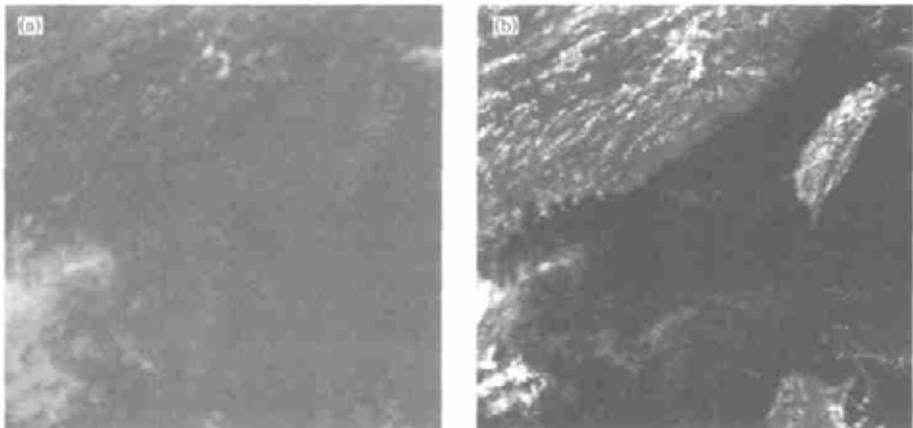


图 6 1996 年 6 月 11 日 03 :00 我国东南沿海的卫星云图

(a) 红外图像,(b) 可见光图像

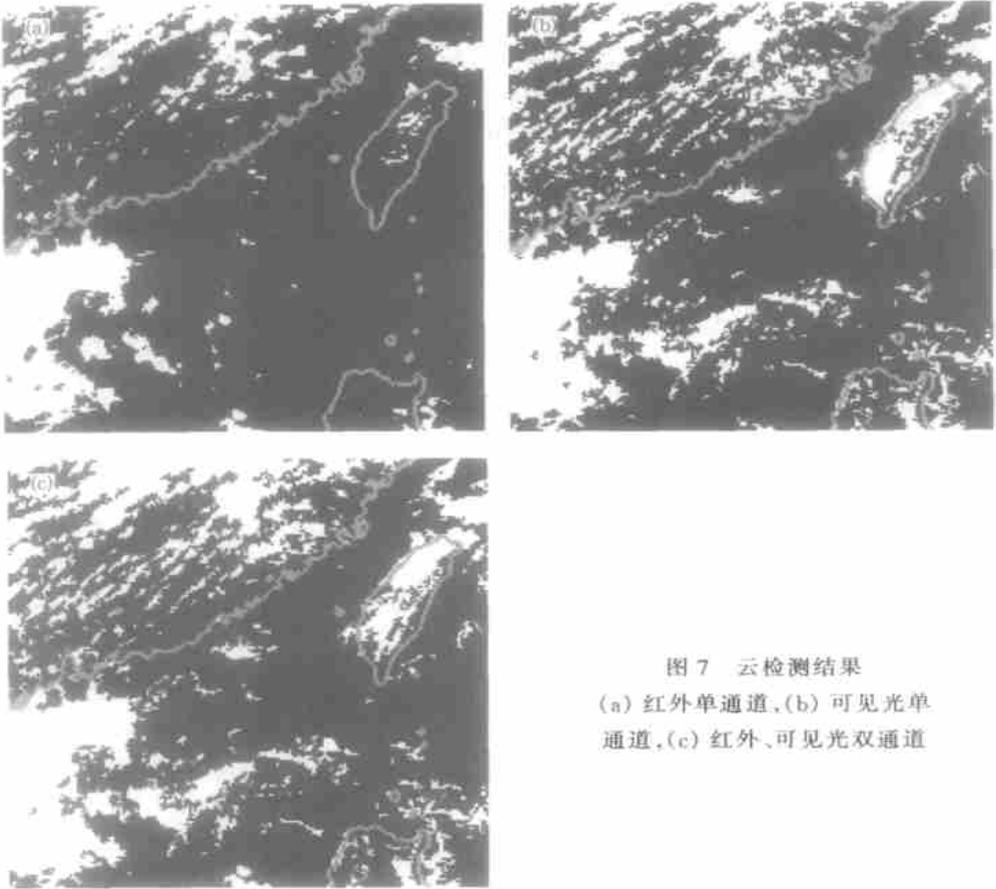


图7 云检测结果
(a) 红外单通道,(b) 可见光单通道,(c) 红外、可见光双通道

3 影响云检测结果的原因分析

3.1 算法不成功的原因分析

动态阈值法假定在被分析像元阵中存在晴空像元。在进行直方图统计时,有峰与晴空像元相对应。用目视图像与分析结果对比作为真实性检验,注意到算法不成功与以下事实有关:

(1) 若像元阵中云像元的个数超过 80%,晴空像元极少,那么在直方图上往往找不到地表特征所对应的峰值点,从而找不到区分云和地表的动态阈值。

(2) 如果地表物理特性均一,直方图中地表群落呈单峰形态,此时容易找到动态阈值;如果地表的物理特性起伏很大,例如既有沙漠,又有植被,则直方图会出现两个代表地物的峰,分别对应沙漠和植被,这时容易将沙漠误判成云。

(3) 若像元阵内有积雪,由于雪具有较高的反射率和较低的温度,容易将雪误判成云。

(4) 图像是连续的,而相邻两个像元阵的动态阈值不连续。在这种情况下,像元阵的

边缘可能会出现判识结果不连续,即俗称的马赛克现象。这种现象在高纬度地区出现较多。

(5) 在中低纬度地区,结合使用可见光通道和红外通道,该算法具有较好的云判识结果。在高纬度,地表温度较低,有积雪存在,且光照条件差,只用红外一个通道来判识云。这些因素使得高纬度地区云检测的精度降低。

以上事实说明,直方图分析的结果,与目标物本身的物理性质关系甚大。对目标物观测特征的了解,是设计直方图分析方法的基础。下面对直方图分析方法的设计进行一些解剖。

3.2 像元阵大小对云检测结果的影响——海陆交界处和圆盘图边缘的分析

进行直方图分析的目的,是从中找到区分云和地物的阈值。找阈值的过程是从地物(低反射率,高亮温)一端开始的,因此找到合理阈值的必要条件是像元阵中存在一定数目的地物像元。在海陆交界处的 32×32 像元阵分为陆地区域和海洋区域各自进行分析。若陆地(海洋)区域含有的像元数太少,就分析不出合理的结果。下面以 1996 年 6 月 11 日 03:00 的 GMS-5 的红外通道图像为例,对陆地和海洋动态阈值确定情况进行分析(表 2)。

表 2 动态阈值确定情况的分析

像元阵中同类像元的个数	陆地像元阵			海洋像元阵		
	像元阵的个数	占陆地像元阵总个数的百分比/ %	找到动态阈值的个数	像元阵的个数	占海洋像元阵总个数的百分比/ %	找到动态阈值的个数
1 ~ 100	283	21.89	13	178	5.51	21
101 ~ 200	83	6.42	43	60	1.86	44
201 ~ 300	76	5.88	55	56	1.73	46
301 ~ 1023	418	32.33	406	673	20.82	670
1024	433	33.49	433	2265	70.08	2265
合计	1293		950	3232		3046

像元阵中陆地像元数少于 100 的有 283 个,占陆地像元阵总个数的 21.89%,其中找到动态阈值的像元阵个数仅为 13,即当陆地像元的个数少于 100 时,找到动态阈值的比例不到 10%。

像元阵中的海洋像元数少于 100 的有 178 个,占海洋像元阵总个数的 5.51%,其中找到动态阈值的像元阵个数仅为 21,即当海洋像元的个数少于 100 时,找到动态阈值的比例不到 12%。

随着分析区域内像元个数的增加,找到动态阈值的比例迅速增加。当分析区域内像元的个数超过 300 时,基本上都能找到动态阈值。因此分析区域的大小至少要包含 300 个像元左右。

3.3 像元阵大小的选择对云检测结果的影响

像元阵越大,相邻像元阵的动态阈值差别就越大,而图像是连续的,因此像元阵的边缘容易出现马赛克现象。如果像元阵设得很小,像元数目太少,使得分析结果不具备统计特性。 32×32 像元为一个像元阵是比较合理的选择。此时像元数目足够多,具有统计特性,且每一像元阵所对应的实际地表区域不是太大,像元阵内地表反射率和温度变化相对小。

以1996年6月11日03:00的GMS-5的红外通道为例,当像元阵大小取 32×32 时,下垫面为陆地的像元阵找到动态阈值的比例为73.47%,采用二分法得阈值的比例为0.39%,采用默认阈值的的比例为26.14%;下垫面为海洋的像元阵找到动态阈值的比例为94.25%,采用二分法得阈值的比例为1.67%,采用默认阈值的的比例为4.08%。

采用二分法得到阈值的比例很少,这是因为当没找到陆地或海洋峰值点,即没找到动态阈值时,二分法得到的阈值多半已经超出了阈值的上下限,只能将阈值设为默认值。

3.4 平滑间距的选取对云检测结果的影响

以1996年6月11日03:00 GMS-5的红外通道为例,表3统计了不同的平滑间距对原始直方图进行平滑时,采用动态阈值的像元阵占像元阵总数的比例,未找到动态阈值而采用二分法所得的阈值的比例及当以上两种方法所得的阈值均不在合理范围内,而采用默认阈值的比例。

表3 选取不同的平滑间距时阈值为动态阈值、二分法阈值、默认阈值的比例

大平滑 间距值	小平滑 间距值	陆地像元阵/ %			海洋像元阵/ %		
		动态阈值	二分法	默认阈值	动态阈值	二分法	默认阈值
6	4	66.98	0.46	32.56	91.65	0.19	8.17
8	6	70.38	0.08	29.54	92.91	0.68	6.41
12	8	73.47	0.39	26.14	94.25	1.67	4.08
16	12	76.33	0.31	23.30	94.71	3.43	1.85
20	15	78.34	0.46	21.19	95.02	4.05	0.93
30	20	80.59	1.55	17.86	95.14	4.25	0.61

平滑间距取4,8,15时,寻找动态阈值的步长分别为0.8 K,1.6 K,3 K。可见平滑间距越小,确定的动态阈值就越精确。但平滑间距太小时,地表像元在直方图曲线上可能形成双峰或多峰,此时找不到正确动态阈值的可能性增大,从而检测出的云偏多,即易将一部分地表误判为云。

GMS-5 红外通道图像仪器的分辨率在入瞳等效黑体亮温大于263 K时,大约是0.3~0.5 K,因此当寻找动态阈值的步长越小,每个步长内含有的样本数越少。为了使分析结果有代表性,本文假定分析出的地表群落至少应含有32个像元(大于像元阵中像元总数的 $1/32$)。当分析步长太小时,会因为样本数太少而无法确定动态阈值,从而使找到动态阈值的百分比降低。陆地区域的像元阵找到动态阈值的百分比随平滑间距的减小而减小,比海洋区域快得多。这是因为海洋的特性较均一,地表群落较集中,同一步长内的样本数相对多,有利于找到动态阈值。

对GMS-5观测图像,红外通道对原始直方图进行二次平滑时小平滑间距取8,寻找动态阈值的步长为1.6 K,可见光通道小平滑间距取12,寻找动态阈值的步长为1.2%,既保证直方图聚类分析得到的类别中,含有足够多的样本,又使得动态阈值的步长相对小,保证了分析精度。

4 小结

(1) 双通道动态阈值法是一个相对简单、完全自动化的云检测方法。本文的研究工

作表明,当不用历史资料、不知道区域先验气候值和辐射值时,有可能获得合理的结果。

(2) 分析区域的大小至少要包含 300 个像元左右。以每 32×32 像元为一个像元阵是比较合理的选择,保证了直方图聚类分析得到的类别中,含有足够多的样本。分析像元阵选取过大时,容易出现马赛克。

(3) 对 GMS-5 云图,红外通道对原始直方图进行二次平滑时小平滑间距取 1.6 K,可见光通道小平滑间距取 1.2%,即保证每个步长内的样本数足够多,又使得确定的动态阈值的步长相对小,保证了分析精度。

(4) 海洋上由于下垫面物理性质较接近,在进行直方图统计时,峰的曲线较光滑,所以容易确定动态阈值。在陆地区域,地表特性相差大,陆地特征在直方图上表现为多个峰,不利于正确地确定动态阈值。

(5) 在中低纬地区白天,结合使用可见光通道和红外通道,该算法具有较好的云判别结果。但在中低纬地区夜间,以及高纬度地区仅用红外通道时,云检测的准确度降低。地表温度的降低,积雪的存在,这些因素也使得高纬度地区云检测的准确度降低。

(6) 当没找到动态阈值时,绝大部分情况是采用了默认值。因此提高默认值的准确度就等于提高了云检测的准确度。在今后的工作中,若对不同区域,特别是对应高纬度地区,利用气候值来分别设定阈值的上下限和默认值,将可以提高云检测的准确度。

参 考 文 献

- 1 Rossow W B, Garder L C. Validation of ISCCP cloud detections. *J Climate*, 1993, **6**: 2370 ~ 2393.
- 2 Stowe L L, Davis P A, McClain P E. Scientific basis and initial evaluation of the CLAVR-1 global clear/cloud classification algorithm for advanced very high resolution radiometer. *J Atmos Ocean Technol*, 1999, **16**: 656 ~ 681.
- 3 Desbois M, Seze G, Szejwach G. Automatic classification of clouds on meteosat imagery, Application to high-level clouds. *J Climate Appl Meteor*, 1982, **21**(3): 401 ~ 412.
- 4 Saunders R W, Kriebel K T. An improved method of detecting clear sky and cloudy radiances from AVHRR data. *Int J Remote Sens*, 1988, **9**: 123 ~ 150.
- 5 Peak J E, Tag P M. Segmentation of satellite imagery using hierarchical thresholding and neural networks. *J Appl Meteor*, 1994, **33**: 605 ~ 616.
- 6 Nowcasting and very short range forecasting SAF, Prototype Scientific Description for Météo France/ CMS. SAF/ NWC/ MFCMS/ MTR/ PSD, Issue 1, Rev. 1, 2000.
- 7 Alan V D V, William J E, Senior Member. An automated dynamic threshold cloud masking algorithm for daytime AVHRR images over land. *IEEE Transactions on Geoscience and Remote Sensing*, 2002, **40**(8): 1682 ~ 1694.

A BI-CHANNEL DYNAMIC THRESHOLD ALGORITHM USED IN AUTOMATICALLY IDENTIFYING CLOUDS ON GMS-5 IMAGERY

Liu Xi¹⁾²⁾ Xu Jianmin³⁾ Du Bingyu¹⁾

¹⁾ (*Nanjing Institute of Meteorology, Nanjing 210044*)

²⁾ (*The Third Institute of Oceanography, State Oceanic Administration of China, Xiamen 361005*)

³⁾ (*National Satellite Meteorological Center, Beijing 100081*)

Abstract

A bi-channel dynamic threshold algorithm is used to identify clouds on GMS-5 images. Threshold values of clouds and surface objects are gained in visible and infrared window channels by means of a statistic histogram analysis made of 32×32 pixel matrixes, followed by cloud recognition done from pixel to pixel in each of the matrixes. The effect on cloud recognition of the chosen segment and bin size for dynamic smoothing in obtaining threshold values is concerned in terms of the histogram analysis. Results show that it is appropriate to choose 32×32 as segment size, thus ensuring a sufficiently large number of samples for each kind in histograms clustering analysis; when the histogram is smoothed it is appropriate to choose the small bin at 1.2% (1.6 K) for the visible (IR) channel so that a relatively short step length is applied in determining dynamic threshold value for required accuracy. Test of the results by visually examining images shows that with visible and IR data available the recognition precision from the algorithm is quite good at middle and low latitudes but the precision is not so good at higher latitudes where surface is colder, snow cover is large and the sun's elevated angle is low.

Key words: Cloud recognition Histogram Threshold