

文章编号: 1007 4619(2006)03 0415 06

一种基于区域生长原理的油库目标聚类定位方法

蔡红苹, 蒋咏梅, 粟毅

(国防科学技术大学 电子科学与工程学院, 湖南 长沙 410073)

摘要: 利用油罐目标相对集中而虚假目标相对较离散的分布特点, 提出了一种基于区域生长原理的聚类方法。对于实际的遥感光学图像, 所提方法既能将油罐目标聚类定位油库区域, 又能自动去除虚假油罐目标。实验表明, 该方法在油库检测定位上准确率高, 是一种实用有效、运算量小的聚类方法。

关键词: 区域生长; 聚类; 油库检测; 遥感图像

中图分类号: TP751.1/TP79 **文献标识码:** A

A Clustering Method for Localizing Oil Houses Based on Region-growing

CAI Hong-ping JIANG Yong-mei SU Yi

(School of Electronic Science and Engineering, National University of Defense Technology, Hunan, Changsha 410073, China)

Abstract Oil storage tanks are usually distributed closely in groups, while the false targets generally distract from each other. A region-growing clustering method is proposed in this paper, which can not only localize oil tanks but also remove the false oil tank targets automatically. The experimental results with real remote sensing images indicate that this method can localize oil house accurately, effectively, and at meantime is a practical and fast clustering method.

Key words region-growing; clustering; oil house detection; remote sensing image

1 引言

油库及类似目标检测定位, 在城市规划和军事上有着重要的意义。油库一般是成片的油罐目标群, 油罐直径范围 3—80m, 油罐高度范围 2—30m。在光学遥感图像中, 油罐的识别特征十分明显, 通常

呈现颜色均一的椭圆形状, 是油库检测定位最可靠的依据。因此, 在实现油库的自动检测定位时, 可以采用的方法是先检测出椭圆, 然后根据油库内油罐相对集中分布的先验知识, 将椭圆进行聚类, 从而自动定位油库区域, 检测步骤如图 1 所示。本文研究的内容主要是如何从提取出的椭圆信息中进行油库提取, 其步骤如图 1 中画虚线的部分所示。

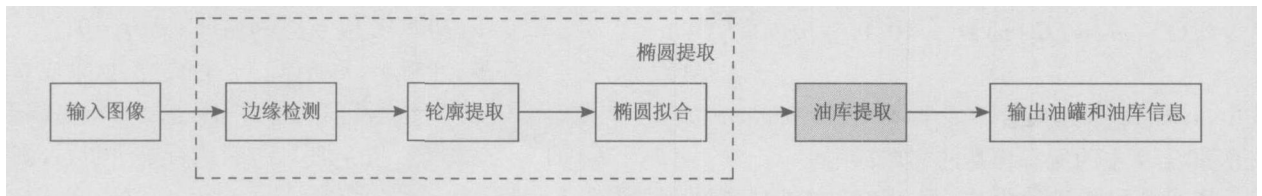


图 1 油库检测定位步骤图

Fig. 1 Flow chart for detecting and localizing oil houses

收稿日期: 2004-11-01; 修订日期: 2005-01-27

作者简介: 蔡红苹(1980—), 女, 分别于 2001 年和 2004 年在国防科技大学理学院获计算数学专业学士和硕士学位, 现为该校电子科学与工程学院在读博士, 主要从事遥感图像目标识别、融合算法研究。E-mail: hongpingca8014@yahoo.com.cn

目前成熟的聚类算法(如 C 均值算法^[1]和 FODATA 算法等^[2])可以将不同的油库分离开,然而存在的一个问题是,在椭圆检测过程中,会把一些不是油罐的椭圆目标也检测出来。并且由于虚假目标的存在会使油库定位不准确,并使油库区域扩大。然而现有的聚类算法通常没有考虑到自动去除虚假目标的问题。为此,本文提出了一种在聚类时将虚假油罐目标自动去除的聚类方法,这种方法引用了区域生长的图像分割算法思想。通过实验表明,该方法在油库检测定位上准确率高。

2 基于区域生长原理的聚类方法

区域生长是一种图像分割算法,其基本原理是将具有相似性质的像素集合起来构成区域。具体地讲,是先对每个需要分割的区域找一个种子像素作为生长的起点,然后将种子像素周围邻域中与种子像素有相同或相似性质的像素合并到种子像素所在的区域中。将这些新像素当作新的种子像素继续进行上面的过程,直到再没有满足条件的像素被包括进来为止^[3,4]。

在由检测出的椭圆定位油库的过程中,引用了区域生长原理,把椭圆看成是图像分割中的像素,目的就是要把聚集在一起的椭圆通过区域生长聚成一类。首先将每个椭圆称作是样本,其位置是椭圆中心点的位置,为了简化算法,不考虑椭圆的大小。利用区域生长原理将具有相邻特征的样本集合起来构成类,其生长准则是以 r_0 为半径的邻域。具体做法是将一个种子样本 i_0 作为生长的起点,然后将以种子样本 i_0 为中心、 r_0 为半径的邻域样本集 B_{i_0} 合并到一类 C_k , 将 C_k 中的所有样本(即 $C_k(p) \in C_k, p=0, 1, \dots, num(C_k) - 1$)当作新的种子样本,寻找所有样本 $C_k(p)$ 的 r_0 半径的邻域,将这些邻域点合并到 C_k 类,即

$$C_k = C_k \cup Neighbor[C_k(p)] \quad p=0, 1, \dots, num(C_k) - 1 \quad (1)$$

式中, $num(D)$ 表示集合 D 的样本个数, $Neighbor[D]$ 表示集合 D 中的所有样本的邻域样本集。

通过此过程继续生长,直到再没有满足条件的样本归入该类。在还未聚类的样本集 ϕ 中,寻找一个新的种子样本重复上面的步骤,直到所有的样本都被归入某类为止。如图 2 所示,样本点间的连线表示两样本是关于 r_0 半径相邻。因为油罐通常聚集在一起,分布相对较集中,而虚假目标分布相对较

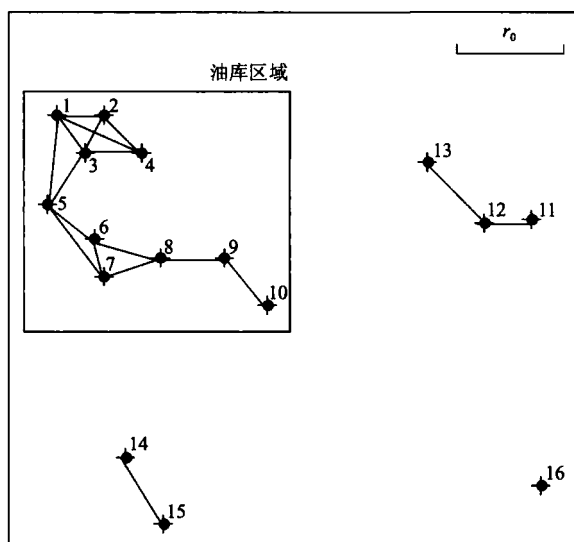


图 2 基于区域生长原理的聚类方法示意图

Fig 2 The clustering method based on region growing

分散,所以设置类内最少样本数目 N_0 , 椭圆数目小于 N_0 的类被认为是虚假目标去除掉,从而把样本数大于 N_0 的类标定为油库区域。

基于区域生长原理的聚类算法步骤描述如下(如图 3 所示):

第 1 步,输入样本集 E , 初始化变量:类内最少样本数目 N_0 , 未分类样本集 $\phi = E$, 类别 $k=0$

第 2 步,计算每两个样本间的距离 $d_{ij} (i, j=1, 2, \dots, N)$, 也就是计算距离矩阵 D ;

第 3 步,比较距离矩阵第 i 列的值和邻域半径 r_0 的大小,与第 i 个样本距离小于 r_0 的样本就属于第 i 个样本的邻域集 $B_i (i \in E)$;

第 4 步,寻找具有最多邻域点的样本: $i_0 = \arg \max_i \{ num(B_i), i \in \phi \}$;

第 5 步,判断 $num(B_{i_0}) < 2$ 或者 $num(\phi) < N_0$ 是否成立,如果成立,则算法结束;否则,转入下一步;

第 6 步,初始化 $C_k = B_{i_0}$ 和循环变量 $p=0$

第 7 步,判断 $p < num(C_k)$ 是否成立,如果成立,则 $C_k = C_k \cup Neighbor[C_k(p)]$ 并将增加的样本添至 C_k 的末尾,循环变量 p 加 1,重新执行第 7 步;否则,转下一步;

第 8 步,从未分类样本集去除第 k 类样本,即 $\phi = \phi - C_k$;

第 9 步,判断 $num(C_k) < N_0$ 是否成立,如果成立,则转至第 4 步;否则, k 加 1 后转至第 4 步,同时输出第 k 类样本集 C_k 。

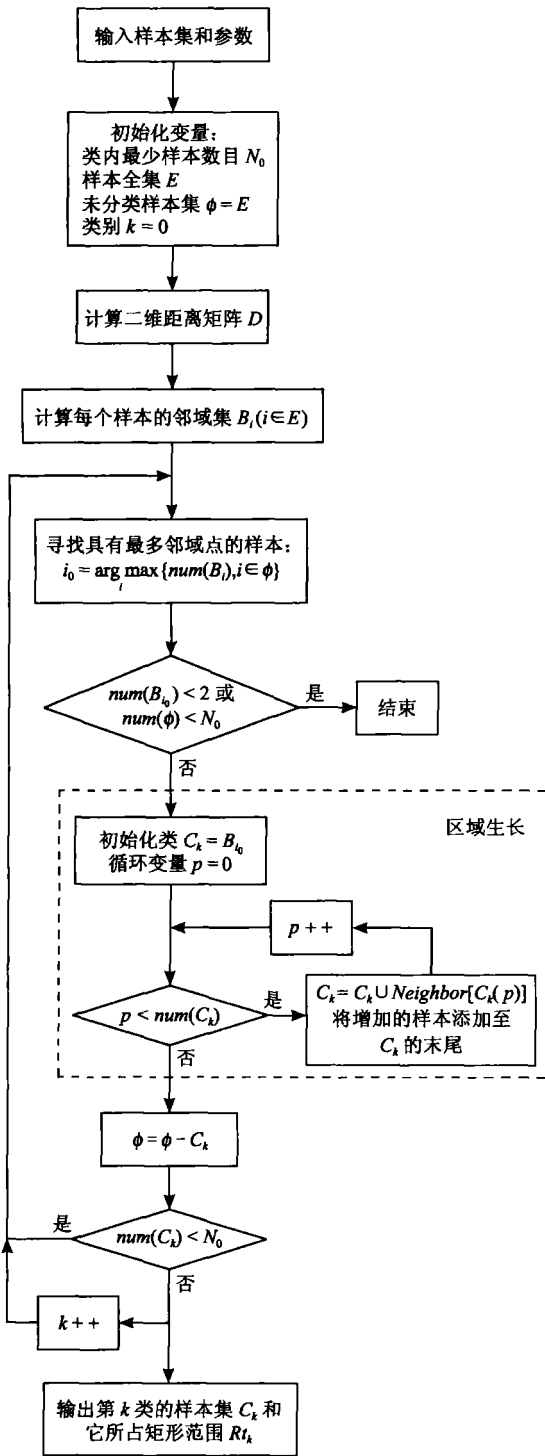


图 3 基于区域生长原理的聚类方法流程图
Fig. 3 Flow chart of the clustering method based on region growing

算法能否有效地聚类油库且去除虚假目标, 与邻域半径 r_0 的大小选择有很大关系, 我们选择 r_0 的原则是满足以下几个条件:

- (1)使同一个油库内的油罐目标至少能与两个

- 油罐(包括自身)相邻;
- (2)使不同的油库不相邻;
- (3)油库和虚假目标不相邻;
- (4)虚假目标构成的类的样本数目少于设置的类内最少样本数目。

由于事先并不知道图像内油罐之间的距离, 更不知道虚假目标的信息, 所以 r_0 的选择只能是根据经验来定, 在油库检测中 r_0 设定为与分辨率成反比:

$$r_0 = \frac{C}{p} \quad (2)$$

式中, C 是用户设置的常数, 通常是 300—600 间的整数, p 表示图像分辨率。

3 计算量分析

该聚类方法主要的计算量集中在寻找每个样本点的半径为 r_0 邻域中, 也就是计算每两个样本之间的距离, 然后与 r_0 比较。没有必要每计算一个样本的邻域时就计算一次它与所有其他样本的距离, 这样会有大量重复计算, 每两个样本间的距离只计算一次即可, 也就是在寻找邻域之前只需先计算一个二维距离矩阵 D :

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \dots & \dots & \dots & \dots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{bmatrix} \quad (3)$$

式中, n 表示样本总数。该矩阵是一个对角线为 0 的对称矩阵, 因此只用计算上三角部分元素的值。这里的距离采用欧氏距离, 由于求距离的目的只是为了与 r_0 比较大小, 所以可以用距离的平方与 r_0^2 比较即可, 从而增加一次乘法运算, 避免了 $n(n-1)/2$ 次开方运算, 因此总的计算量为: $n(n-1)+1$ 次乘法、 $\frac{3n(n-1)}{2}$ 次加法和若干次求最大值运算(最坏的情况下是 $n-1$ 次)。

该聚类方法没有迭代运算, 是一种计算量小的方法, 从后面的实验结果可以看出, 该聚类方法的运算时间基本可以忽略不计。

4 实验结果

表 1 列出了使用本文提出的聚类方法对 7 幅不同的含油库遥感图像的油库检测定位实验结果。

表 1 基于区域生长原理的聚类方法实验结果

Table 1 Experimental results of the clustering method based on region growing

图像名	01. bmp	02. bmp	03. bmp	04. bmp	05. bmp	06. bmp	07. bmp
图像大小	971×1832	1945×1542	1858×1306	840×582	800×755	1363×1231	1858×1306
聚类时间	0.0	0.0	0.0	0.0	0.0	0.0	0.0
椭圆数目	141	200	106	77	74	39	131
油罐数目	95	157	91	70	71	6	108
油库数目	6	4	5	2	2	1	5
虚警油库数目	1	0	0	0	0	0	1
漏检油库数目	0	0	0	0	0	0	1
备注	2 个油库位置略有偏差, 主要原因在于椭圆提取步骤	检测基本准确	检测基本准确	一小片漏检, 主要原因在于前面的椭圆提取步骤	一小片漏检, 主要原因在于前面的椭圆提取步骤	检测准确	漏检主要原因在于前面的椭圆提取步骤

图 4 显示了部分图片的检测定位结果, 红色椭圆标出了提取出的所有椭圆, 白色矩形框标出了聚类后油库的位置和大小, 框内的椭圆才被认为是油罐。椭圆提取步骤中用到了嵌入置信度的边缘检测算法^[5]和彭扬的椭圆快速提取算法^[6]。油库提取步骤中类内最少样本数目 N_0 设置为 3—6 邻域半径 r_0 设置为 300—600 分辨率。聚类时间 t 显示都为 0.0 也就是在 $t < 10^{-3}$ s 可以忽略不计。存在有部分油罐区域漏检的现象, 主要原因是椭圆提取步骤中椭圆提取太少而导致的, 而不是聚类方法的原因。因此, 从实验数据和实验图片可以看出, 本文提出的聚类方法能有效地将油罐目标聚类定位油库区域, 又能自动去除虚假油罐目标, 准确率较高, 计算速度快。

5 讨 论

本文提出的方法具有以下特点:

(1) 基于区域生长的聚类过程, 还可以认为是寻找连通区域的过程, 在这里连通区域的定义是, 在同一类里的任何两个样本, 可以通过若干个样本的邻域连接起来, 从而一个连通区域就是一类。

(2) 由于分辨率的设置大小会影响到前面的椭圆提取, 通常分辨率设置越大检测出的弱小边缘就越多, 因此检测出的椭圆就越多, 所以在油库检测定位过程中分辨率的设置可以比实际分辨率略大, 从而使油罐尽量少漏检, 而增加的虚假目标可以通过

区域生长的油库聚类方法去除掉。

(3) 实验结果表明基于区域生长的聚类方法有以下三个优点: 在油库定位的同时可有效地去除虚假目标, 定位准确率高; 可以允许前面的油罐目标检测有一定漏检和虚警; 计算量小, 不需要迭代。

(4) 该方法对邻域半径 r_0 的选择比较敏感, 如何通过某种算法自动调节 r_0 到最佳的邻域半径值尚需进一步研究。

参 考 文 献 (References)

- [1] Macqueen J. Some Methods for Classification and Analysis of Multivariate Observations [A]. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability [C], 1967.
- [2] Ball G, Hall D. A Clustering Technique for Summarizing Multivariate Data [J]. *Behavioral Sciences* 1976 (12): 153—155.
- [3] Hong T H, Rosenfeld A. Compact Region Extraction Using Weighted Pixel Linking in a Pyramid [J]. *IEEE-PAMI* 1984 6(2): 222—237.
- [4] Zhang Y J. Image Segmentation [M]. Beijing: Science Press 2001. [章毓晋. 图像分割 [M]. 北京: 科学出版社, 2001.]
- [5] Meer P, Georgescu B. Edge Detection with Embedded Confidence [J]. *IEEE Trans Pattern Anal Machine Intell*, 2001, 23(12): 1351—1365.
- [6] Peng Y, Hu F Q, Li J G. Fast Extraction of Ellipses [J]. *Journal of Shanghai Jiaotong University*, 1998, 32(9): 61—63. [彭扬, 胡福乔, 李介谷. 椭圆特征的快速提取 [J]. 上海交通大学学报, 1998, 32(9): 61—63.]



(a)



(b)

图 4 基于区域生长原理的聚类方法定位油库实验图片结果

(a) 02 bmp (b) 04 bmp

Fig 4 Image results of localizing oil houses by the proposed methods

(a) 02 bmp (b) 04 bmp