

基于粗糙集的图像聚类分割方法研究^{*}

刘 岩¹, 岳应娟², 李言俊¹, 张 科¹

(1. 西北工业大学 航天工程学院, 陕西 西安 710072;

2. 第二炮兵工程学院 基础部, 陕西 西安 710025)

摘 要: 结合粗糙集理论和 K -均值聚类算法, 提出了一种图像分割方法。将原图像按等价关系进行划分, 基于属性约简的概念对不同区域按照不可分辨关系分类。分割结果表明, 文中方法是一种有效的图像分割方法, 具有良好的鲁棒性。

关键词: 粗糙集; K -均值聚类; 图像分割

中图分类号: TP391 **文献标识码:** A **文章编号:** 1007-2276(2004)03-0300-03

Application of Rough Set and K -means clustering in image segmentation^{*}

LIU Yan¹, YUE Ying-juan², LI Yan-jun¹, ZHANG Ke¹

(1. College of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China;

2. The Second Artillery Engineering Institute, Xi'an 710025, China)

Abstract: Rough Set theory is a new mathematical tool to deal with problems on vagueness and uncertainty. An image segmentation method based on Rough Set theory and K -means clustering is presented. The original image is segmented according to the relation of equal value. By applying value reduct to the attribute values, different regions are classified based on indiscernibility. The experimental results indicate that the method can improve veracity and stability of image segmentation.

Key words: Rough Set; K -means clustering; Image segmentation

0 引 言

图像分割是把目标从背景中提取出来, 为更高层次的图像处理和识别打下基础。聚类方法经常用于分割边界模糊的图像, 利用此方法进行图像分割可看作是对阈值分割概念的推广。常用的聚类方法有 K -均值聚类算法、FCM 算法和 ISODATA 聚类算法等。

K -均值聚类算法的优点是能够动态聚类, 具有一定的自适应性。但是, 其结果易受聚类中心个数 K 及初始聚类中心的影响, 同时也受样本的几何形状及排列次序的影响。因此, 算法能否收敛取决于样本的特性和其能否形成不同区域的个数。粗糙集作为一种新的处理模糊和不确定知识的数学工具, 目前已广泛应用于模式识别、人工智能、图像处理和数据分析等领域。粗糙集体现了集合中对象的不可区分性, 即由

收稿日期: 2003-07-08; 修订日期: 2003-10-18

^{*} 基金项目: 高等学校博士点专项科研基金(20020699014)

作者简介: 刘岩(1973-), 男, 江苏镇江人, 讲师, 博士生, 主要研究方向为精确制导、图像处理、模式识别。

于知识的粒度而导致的粗糙性。图像信息具有较强的复杂性和相关性,在处理过程中经常出现不完整性和不精确性问题,将粗糙集理论应用于图像的处理和理解,具有比硬计算方法更好的效果。

文中提出一种基于粗糙集理论的 K 均值聚类图像分割方法。首先,利用 K 均值聚类将原图像分为 K 个类,且此结果可描述图像每一像素的属性;其次,所有像素属性决定等价关系,并且根据所有不可分辨关系,将图像分为一些小区域;然后根据这些区域的不同点进行属性约简;最后,由区域的不同点决定等价关系,通过综合这种等价关系可得到满意的分割结果。

1 K -均值聚类和粗糙集理论基础

1.1 K -均值聚类算法

令 $x = (x_1, x_2)$ 代表一个像素坐标, $g(x)$ 代表此像素灰度值, K -均值聚类算法的目的是使下式的值达到最小:

$$E = \sum_{i=1}^K \sum_{x \in Q_j^{(i)}} g(x) - \mu_j^{(i+1)} \quad (1)$$

式中 $Q_j^{(i)}$ 表示在第 i 次迭代后赋给类 j 的像素集合; μ_j 表示第 j 类的均值。具体步骤如下:

- (1) 任意选 K 个初始类均值 $\mu_1^{(1)}, \mu_2^{(1)}, \dots, \mu_K^{(1)}$;
- (2) 在第 i 次迭代时,根据下述准则将每个像素都赋给一个 K 类,即 $x \in Q_j^{(i)}$, 如果 $g(x) - \mu_j^{(i)} < g(x) - \mu_j^{(i)}$;
- (3) 对 $j=1, 2, \dots, K$, 更新类均值 $\mu_j^{(i+1)}, \mu_j^{(i+1)} = 1/N_j \sum_{x \in Q_j^{(i)}} g(x)$, 其中 N_j 是 $Q_j^{(i)}$ 中的像素个数;
- (4) 如果对所有的 $j=1, 2, \dots, K$, 有 $\mu_j^{(i+1)} = \mu_j^{(i)}$, 则算法收敛,结束。否则退回步骤(2)继续下一次迭代。

1.2 粗糙集的概念

定义1 一个近似空间(或知识库)定义为一个关系系统(或二元组):

$$K = (U, R) \quad (2)$$

式中 $U \neq \emptyset$ ($\emptyset$ 为空集), 是一个被称为论域的个体集合; R 是 U 上等价关系的一个族集。

给定一个有限的非空集合 U 称为论域, R 为一等价关系, 则称知识库 $K = (U, R)$ 为一个近似空间。设 x 为 U 中的一个对象, X 为 U 的一个子集, $R(x)$ 表示所有与 x 不可分辨的对象所组成的集合。当 X

能用 R 的属性确切地描述时, R 可定义 X , 称 X 为 R 精确集; 当 X 不能用 R 的属性确切地描述时, R 不可定义 X , 称 X 为 R 非精确集或 R 粗糙集。

定义2 设 $P \subseteq R$, 且 $P \neq \emptyset$, P 中所有等价关系的交集称为 P 上的一种不可分辨关系, 记作 $ind(P)$, 即:

$$[X]_{ind(P)} = \bigcap_{P \in P} [X]_P \quad (3)$$

定义3 设给定知识库 $K(U, R)$, 对于每个子集 $X \subseteq U$ 和一个等价关系 $R \in ind(K)$, 可以根据 R 的基本集合描述来划分集合 X :

$$R_-(X) = \{Y \mid Y \subseteq U / ind(R) : Y \subseteq X\} \quad (4)$$

$$R^+(X) = \{Y \mid Y \subseteq U / ind(R) : Y \cap X \neq \emptyset\} \quad (5)$$

$$bn_R(X) = R^+(X) - R_-(X) \quad (6)$$

式中 $R_-(X)$ 和 $R^+(X)$ 分别为 X 的 R 下近似和 R 上近似; $bn_R(X)$ 为 X 的边界。集合的下近似是包含给定集合中所有基本集的集合, 集合的上近似是包含给定集合元素中所有基本集的最小集合。显然, 当 $bn_R(X) \neq \emptyset$ 时, X 是一个不精确的概念。

2 基于粗糙集理论的 K -均值聚类算法

2.1 属性提取

首先用 K -均值聚类法将图像分为 K 个类, 并从中确定每个像素的属性值, 然后根据属性值得到不可分辨关系:

$$\{U/R_{x_1}, U/R_{x_2}, \dots, U/R_{x_n}\}$$

即得到等价关系。像素属于同一区域称为一类。

2.2 属性约简

基于属性约简的概念计算出每个属性的权值, 并用权值决定区域内的不同点。每一个决策都是基于类的等价关系。

设 R 是等价关系的一个族集, U 为论域, x_i 为 U 中的对象, a_k 是第 k 个属性, v_i 是族集 R 上的对象 x_i 的属性约简。 L 是 U 上的所有属性约简的数值。属性 a_k 的权值 w_k 定义为:

$$w_k = 1 - \frac{\sum_{x_i \in U} \omega(v_i, a_k)}{L} \quad (7)$$

$$\omega(v_i, a_k) = \begin{cases} 1 & \text{如果 } v_i(a_k) \text{ 是可省的属性} \\ 0 & \text{如果 } v_i(a_k) \text{ 是不可省的属性} \end{cases}$$

通过分配每一个属性的权值, 可计算各个类间的

相异度。设 $U = \{x_1, x_2, \dots, x_n\}$ 为论域, $A = \{a_1, a_2, \dots, a_m\}$ 是一组属性, $a_k(x_i)$ 定义为属性 a_k 在类 x_i 上对应的值。类 x_i 和 x_j 的相异度 $d_N(x_i, x_j)$ 定义为:

$$d_N(x_i, x_j) = \sum_{k=1}^m (w_k \times \lambda_{ij}^k) \quad (8)$$

$$\lambda_{ij}^k = \begin{cases} 0 & \text{如果 } a_k(x_i) = a_k(x_j) \\ 1 & \text{如果 } a_k(x_i) \neq a_k(x_j) \end{cases}$$

且 $d_N(x_i, x_j) = d_N(x_j, x_i)$, $d_N(x_i, x_i) = 1$ 。

2.3 区域分类

根据区域的不同点将相同和不同的区域分类。等价关系将论域分为两类,一类包含相似区域,而另一类包含不同区域。对于类 $x_i \in U$, 定义等价关系 R_j 为:

$$U/R_i = \{\{x_j \mid s(x_i, x_j) \leq Th\}, \{x_j \mid \text{others}\}\},$$

for all $j(1 \leq j \leq n)$

其中 Th 是一个不同的阈值,用来区分 x_i 和 x_j 是否相似的区域。最后,利用所有的等价关系的交集得到不可分辨关系 R :

$$[X]_{ind(R)} = \bigcap_{R_i \in R} [X]_{R_i}$$

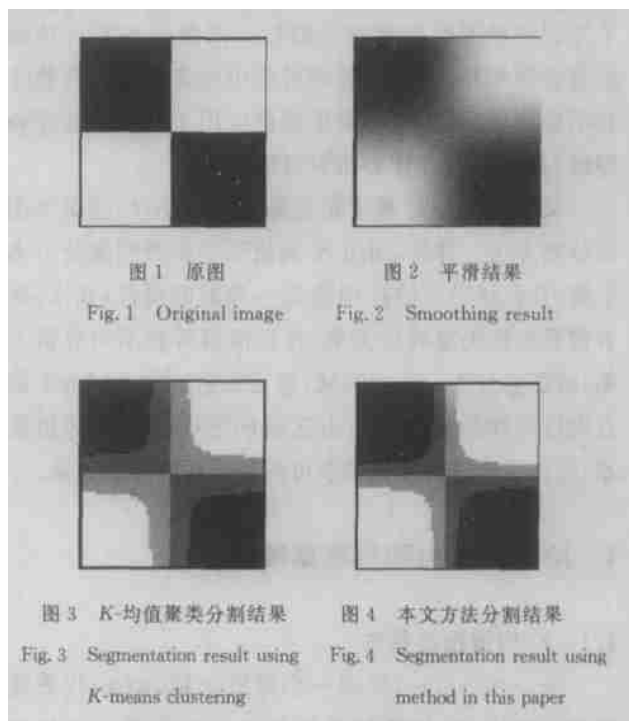
定义相似区域是基于不可分辨关系 R 上的相似的类,合并相似区域,继而得到最后的分割结果。

3 实例分析

图 1 为原图像,图 2 为对图 1 进行平滑 60 次后所得图像,图 3 为对图 2 用 K -均值聚类图像分割方法所得到的结果,图 4 为利用文中方法,用灰度值、位置和灰度通道三个属性值和两个不同的阈值对图 2 进行处理所得的结果。从结果中可以看出,两种方法都将图 2 分为 4 类,分割后的图像包含 4 个灰度级,每一个灰度级代表不同的区域。文中方法的分割结果明显要比用 K -均值聚类图像分割方法所得到的结果更细致,更接近图 1 的原貌。

4 结束语

粗糙集理论还处在不断发展和完善的过程中,尚



有许多理论问题亟待解决。但用该理论处理不确定性问题的独特方式以及与其他理论的较好融合,将有利于其在应用领域特别是智能控制领域的进一步发展。粗糙集理论为智能信息提供了有效的处理手段,尤其是对特征提取,简化信息和不精确知识的表达、归纳等十分有效。本文基于粗糙集理论和 K -均值聚类方法提出了一种新的图像分割方法,得到了令人满意的分割结果。

参考文献:

- [1] Selim S Z, Ismail M A. K -means-type algorithm[J]. IEEE Trans Pattern Anal Mach Intell, 1994, 6(1): 81-87.
- [2] Tsumoto S. Automated extraction of medical expert system rules from clinical databases based on rough set theory[J]. Information Science, 1998, 112: 67-84.
- [3] 曾黄麟. 粗糙集理论及其应用[M]. 重庆:重庆大学出版社, 1998.
- [4] 郑南宁. 计算机视觉与模式识别[M]. 北京:国防工业出版社, 1998.
- [5] 章毓晋. 图像分割[M]. 北京:科学出版社, 2001.
- [6] 王珏, 苗夺谦. 关于 Rough Set 理论与应用的综述[J]. 模式识别与人工智能, 1996, 9(4): 337-344.