

Cloud Cover Estimation Based on Image Segmentation

Yongqiang Zhan, Lijia Zhou, Guanlei Xu, Xianpeng Liu

Department of Military Oceanography, PLA Dalian Naval Academy, Dalian
Email: zyq0214@tom.com

Received: Apr. 15th, 2014; revised: Apr. 17th, 2014; accepted: Apr. 23rd, 2014

Copyright © 2014 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Cloud is one of the important meteorological elements that could be perceived by eyes. Cloud observation could help us make weather forecast. Cloud cover estimation is part of cloud observation. The paper uses the method of threshold segmentation of digital image processing to estimate cloud amount. Compared with traditional artificial and ordinary instrument method, the measurement is more convenient and practical. The paper analyzes the result of three methods of image segmentation which are Otsu, max entropic thresholding and Mathematics expectation, and puts forward a modified method. The modified method works better on cloud image which shows higher cloud.

Keywords

Cloud Image Based on Ground, Image Segmentation, Saturation, Otsu, Max Entropic Thresholding, Mathematics Expectation

基于图像分割的云量估计

战勇强, 周立佳, 徐冠雷, 刘现鹏

海军大连舰艇学院军事海洋系, 大连
Email: zyq0214@tom.com

收稿日期: 2014年4月15日; 修回日期: 2014年4月17日; 录用日期: 2014年4月23日

摘要

云是天空中人类可以直接通过肉眼感知到的主要气象要素之一，云的探测能够帮助人们识别阴晴风雨，预知天气变化。云量估计是云探测组成部分。本文主要采用数字图像处理技术中阈值分割方法估计云量，相比传统人工目测方法和一般器测方法具有便利性和实用性。本文利用地基的可见光范围的云图，在饱和度空间运用分割算法估计云量。对比分析了最大类间方差法、最大熵法、基于数学期望分割法的分割效果，并提出了改进算法。该方法在云底高度较高的云类获得的分割效果更符合云量估计标准。

关键词

地基云图，图像分割，饱和度，最大类间方差法，最大熵法，数学期望法

1. 引言

云是指停留在大气层中的水滴或冰晶的集合体，是地球水汽循环的重要因子，是影响地球大气辐射平衡、热量平衡的重要因素[1]。云和天气的关系十分密切，云的特征常常可以直接反映天气的变化。云的变化作为预报天气的重要参考。云的外形、数量、高度等信息是大气环境复杂变化的一种表现。云量探测研究领域，根据所用光谱波段不同，主要可以划分为红外波段测云和可见光波段测云两个方向。红外波段测云的主要思想是将云看作为比较强的红外发射体，选择适当的设备通过探测这一部分红外辐射，结合亮温数据，来作为识别云层的主要方法[2]。对于可见光波段云量的计算，主要的依据是大气分子和云粒子对可见光的散射关系不同所造成颜色变化，再从区分颜色角度进行云量估计。

Yamashita 等[3]提出了用天空指数来表述获得的全天空图像，在 RGB 彩色空间中，将此指数定义为 $(B-R)/(B+R)$ 。经统计选定统计值作为阈值分辨晴空和云点，这一方法容易误判天边象元。其主要误差来自薄云和气溶胶的存在。霍娟[4]等人利用全天空数字相机观测云量，选取蓝色通道值和亮度的比值作为判断云的阈值标准。在能见度大于 15 km 时效果较好，对于卷云类的薄云难以分辨。Souza-Echer[5]在利用视张角为 72° 的测云仪器观测南极地区上空云时，将 RGB 图像转换到 HIS 空间，然后，计算图像饱和度并作为云点识别的因子，采用平行六面体分类法进行云、晴空和不确定点 3 类像元的识别。与人工目测分析结果相比，此方法对晴空的估计偏高。

在地基云图计算云量研究中，还鲜有人从云图饱和度直方图入手通过阈值分割来估计云量。本文利用地基相机拍摄的云图，在 HIS 彩色空间[6]的饱和度分量中利用阈值分割进行云量估计的，本文对比分析了最大类间方差法(Otsu)[7]、最大熵法[8]、基于数学期望法[9]，并提出的改进算法进行云图分割实验效果。寻找适合某一类型云的分割方法进行云量的估计。实验表明，使用该种方法同人工识别云的结果一致性较好。

2. 云量观测与饱和度通道选取

云的观测包括总云量、低云量、云状、低云高等要素。其中云量以天空被云遮挡的成数表示，用十分法估计。全天乌云或有云但不到天空的 $1/20$ ，记为“0”；占全天空的 $1/10$ ，记为“1”，云占全天空的 $2/10$ ，记为“2”；其余依次类推，全天为云遮盖无缝隙，记为“10”，由少量缝隙可见蓝天，则记为“ $\square 10$ ”，对于观测时有雾，天顶不可辨的情况，云量记为“10” [10]。目前地面对云的定量测量，特别

是非降水云特征的观测仍是弱项,对云量、云状的观测,还是以人工观测为主。在气象日常工作中,目测是唯一估计云量的方法[11]。

云分类中,根据云的底部高度把云分为低、中、高三种,本文主要感兴趣的是中、高云系,并且关注云的外貌颜色与其类别的关系。一般情况下,由于云粒子和大气分子对光散射关系不同,在有云的天空中,尤其是高云,云的颜色较浅,呈现为白色,天空颜色较深,呈现为蓝色。在相应的饱和度通道中,云的饱和度一般要高于天空。由于在 HSI 空间中,色调、饱和度以及亮度三个通道的量值是相对独立的,应用饱和度通道可以降低色调、光照强度的影响,处理起来更加合理。

3. 分割算法

阈值分割算法核心技术是阈值确定,阈值分割算法都是基于图像的灰度直方图。将图像像素的灰度值与该阈值进行比较,根据比较的结果将像素进行分类,从而达到分割的目的。本文选用最大类间方差法、最大熵方法和基于灰度数学期望的方法。这些方法对于地基云图中云的识别效果因云的性质而存在差异,表现出不同的分割效果。

3.1. 最大类间方差法

最大类间方差法是一种使一副图像中目标和背景间的方法最大化而确定分割阈值的方法。此方法应用于云的分割步骤就是选取一个阈值 T ,分别计算预分割的云和天空两部分的类内方差和类间方差,然后通过选择 T 使得类间方差和总方差的比值最大,进而确定最佳阈值 T 。

设图像像素点的数目为 N ,灰度值范围为 $[0, L-1]$,对应某个灰度级 i 的像素点数目为 N_i ,则整幅图像的像素点数目为: $N = N_0 + N_1 + \dots + N_{L-1}$,不同灰度级上像素出现的概率可以表示为: $P_i = N_i/N$, $i = 0, 1, 2, \dots, L-1$,这里, $\sum_{i=0}^{L-1} P_i = 1$ 。

云图被阈值 T 分为云(表示为 C)和天空(表示为 S),云的饱和度值普遍大于天空,其相应的灰度值同样普遍大于天空,因此 C 类包含着 $[0, T]$ 范围的像素,而 S 则包含 $[T+1, L-1]$ 范围的像素。

整幅图像的均值可以表示为:

$$\mu_T = \sum_{i=0}^{L-1} iP_i \quad (1)$$

云类和天空类的均值 μ 和出现的概率 ω 可以由下式给出:

$$\omega_c = P(C) = \sum_{i=0}^T P_i = \omega(T) \quad (2)$$

$$\mu_c = \sum_{i=0}^T iP_i / \omega_c = \mu(T) / \omega(T) \quad (3)$$

$$\omega_s = P(S) = \sum_{i=T+1}^{L-1} P_i = 1 - \omega(T) \quad (4)$$

$$\mu_s = \sum_{i=T+1}^{L-1} iP_i / \omega_s = \frac{\mu_T - \mu(T)}{1 - \omega(T)} \quad (5)$$

其中 $\mu(T) = \sum_{i=0}^T p_i$ 。不难看出, $\mu_T = \omega_c \mu_c + \omega_s \mu_s$,而且 $\omega_c + \omega_s = 1$,在 T 取 $[0, L-1]$ 范围内任意值都成立。而云类和天空类的方差由下式给出:

$$\begin{cases} \sigma_c^2 = \sum_{i=1}^T (i - \mu_c)^2 P_i / \omega_c \\ \sigma_s^2 = \sum_{i=T+1}^{L-1} (i - \mu_s)^2 P_i / \omega_s \end{cases} \quad (6)$$

下面定义类内方差为 σ_1^2 。类间方差为 σ_2^2 、总体方差为 σ_a^2

$$\sigma_1^2 = \sigma_c^2 \omega_c + \sigma_s^2 \omega_s \quad (7)$$

$$\sigma_2^2 = \omega_c (\mu_c - \mu_T)^2 + \omega_s (\mu_s - \mu_T)^2 = \omega_c \omega_s (\mu_c - \mu_s)^2 \quad (8)$$

$$\sigma_a^2 = \sigma_1^2 + \sigma_2^2 \quad (9)$$

此时引入 $Y(T) = \sigma_s^2 / \sigma_c^2$ ，当 $Y(T)$ 最大值时，所得的 T 就是最佳阈值。

3.2. 最大熵方法

熵作为信息论的重要概念，常用来描述体系混乱程度，能量分布的越均匀，熵就越大，学者 Pun[5] 使用关于香农熵的这一概念，在用图像的灰度直方图表示图像的基础上，对图像的熵作了定义，并以此用来分割图像。熵的应用在云图分割中的最大熵方法的思想是将图像中的云和天空分成两类，选择适当的阈值进行分割，使得其每一类的总熵最大。

对于一副云图，像素点个数与出现概率的定义同 3.1 中的方法。在云类和天空类的两个部分之中，其概率分布分别为：

云区域： P_i / P_T ，其中 $i = 0, 1, 2, \dots, T$

天空区域： $P_i / (1 - P_T)$ ，其中 $i = T + 1, T + 2, L - 1$

其中 $P_T = \sum_{i=0}^T P_i$ 。

定义云熵和天空熵如下式：

$$H_c(t) = - \sum_{i=0}^T \frac{P_i}{P_T} \lg \frac{P_i}{P_T} \quad (10)$$

$$H_s(t) = - \sum_{i=T+1}^{L-1} \frac{P_i}{1 - P_T} \lg \frac{P_i}{1 - P_T} \quad (11)$$

那么一副云图的总熵表示为：

$$H(t) = H_c(t) + H_s(t) \quad (12)$$

当 $H(t)$ 取最大值时，对应的 t 就是最大熵阈值法所获得的最佳阈值。

3.3. 基于数学期望的分割方法及改进算法

对于经典的最大类间方差方法和最大熵方法对云图的分割都达到了不错的效果，有一定的普适性。同样从云图饱和度通道的灰度直方图入手，理想情况下的基于直方图的分割应该有显著的“峰”和“谷”，而除非恶劣天气条件下，一般的地基云图表现的颜色比较单纯而且色调相近，尤其是云底高度比较高的情况，例如卷云、高积云等，它们一般形成在海拔比较高的地区，也产生在海拔低但是寒冷地区，高云大都是由细小且稀疏的冰晶构成，故比较薄而且具有较好的透光性，云的边缘色彩也相对其他云较淡。

最大类间差方法和最大熵方法对此类云的获得量普遍偏少。这两类方法虽然可以作为云的识别用，但是不能很好的反映图像的几何结构，特别对于薄云容易漏检。针对地基云图的灰度直方图没有明显多

峰值的几何结构，我们引入几何重心的概念，在云图的灰度直方图中，灰度级为 $[0, L-1]$ 对于预设的阈值为 T 将直方图分成了两部分，分别是云类 C 和天空类 S ，两类中不同灰度级的元素 i ，其中 $i \in [0, L-1]$ ，到达重心 T 的矩可以分别表示为：

$$|T-i|p(i), \text{其中 } i = 0, 1, 2, \dots, L-1 \quad (13)$$

$p(i)$ 表示灰度值为 i 的像素点的个数。当云类和天空类相对于预选阈值 T 的矩相等时，如式(13)所示，我们即认为找到了表征图像几何形状的“重心”，由重心所分割的两部分相对于重心达到平衡，即认为获得了满意的分割效果。

$$\sum_{i=0}^T (T-i)p(i) = \sum_{i=T+1}^{L-1} (i-T)p(i) \quad (14)$$

可得：

$$T = \frac{\sum_{i=0}^{L-1} ip(i)}{\sum_{i=0}^{L-1} p(i)} \quad (15)$$

从另外方差最大化的角度，也同样可以解释该方法：同样的定义，图像的方差可以表示为：

$$\sigma^2(i) = \sum_{i=0}^{L-1} (i-T)^2 p(i) \quad (16)$$

当 $\sigma^2(i)$ 取得极大值时，对式(16)求导数，同样可以得到式(14)的结果。

在实际云图分割实验中，对于卷云、高积云等颜色比较淡，边缘相对模糊的地基云图，该方法优于最大类间方差法和最大熵方法，云量判别多于前两种方法，符合视觉对云量估计的标准。但是仍然不能满足我们云量判别的要求，在基于数学期望的基础上，提出了改进，由于每个灰度值同预选阈值的距离 $|T-i|$ 表征该像素灰度同阈值的差异程度，我们希望通过调整它的权重，放大灰度差引起的效应，获得来获得更好的分割效果。定义新的矩：

$$\exp(|T-i|) \cdot p(i), \text{其中 } i = 0, 1, 2, \dots, L-1 \quad (17)$$

即要求：

$$\sum_{i=0}^T \exp(T-i) \cdot p(i) = \sum_{i=T+1}^{L-1} \exp(i-T) \cdot p(i) \quad (18)$$

虽然此式的解析解不容易表答出，但是计算机可以采用三点快速收缩算法通过计算式(19)来快速获取阈值。修正后的分割方法相对原方法对云图中薄云、淡云的分割有更好的效果。

$$\min \left| \sum_{i=0}^T \exp(T-i) \cdot p(i) - \sum_{i=T+1}^{L-1} \exp(i-T) \cdot p(i) \right| \quad (19)$$

4. 实验验证与分析

使用本文提到的几种方法对云图进行验证。地基云图选自于《中国云图》中具有代表意义的个例：

实验一：图 1~图 3

实验二：图 4~图 6

从两组实验中分别选取了具有代表性的避光层积云和高积云为例，并分别图(a)、(e)是最大类间方差法分割图，图(b)、(f)是最大熵方法分割图，图(c)、(g)是基于数学期望方法分割图，图(d)、(h)是本文改进的方法。通过对比四种方法和人工目测结果(如表 1)可以看出，四种方法都已经大致将云类和天空类分



Figure 1. Stratocumulus opacus
图 1. 避光层积云

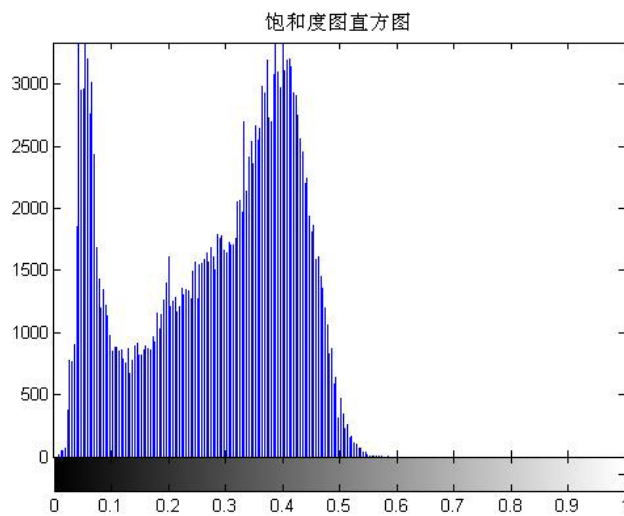
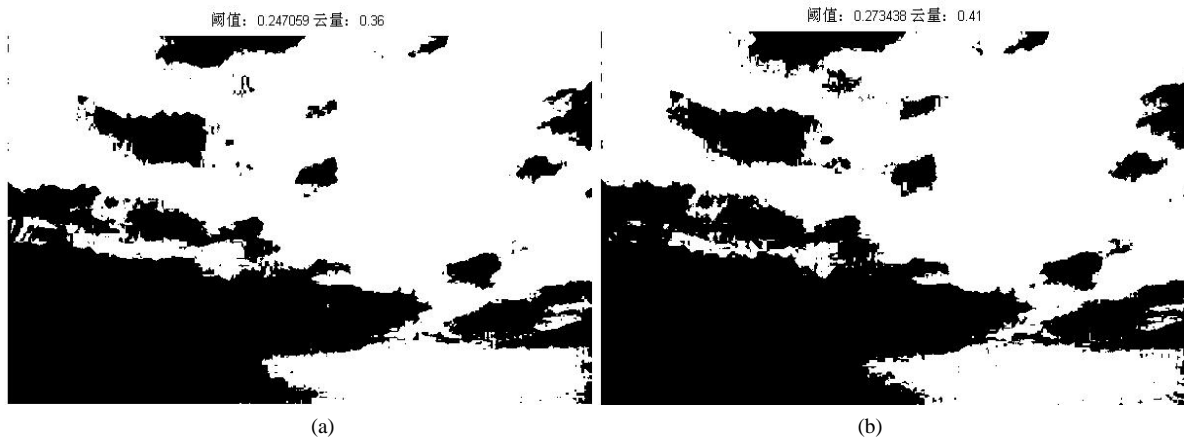


Figure 2. Histogram of saturation of the cloud image
图 2. 云图饱和度通道直方图



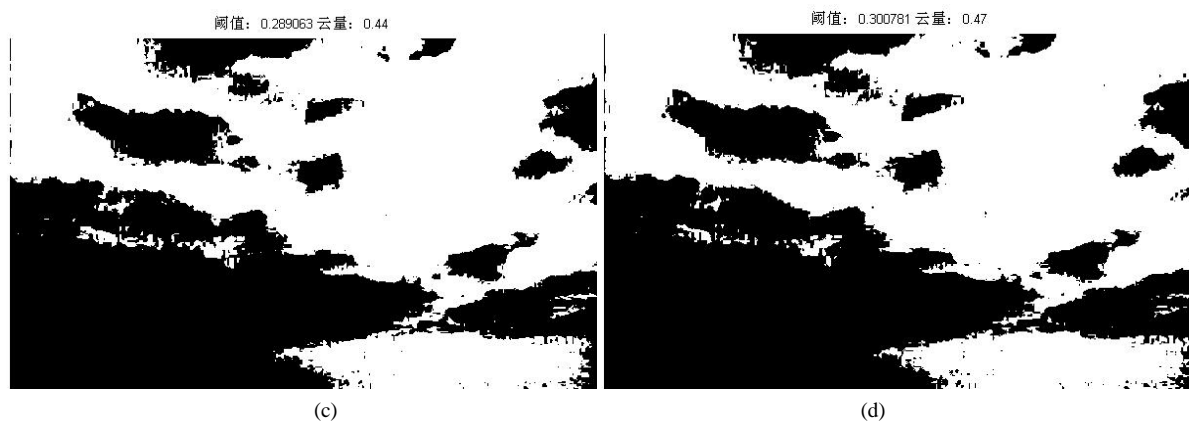


Figure 3. Impact figure of four methods of segmentation
图 3. 四种方法的分割效果图



Figure 4. Altocumulus
图 4. 高积云

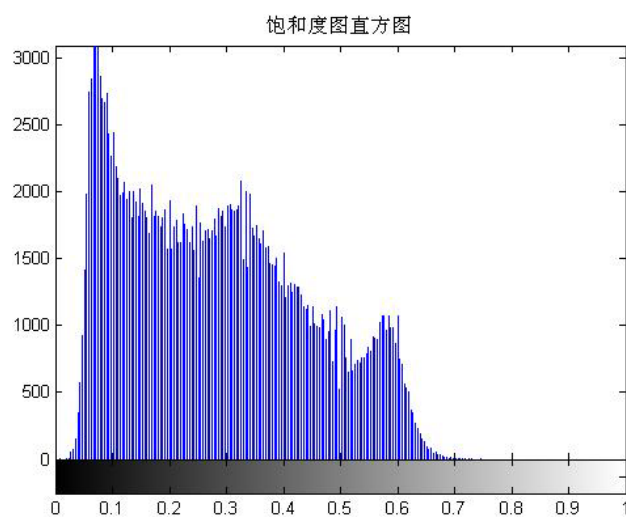


Figure 5. Histogram of saturation of the cloud image
图 5. 云图饱和度通道直方图

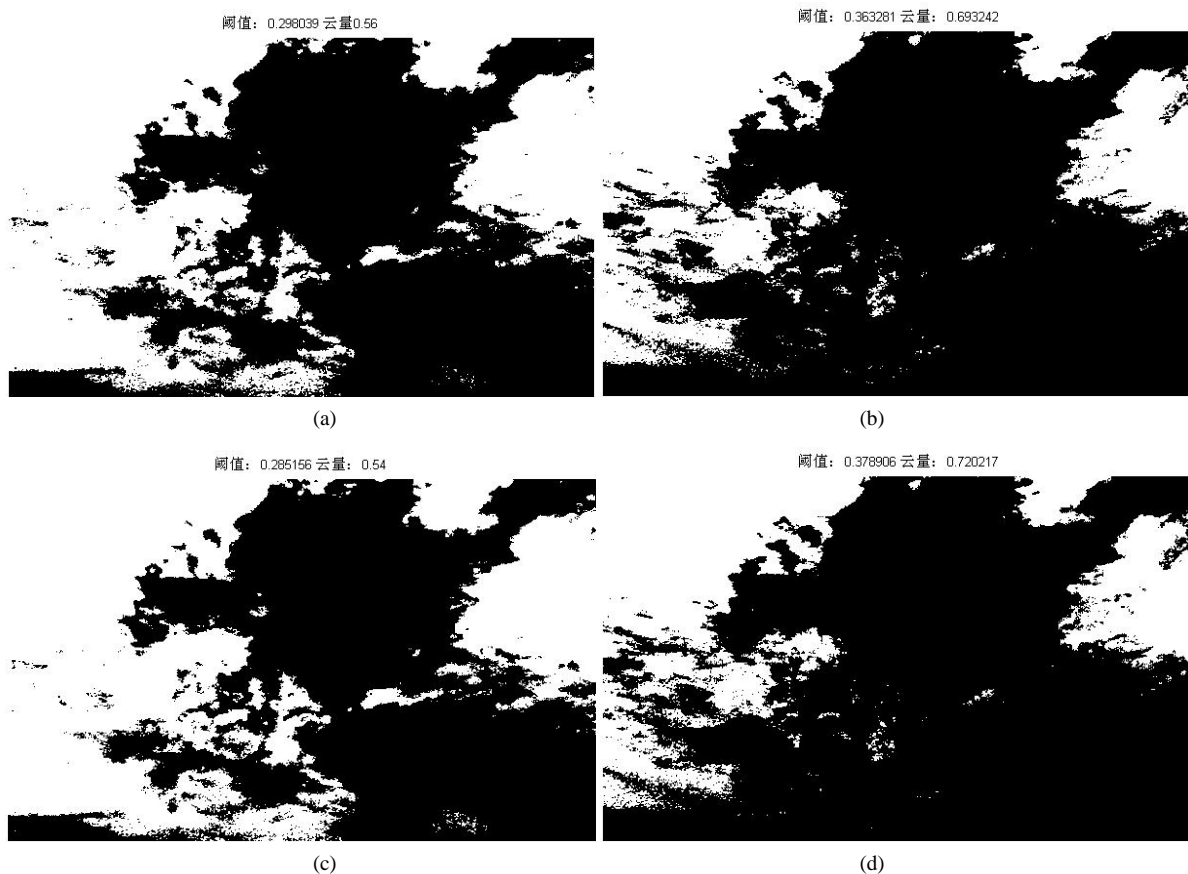


Figure 6. Impact figure of four methods of segmentation
图 6. 四种方法阈值分割效果图

Table 1. Comparison of different measurements
表 1. 不同方法对比结果

个例 \ 方法	最大熵法	最大类间方差法	数学期望法	改进算法	人工目测
避光层积云	4.1	2.5	2.9	4.7	5.3
高积云	6.9	3.0	5.4	7.2	7.3

割出来，最大类间方差法、最大熵方法和基于数学期望方法对于云图中白亮的主成分分割效果好，但是对于云与天空过渡的边缘部分明显偏少。改进方法的结果最符合观测实际。单纯从图像分割角度，前三种方法的效果都做到了突出图像中的重点区域，即亮而白的云区域，但是在云量的实际观测中，消散的颜色淡的云同样应归于云类，因此从云量观测的角度来讲，改进的方法获得的分割图更符合判别云量的标准。

5. 结语

云量的观测是气象观测的重要组成部分。本文利用地基云图资料，从饱和度通道的直方图入手，选取不同的分割算法进行云图的阈值分割并比较算法之间的优劣。另外从图形几何角度重新定义了期望算法并提出改进。改进的算法相对于其他算法在部分高云云图上获得了比较好的分割效果，对于薄云、淡云的分割更具优势，同作为标准的人工目测结果最为接近。利用地基云图结合图像处理方法来观测云量

的相对于目前舰艇保障普遍使用的人工目测方法,大大提高了效率,仅依靠摄像机和计算机以及配套算法就可以实现白天云量的连续观测。方法的不足之处在于观测目标局限性,雾霾、雷暴等恶劣天气中的观测效果不佳,同时继续改进以获得更具普适性的算法。

基金项目

国家自然科学基金(6125006)资助课题。

参考文献 (References)

- [1] 徐兴奎 (2012) 中国区域总云量和地云量分布变化. *气象*, **1**, 90-95.
- [2] 胡树贞, 马舒庆, 陶法, 等 (2013) 基于红外实时阈值的全天空云量观测. *应用气象学报*, **2**, 179-188.
- [3] Yamashita, M., Yoshimura, M. and Nakashizuka, T. (2004) Cloud cover estimation using multitemporal hemisphere imageries. *Proceedings of 7th Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS04)*, Istanbul, 818-821.
- [4] 霍娟, 吕达仁 (2002) 全天空数字相机观测云量的初步研究. *南京气象学院学报*, **2**, 242-246.
- [5] Souza-Echer, M.P., Pereir, A.E.B., Bins, L.S., et al. (2006) A simple method for the assessment of the cloud cover state in high latitude regions by a ground-based digital camera. *Journal of Atmospheric and Oceanic Technology*, **23**, 437-447.
- [6] Gonzalez, R.C. (2012) *Digital image processing*. Publishing House of Electronics Industry, Beijing.
- [7] Otsu, N. (1979) A threshold selection method from gray-level histogram. *IEEE Transactions on Systems, Man and Cybernetics*, **9**, 62-66.
- [8] Pun, T. (1980) A new method for gray-level picture thresholding using the entropy of the histogram. *Signal Processing*, **2**, 223-237.
- [9] 尹沧涛, 何平, 等 (2012) 基于图像序列期望特征的阈值分割方法. *计算机应用研究*, **1**, 376-378.
- [10] 侍茂崇, 高郭平, 鲍献文 (2008) *海洋调查方法导论*. 中国海洋大学出版社, 青岛.
- [11] 国家气象局气候监测应用管理司 (1995) *气象仪器和观测方法指南*. 气象出版社, 北京.