

基于支持向量机的遥感大雾判识

刘年庆 蒋建莹 吴晓京

(国家卫星气象中心,北京 100081)

提 要: 提出了一种基于支持向量机的卫星遥感数据大雾判识方法:首先通过对风云 1D 卫星大雾区域各通道辐射值出现频次进行概率统计,利用其阈值来粗判识大雾;然后在粗判识的基础上通过支持向量机的方法进行大雾细判识;最后利用腐蚀和膨胀的图像处理技术对判识后的图像进行优化处理。在对我国 2006 年 9—12 月的 65 条监测到大雾的风云 1D 轨道的探测数据进行分析之后,发现大雾判识结果与专家标记吻合。检验结果表明,利用 1、2、4、6、7、10 通道组合进行粗判识的结果最好,5 交叉正确率为 89.9849%,TS 评分为 74.04%。利用上述方法对个例的分析检验表明,基于支持向量机的遥感大雾判识方法是切实可行的。

关键词: 遥感大雾判识 概率统计 支持向量机 腐蚀膨胀

Fog Judgment Based on the Support Vector Machine by Remote Sensing Data

Liu Nianqing Jiang Jianying Wu Xiaojing

(National Satellite Meteorological Center, Beijing 100081)

Abstract: A method is put forward to recognize the fog based on the support vector machine, according to the satellite remote sensing data. Firstly, the probability statistics method is used to roughly judge the fog, according to the frequency of the fog areas appearing at different channels of FY-1D satellite; secondly, based on the former judgment, the support vector machine is applied to judge the fog carefully; lastly, erosion and dilation techniques are used to optimize the result of the second procedure. From September to December in 2006, 65 overpasses of FY-1D satellite data including fog areas are analyzed, and the judged fog areas are found to correspond well to the experts' experience. And the result shows that the combination of 1, 2, 4, 6, 7 and 10 channels is the best of judgment. The 5-fold cross-validation is 89.9849% and the TS score is

74.04%。This method is also used to recognize the fog during other time, and found that this method is excellent.

Key Words: fog judgment by remote sensing data probability statistics support vector machine erosion and dilation

引 言

雾是一种重要的天气现象,它由漂浮在近地面的极细小的水滴(或冰晶粒子)组成,大雾的出现对交通运输及人们的日常生活造成极大影响,并且严重污染了大气。近年来对雾的研究逐步成为热点,众多学者从大雾的天气背景、物理化学结构、局地大雾的宏微观结构及物理成因等方面进行了许多研究^[1-5]。自 1960 年代 Estoque 等^[6]建立大气边界层的数值模式后,雾的数值模拟工作相继开展起来,利用复杂地形上的二维雾模式、三维雾模式、中尺度数值模式系统 MM5^[7-9]等,对大雾的形成和维持机制进行了细致分析。

遥感技术是当前人类研究地球资源环境的一种主要技术手段,卫星资料具有覆盖范围广、信息量大、重复频率高、客观形象、信息源可靠、资料获取方便、费用低廉等诸多优势,是其它各种常规观测方法所无法替代的。遥感图像的分类判识是遥感资料应用的基础,一直被从事遥感图像处理的广大科技人员所重视。由于从地面发射或反射的辐射通过大气介质向上传播,不同成分、不同温度的非均匀大气对不同波段的电磁辐射就有着不同的吸收、辐射和散射特性,通过分析云、雾粒子在卫星仪器不同通道的辐射和反射特性,使得卫星遥感监测大雾成为可能^[10-12]。

1990 年代以来,人工神经网络模型在遥感图像分类中的应用取得了很大的进展,但神经网络算法有其固有的缺点,如网络结构的确定尚无可靠的规则,易陷入局部极小

等^[13-14]。为了提高遥感图像分类精度,还必须寻找一些新的学习算法。支持向量机(SVM, Support Vector Machine)是 Vapnik^[15]等人根据统计学理论提出的一种新的通用学习方法,它建立在统计学理论的 VC 维理论和结构风险最小原理基础上,能较好地解决小样本、非线性、高维数和局部极小点等实际问题。近年来在模式识别、回归分析和特征提取等方面得到了很多应用,正逐渐在各行各业显示出强劲的潜力。陈永义^[16]等人 2004 年开始将 SVM 引入气象领域,为我国气象资料处理和预报方法研究开拓了一个新方向。

支持向量机方法继承了最大间隔超平面、Mercer 核、凸二次规划、稀疏解和松弛变量等多项技术,可望避免传统用经验判识大雾区域求解方法中繁琐的设计与编程,而只需要收集一定量的有输出值的样本数据,再运行一个现成的算法来学习输入到输出的映射,为研究大雾判识开辟了一个新的思路。

1 支持向量机基本原理及其回归算法的实现

支持向量机是在高维特征空间使用线性函数假设空间的学习系统,它由一个来自最优化理论的学习算法训练,该算法实现了一个由统计学习理论导出的学习偏置^[17]。

其基本思想简单地说就是升维和线性化,在约束条件,

$$y_i(\omega^T \phi(x_i) + b) \geq 1 - \zeta_i, \zeta_i \geq 0 \quad (1)$$

下完成优化目标

$$\min_{\omega, b, \zeta} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l \zeta_i \quad (2)$$

其中, $\phi(x_i)$ 表示样本到高维空间的映射, ω 是权向量, C 是正则化参数, 表示惩罚系数。对于任何一个样本 x_i , 如果分类超平面 $y_i(\omega^T \phi(x_i) + b) > 0$, 则称将样本正确分类; 如果 $y_i(\omega^T \phi(x_i) + b) \geq 1$, 说明它以几何间隔 $\frac{1}{\|\omega\|}$ 将样本正确分类, 否则用 ζ_i 记录该样本离所需间隔 $\frac{1}{\|\omega\|}$ 的相差程度, 将小于几何间隔的样本都记录为分类有错误。

从目标函数可以看出, 最小化 $\frac{1}{2} \omega^T \omega$, 要求分类超平面的间隔 $\frac{1}{\|\omega\|}$ 尽可能的大, 统计学习理论证明间隔越大对应的泛化性能越好。最小化 $\sum_{i=1}^l \zeta_i$ 对应小于间隔的样本错误程度之和。提出这项, 主要是考虑到有时数据在高维仍然线性不可分的情况。那么系数 C 就控制了这两种情况的折中, C 的大小表明对错误惩罚程度的多少, 一般通过交叉验证来确定 C 的值。

注意到在约束不等式 $y_i(\omega^T \phi(x_i) + b) \geq 1 - \zeta_i$ 中含有 $\phi(x_i)$, 它表示将样本映射到高维空间, 如果选用 Gauss 核函数:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (3)$$

其中 γ 为核函数参数, 那么由 $K(x_i, x_j) \leq$

$\phi(x_i) = \langle \phi(x_j) \rangle$, 就定义了这种核映射。

2 大雾判识方法

大雾自动判识方法可以实现对大雾区域的标准化选取, 使其不受个人判识经验的影响, 主要分成三个部分: 首先是通过卫星不同通道辐射值的概率统计来粗判识大雾区域; 然后在粗判识的基础上通过支持向量机的方法进行细判识; 最后利用腐蚀和膨胀的图像处理技术对判识后的图像进行优化处理。

2.1 利用概率统计进行粗判识

在对大雾进行粗判识之前, 首先来讨论风云 1D 卫星 10 个通道的辐射特性。这里选取 2006 年 9—12 月共 65 次监测到大雾的卫星轨道探测数据, 大雾区域的可见光和近红外通道、红外通道数值出现频次统计如图 1(见彩页)所示。

由图可见: 大雾区域在不同通道上的数值分布存在明显不同, 但在同一通道内有近似的正态分布。考虑到噪声等原因, 去掉出现频次低于 1% 的区域, 把出现频次较高的区域作为阈值的上下限, 经过统计, 风云 1D 大雾区域 10 个通道值的上下限如表 1 所示。

表 1 风云 1D 卫星大雾区域各通道值上下限及其均值、方差

通道	1	2	3	4	5	6	7	8	9	10
波长(μm)	0.58~0.68	0.84~0.89	3.55~3.93	10.3~11.3	11.5~12.5	1.58~1.64	0.43~0.48	0.48~0.53	0.53~0.58	0.90~0.965
类别	可见光	近红外	中波红外	长波红外	分裂窗	近红外	可见光	可见光	可见光	近红外
下限	163	193	2779	2694	2681	191	158	191	174	97
上限	709	803	3155	2899	2888	954	514	641	653	483
均值	346.6	384.7	2899.4	2806.4	2795.6	423.2	302.9	371.8	344.2	226.9
方差	85.14	87.02	68.90	46.55	47.14	107.93	65.27	79.41	76.23	63.54

如果一个探测点的 10 个通道值都在阈值上下限范围内, 则此点被认为是雾。利用以上通道阈值进行筛选, 可以有效地去掉非

大雾区域, 去除比例达到 75.51%, 如表 2 所示, 但是剩下的区域不一定是大雾区域, 需要进一步判识。

表 2 单通道/混合通道粗判识去除比例

	单通道判识 去除范围比例	混合通道判识 去除范围比例
按像元平均	75.51%	78.59%
按轨道平均	74.68%	78.00%

以上是通过 10 个通道各自划定的阈值进行筛选的方法,但这种方法并没有考虑到各个通道之间的关系。10 个通道可以看作是一个 10 维空间,由于构造 10 维空间很困难,可以利用两个通道构成的二维空间进行研究,10 个通道共构成 45 个二维空间来表示这个 10 维空间。

图 2 是雾区可见光通道 1 和红外通道 3 的散点图,从图中散点的分布可以看出,左上角和右下角的空白处虽然都同时满足通道 1 和通道 3 阈值的上下限关系,但却不存在雾区点。这说明雾区点的值不仅要满足通道阈值条件,而且各通道之间还存在一定的特殊关系。对 10 个通道所构成的 45 组两两组合的散点图进行分析后,发现其中的 28 对组合存在比较明显的分布关系(图略)。在二维平面内用直线划分出大雾点所在区域,在前面单通道判识的基础上,再加上通道组合的判定来重新判识大雾区域。加上此限制条件后,可使大雾可能出现的范围缩小到轨道覆盖范围的 21.41%,如表 2 所示。

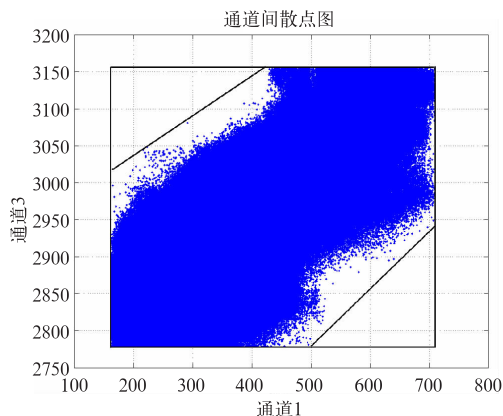


图 2 大雾区域通道值散点分布图

在粗判识留下的区域上,再利用支持向量机对大雾区进行细判识。由于对每一个点来说,粗判识的速度要远远高于支持向量机的判识速度,通过粗判识去掉了大部分非雾区域,只在可能出现大雾的区域进行支持向量机判识,大大提高了判识效率。

2.2 利用支持向量机对大雾进行细判识

下面重点介绍利用支持向量机对大雾进行细判识的过程。利用 SVM 来进行大雾判识主要分为四步,分别是:挑选样本、调整参数、建立模型和数据预测。首先是选取样本点,样本要尽量选得有代表性,从之前的结果来看(图略),粗判识已经很好地把陆地和海洋都去掉了,剩下一些不易区分的云区,可以在云上多取一些非雾区的样本,这样更有利于支持向量机进行有针对性的判识。

样本取好后要对样本数据进行归一化处理。首先求出所挑选的雾区样本和非雾区样本的均值:

$$\mu = \sum_{i=1}^n \frac{x_i}{n} \quad (4)$$

然后计算标准差:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2} \quad (5)$$

最后进行归一化处理:

$$\bar{x}_i = \frac{x_i - \mu}{\sigma} \quad (6)$$

其中 x_i 表示第 i 个雾区样本(非雾区样本)在此通道的值, n 表示雾区样本和非雾区样本的总和。

在前面提到,支持向量机训练中,需要调试两个参数。一个正则化参数 C ,一个是核函数参数 γ 。经过不同参数组合的交叉验证,可以找出一个交叉验证正确率最高的组合。得到最优参数之后,利用所生成的决策函数来判别预测样本的类别。

由于利用卫星资料判定某个点是否是雾只与这个点的 10 个通道值有关,选择哪几个通道进行 SVM 判识成为要解决的首要问题。虽然利用支持向量机进行判识的时候不会出现维数灾难,但是由于各通道间匹配存在偏差,有些通道信号质量存在问题,所以并不是通道越多判识就越准确;但是通道太少的时候,信息量不足,也不能进行准确判识。这里,对 1~10 个通道的所有组合进行支持向量机训练,以便观察不同通道数量和组合对于 SVM 判识大雾所造成的影响。训练参数 g 和 c 的范围都是从 10^{-3} 到 10^3 ,步长为 $10^{0.4}$,每一组训练, g 和 c 有 256 种组合,基本覆盖了可能出现最高 5 交叉正确率的区域,训练结果如表 3 所示。

从表 3 可以看出,除了 10 个通道都用的情况,出现最高 5 交叉正确率所用的通道组

合中没有使用第 3 通道,这与第 3 通道数据质量不好正好吻合,质量不好的通道数据对预测没有稳定的贡献。随着通道数量的增加,5 交叉最高正确率有逐渐上升的趋势(图 3),这是因为当通道数量较少时,通道数量的增加会带来信息量的增加,促使判识精度的提高。但当多于 6 个通道时,5 交叉正确率没有进一步的提升,而是趋于一个稳定的状态,这是由于多通道之间存在一定的线性关系和通道匹配误差造成的。从对所有通道的 TS 评分曲线(图 3)可以看出来,通道数量的增加,并没有使 TS 评分随之升高,这说明并不是通道数量越多判识越准确,应该选择能正确反应大雾特征的恰当通道组合来进行判识。从图 3 可以看出,使用 1、2、4、6、7、10 这 6 个通道组合的 5 交叉正确率和 TS 评分都很高。

表 3 支持向量机对风云 1D 卫星不同通道组合的训练结果

所用通道数	可能组合数	5 交叉最高正确率	最高正确率参数 g	最高正确率参数 c	最高 5 交叉正确率所用通道	TS 评分
1	10	69.8035%	0.0158	3.9811	5	38.78%
2	45	75.2834%	10.0000	0.6310	1/6	43.37%
3	120	80.8390%	3.9811	0.1000	1/6/10	57.48%
4	210	85.2230%	3.9811	0.6310	1/5/6/10	62.73%
5	252	88.0197%	0.6310	63.0957	2/4/7/9/10	56.07%
6	210	89.9849%	0.6310	10.0000	1/2/4/6/7/10	74.04%
7	120	90.4762%	0.6310	10.0000	2/4/6/7/8/9/10	70.17%
8	45	91.0431%	0.2512	158.4893	1/2/4/6/7/8/9/10	24.43%
9	10	90.8919%	0.6310	10.0000	1/2/4/5/6/7/8/9/10	24.44%
10	1	90.9297%	1.0000	2.5119	1~10	61.82%

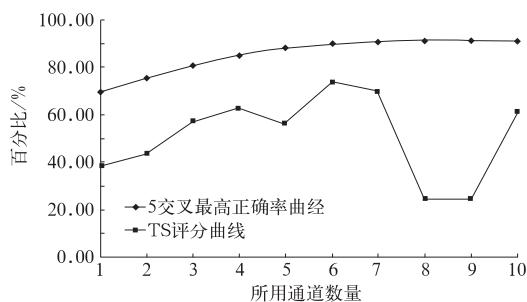


图 3 5 交叉正确率曲线和 TS 评分曲线图

通过支持向量机的细判识,已经达到了很好的效果。但在一些非雾区还存在一些零星的点,这些点可能是雾,也可能是错误判识造成的,由于只关注大片雾区所造成的影响,所以有必要去除这些点状雾区。

2.3 优化细判识后的图像

这里利用自定义的腐蚀和膨胀方法来处理细判识后的图像。标准的腐蚀是“收缩”或

“细化”二值图像中的对象,收缩的方式和程度由一个结构元素控制。简单的说,A 被 B 腐蚀是指 A 中完整包含 B 的区域被 B 的原点代替。这里所用的是根据大雾图像特征自定义的腐蚀方法。由于雾区没有固定的形态,所以用标准的结构元素来腐蚀图像达不到预期的效果。在雾区二值图像中,某点是雾的概率与其周围的环境有关,所以用概率腐蚀,即此点周围一定区域内雾区的比例越大,此点是雾的概率就越高。设定如果在某点周围 5×5 的方格范围内,非雾区点不多于 4 个,则此点被腐蚀后作为雾区保留。

腐蚀处理后,那些零星的雾区点被腐蚀掉了,但是大雾区域的范围也被腐蚀小了,所以通过膨胀来恢复大雾区域的范围。

膨胀是腐蚀的逆运算。标准膨胀是“加粗”或“变长”二值图像中对象的方法,膨胀的方式和程度由一个结构元素控制。简单的说,A 被 B 膨胀是指 A 中的原点被 B 的完整区域所代替。这里对标准的图像膨胀处理进行了改进。由于不是强调某种方向性的膨胀,所以在雾区二值图像膨胀时,不适宜使用结构元素来膨胀,应根据雾区腐蚀前的模版来膨胀,也可以看作是小区域内的图像重构。设定在腐蚀后雾区点 5×5 的方格范围内进行模版的膨胀,对于这个雾区点,将周围 5×5 的区域重构回腐蚀以前的情况。

2.4 个例分析

利用以上方法对 2005 年 8 月 25 日 7 点的风云 1D 轨道进行分析,结果如图 4(见彩页)所示。

图 4(a)是卫星轨道 6、2、1 三个通道的彩色合成图,从图中可以看出中间有部分是雾区,图右边有云和雾的混合区域。图 4(b)是通过通道阈值进行粗判识得到的结果,可以看出,经过简单的粗判识后,去除了大部分的晴空陆地,留下了雾区和一些无法识别的

云雾混合区域。图 4(c)是利用支持向量机的判识结果,可见,通过支持向量机的判识,可以很好地将图右边的雾从云雾混合区域中划分出来,但是对一些零星点存在少量误判和漏判的现象。图 4(d)是图像腐蚀膨胀后的结果,由图可见,大块的雾区又恢复到了之前的样子,而零星的雾区被腐蚀掉了。从原始图和结果图来看,经过上述三个步骤,可以很好地将大雾区域提取出来。

3 小结和讨论

基于支持向量机提出了一种卫星遥感数据对大雾进行客观判识的方法,首先通过对雾区样本不同通道的特征进行统计,划定每个通道的阈值范围,以此来粗判识大雾区域;然后利用支持向量机的方法,选择出最能反应大雾特征的通道组合来对雾区样本和非雾区样本进行训练,利用所生成的模板对可能的大雾区域进行精确判识;最后对判识后的雾区图像进行腐蚀和膨胀的处理,产生最后的大雾区域。

根据 FY1D 的 10 个通道的特性,对我国 2006 年 9—12 月的大雾区域进行分析,结果表明:大雾判识区域与专家标记结果吻合;利用 1、2、4、6、7、10 通道组合进行判识的结果最好,5 交叉正确率为 89.9849%,TS 评分为 74.04%。对个例的分析检验也说明基于支持向量机的遥感大雾判识是切实可行的。

致谢: 文章得到了国家卫星气象中心方宗义研究员的热心指导,特表示谢意!

参考文献

- [1] 李子华,仲良喜,俞香仁. 西南地区 and 长江下游雾的时空分布和物理结构[J]. 地理学报,1992,47(3): 242-251.
- [2] 黄玉生,李子华,许文荣,等. 西双版纳地区冬季辐射雾的初步研究[J]. 气象学报,1992,50(1):112-117.
- [3] 李子华,黄建平,周毓荃,等. 1996 年南京连续 5 天

- 浓雾的物理结构特征[J]. 气象学报, 1999, 57(5): 622-631.
- [4] 康志明, 尤红, 郭文华, 等. 2004 年冬季华北平原持续大雾天气的诊断分析[J]. 气象, 2005, 31(12): 51-56.
- [5] 毛冬艳, 杨贵名. 华北平原雾发生的气象条件[J]. 气象, 2006, 32(1): 78-83.
- [6] Estoque M A. A numerical model of the atmospheric boundary layer[J]. Geophys Res, 1963, 68: 1103-1113.
- [7] 张利民, 李子华. 重庆雾的二维非定常数值模拟[J]. 大气科学, 1993, 17(6): 750-755.
- [8] 石春娥, 杨军, 孙学金, 等. 重庆雾的三维数值模拟[J]. 南京气象学院学报, 1997, 20(3): 308-317.
- [9] 董剑希, 雷恒池, 胡朝霞, 等. 北京及其周边地区一次大雾的数值模拟及诊断分析[J]. 气候与环境研究, 2006, 11(2): 175-184.
- [10] 刘健, 许健民, 方宗义. 利用 NOAA 卫星的 AVHRR 资料试分析云和雾顶部粒子的尺度特征[J]. 应用气象学报, 1999, 10(1): 28-33.
- [11] 孙涵, 孙照渤, 李亚春. 雾的气象卫星遥感光谱特征[J]. 南京气象学院学报, 2004, 27(3): 289-301.
- [12] 陈林, 牛生杰, 仲凌志. MODIS 监测雾的方法及分析[J]. 南京气象学院学报, 2006, 29(4): 448-454.
- [13] Atkinson P M, Tatnall A R L. Neural networks in remote sensing[J]. INT J Remote Sensing, 1997, 18(4): 699-709.
- [14] 李朝峰, 屈颖歌, 夏德深, 等. BP 网络改进模型的性能对比研究[J]. 计算机工程与应用, 2003, 39(19): 120-132.
- [15] Vapnik V N. The Nature of Statistical Learning Theory. NY: SpringerVerlag, 1995. 314pp
- [16] 陈永义, 俞小鼎, 高学浩, 等. 处理非线性分类和回归问题的一种新方法(I)——支持向量机方法简介[J]. 应用气象学报, 2004, 15(3): 345-354.
- [17] 李国正, 王猛, 曾华军译. 《支持向量机导论》, Nello Cristianini 和 John Shawe-Taylor 著[M]. 北京: 电子工业出版社, 2004: 163.

刘年庆等：基于支持向量机的遥感大雾判别

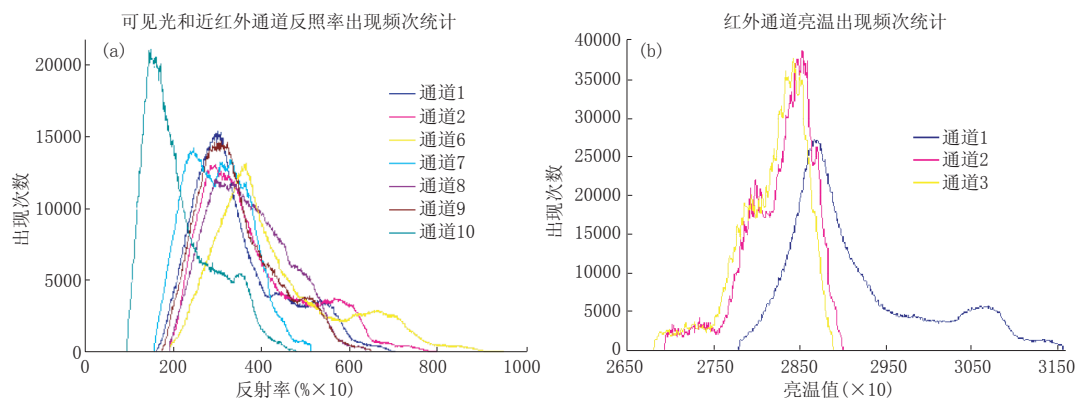
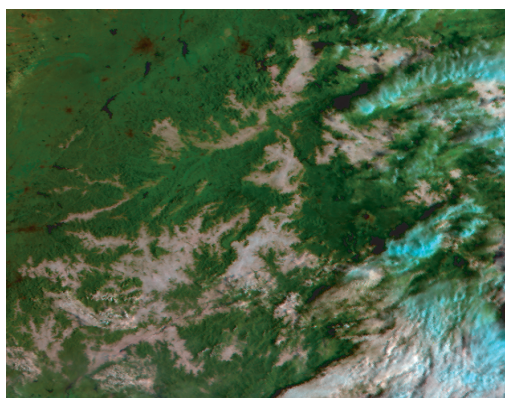
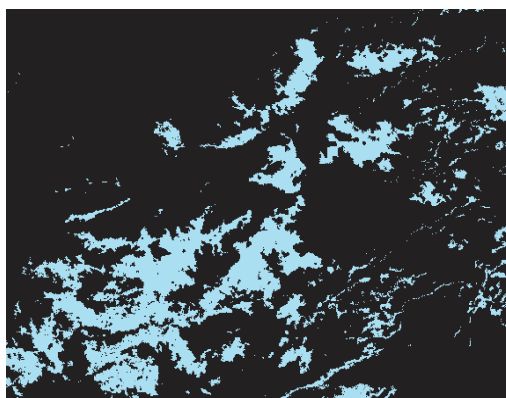


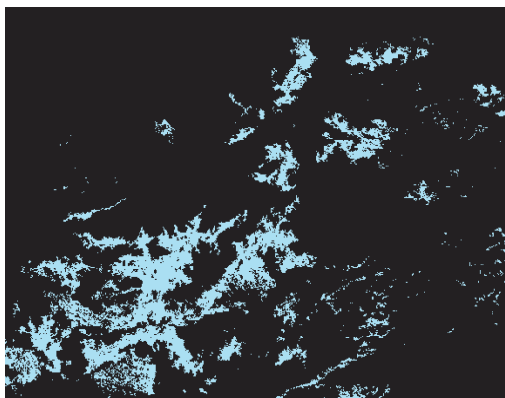
图1 可见光和近红外通道反射率(a)和红外通道亮度温度(b)出现频次统计图



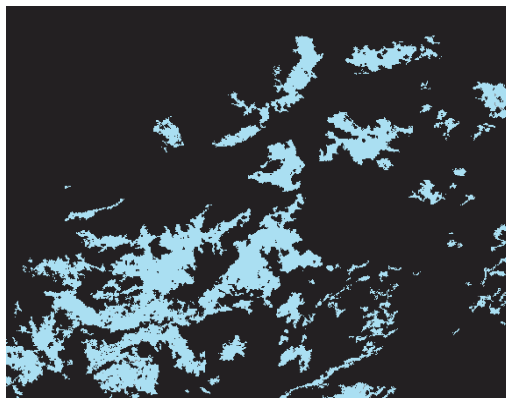
(a) 原始彩色合成图



(b) 粗判识结果示意图



(c) 支持向量机判别结果示意图



(d) 腐蚀膨胀处理结果示意图

图4 轨道图像判识过程图