基于改进聚类的机体损伤识别方法 研究



范杰,司伟森,巴翔

中国南方航空河南公司,河南 郑州 450000

摘 要:为了清晰快速地辨识飞机机体损伤区域,引入*K*均值聚类算法处理损伤图像,通过分析*K*均值聚类算法的局限性, 提出了基于误差平方和差值以及基于像素点变化的聚类迭代终止条件的优化方法,并通过飞机机体损伤图像实例,从损伤 识别效果和运算效率两个方面进行验证。验证结果显示,基于改进聚类的机体损伤识别方法,在保证图像损伤区域识别效 果的同时,明显减少了迭代次数,提高了聚类算法的运算效率,可以满足飞机机体损伤高效识别的要求。

关键词:损伤区域划分;图像识别;K均值聚类算法

中图分类号:V267

文献标识码:A

随着图像处理技术的不断发展,彩色图像中目标区域 提取技术已广泛应用于诸多领域,但由于应用领域的差异 和目标区域的复杂性,提取彩色图像中有意义的目标区域 一直是难点问题。

聚类算法以其简单高效的分类机制成为图像目标区域 高效提取的典型算法。目前,主流的聚类算法主要有两步 聚类、层次聚类和K均值聚类算法^[1]。其中,K均值聚类算 法对大规模图像集的高效处理,使其成为最常用的聚类算 法之一。周宏元等^[2]将K均值聚类引入蝴蝶优化算法,实现 土木工程结构损伤识别。赵杨超等^[3]利用K均值聚类实现 对航空复合材料的敲击检测。刘立坤等^[4]引入聚类算法提 出颤振模态参数自动辨识方法。但由于K均值聚类算法过 分依赖于迭代次数的设定且对离群点和噪声点敏感,聚类 结果经常为局部最优而非全局最优^[5]。

本文面向机体损伤识别方法研究,针对K均值聚类算 法的局限性,研究K均值聚类改进算法,解决迭代次数人工 设定问题,实现损伤区域的精准聚类,进而实现机体损伤的 高效识别。

1 K均值聚类算法

由于服役条件、维护修理不当、设计制造缺陷等因素, 航空结构在服役过程中容易产生小损伤⁶⁶。当造成非常细

DOI: 10.19452/j.issn1007-5453.2022.05.005

微且无法用肉眼进行观察的飞机机体损伤时,需要对损伤 区域进行图像处理来损伤识别^[7]。图像的显示是由不同的 灰度像素点组成的,通过*K*均值聚类法确定锈蚀和完好区 域范围,将图像上的像素点划分到这两类区域中,达到区分 损伤区域的目的,从而方便维修^[8]。

K均值聚类算法常用欧式距离作为其相似性的评价指标^[9],假设给定的数据集 $X = \{x_m | m = 1, 2, \dots, \text{total}\}, X$ 中的样本用d个描述属性 (A_1, A_2, \dots, A_d) ,并且d个描述属性都是连续型属性。数据样本 $x_i = (x_{i1}, x_{i1}, \dots, x_{id}), x_j = (x_{j1}, x_{j1}, \dots, x_{jd})$ 。其中 $, x_{i1}, x_{i1}, \dots, x_{id}$ 和 $x_{j1}, x_{j1}, \dots, x_{jd}$ 分别是样本 x_i 和 x_j 对应d个描述属性 (A_1, A_2, \dots, A_d) 的具体取值。样本 x_i 和 x_j 之间的相似度通常用它们之间的距离 $d(x_i, x_j)$ 来表示,距离越小,样本 x_i 和 x_j 越相似,差异度越小;距离越大,样本 x_i 和 x_i 越不相似,差异度越大^[10]。距离公式为

$$d(x_{i}, x_{j}) = \sqrt{\sum_{k=1}^{d} (x_{ik} - x_{jk})^{2}}$$
(1)

聚类算法使用误差平方和准则函数来评价聚类性能。 给定数据集*X*,其中只包含描述属性,不包含类别属性。假 设*X*包含*k*个聚类子集*X*₁,*X*₂,…,*X*_k;各个聚类子集中的样本 数量分别为*n*₁,*n*₂,…,*n*_k;各个聚类子集的均值代表点(也称聚 类中心)分别为*m*₁,*m*₂,…,*m*_k。则误差平方和准则函数公式为

收稿日期:2021-11-15; 退修日期:2022-01-15; 录用日期:2022-03-03

引用格式: Fan Jie, Si Weisen, Ba Xiang.Research on airframe damage identification method based on improved clustering[J].Aeronautical Science & Technology, 2022, 33(05):32-36. 范杰, 司伟森, 巴翔.基于改进聚类的机体损伤识别方法研究[J]. 航空科学技术, 2022, 33(05):32-36.

$$E = \sum_{i=1}^{k} \sum_{p \in X_{i}} \left\| p - m_{i} \right\|^{2}$$
(2)

式中:k为要聚类的个数;p为样本 X_i 中的元素; m_i 为样本 X_i 的均值。

相似度的计算根据一个簇中对象的平均值来进行。具体为:(1)将所有对象随机分配到 k 个非空的簇中;(2)计算每个簇的平均值,并用该平均值代表相应的簇;(3)根据每个对象与各个簇中心的距离,分配给最近的簇;(4)然后转第(2)步,重新计算每个簇的平均值。这个过程不断重复, 直到满足聚类中心不再改变或者达到最大迭代次数。

2 K均值聚类算法的改进

K均值聚类算法的计算准确度由其迭代次数决定。一般 迭代终止条件是聚类中心不再改变或者达到最大迭代次数, 以节省计算时间^[11-12]。而机体损伤区域识别中,K均值聚类的 迭代后期,聚类中心虽未达到稳定值,但从图像上观察,肉眼 难以分辨输出图像的差异,对损伤辨识效果没有影响。本文 提出基于误差平方和差值以及基于像素点变化的两种改进迭 代终止条件,确定观察图像最佳迭代次数的迭代终止参数。

2.1 基于误差平方和差值的优化方法

K均值聚类算法停止迭代的条件是聚类中心不变或者 达到最大迭代次数,可使用改进误差平方和方法,由于迭代 后期误差平方和变化较小,则设定迭代终止条件为前后两 次迭代过程误差平方和之差小于阈值时即停止迭代,保证 损伤辨识效果的同时减少迭代次数。

误差平方和

$$E = \sum_{i=1}^{k} \sum_{p \in X_{i}} \left\| p - m_{i} \right\|^{2}$$
(3)

式中:X_i为簇集;p为簇集元素;m_i为每次迭代的聚类中心。 误差平方和差值

$$C = E(m) - E(m - 1)$$
(4)

式中:m为相应迭代次数。迭代终止参数为

$$R = \frac{C}{E(m)} \tag{5}$$

式中:C为误差平方和差值。

2.2 基于像素点变化的优化

迭代终止条件还可以从前后两次迭代过程输出的像素 点数目变化入手,设置迭代的终止条件为:前后两次迭代像 素点变化值小于某一值时即输出图像。这样既能保证输出 图像的效果,又能减少迭代次数,减少运算时间。

像素点变化值为

$$S = \left| X(m) - X(m-1) \right| \tag{6}$$

式中:*X*(*m*)为像素点的数目;*m*为相应的迭代次数。比值为

$$B = \frac{S}{A} \tag{7}$$

式中:A为图像总的像素点数目。

2.3 K均值聚类改进算法流程

K均值聚类改进算法流程图如图1所示。具体流程为: (1)初始化簇集,确定要聚类的数据;(2)确定聚类中心数目,随机选取初始化聚类中心;(3)使用式(1)计算数据与聚类中心的距离,按照数据与聚类中心较近原则分配到相应的簇中; (4)使用式(2)计算各簇集中的数据与聚类中心的误差平方和;(5)求出各个簇集中数据的平均值作为新的聚类中心;(6) 判断是否达到规定最大迭代次数或者聚类中心不再改变,如 果不改变完成计算过程,输出迭代次数。



图1 改进K均值聚类流程图

Fig.1 Flow chart of improved K-means clustering algorithm

3 基于改进 K 均值聚类的机体损伤识别方法 验证

为验证基于改进*K*均值聚类的机体损伤识别方法,选 取螺钉锈蚀、蒙皮锈蚀、小孔锈蚀、涂层剥落损伤、丝状锈蚀 和应力锈蚀裂纹等损伤图像为实例,如图2所示。

针对损伤图像实例,分别使用K均值聚类算法和改进 算法进行对比验证,验证结果图像如图3~图8所示,实例a



(a) 螺钉锈蚀



(c) 小孔锈蚀





(b) 蒙皮锈蚀

(c) 丝状锈蚀
 (f) 应力锈蚀裂纹
 图 2 机体损伤图像
 Fig.2 Airframe damage image





 (a) K均值聚类验证结果
 (b) 改进算法验证结果

 图 3 实例 a 验证结果图像

 Fig.3 Verification result image of instance a





 (a) K均值聚类验证结果
 (b) 改进算法验证结果

 图 4 实例 b 验证结果图像







 (a) K均值聚类验证结果
 (b) 改进算法验证结果

 图 5 实例 c 验证结果图像

 Fig.5 Verification result image of instance c





 (a) K均值聚类验证结果
 (b) 改进算法验证结果

 图 6 实例 d 验证结果图像

 Fig.6 Verification result image of instance d





 (a) K均值聚类验证结果
 (b) 改进算法验证结果

 图 7 实例 e 验证结果图像

 Fig.7
 Verification result image of instance e



 (a) K均值聚类验证结果
 (b) 改进算法验证结果

 图 8 实例 f 验证结果图像

 Fig.8 Verification result image of instance f

和实例b验证计算过程见表1。

损伤识别效果方面,由图 3~图 8 的验证结果图像可以看 出,K均值聚类算法的验证结果图像与改进K均值聚类算法的 验证结果图像相比,视觉区分度较小。对于整体色度单一的 实例c,改进K均值聚类算法获得的验证结果图像中,损伤区 域清晰可辨识,与传统K均值聚类算法的验证结果图像相比, 损伤识别效果更优。因此,从损伤识别效果方面来说,改进K 均值聚类算法对于损伤识别领域的应用效果较好。

在运算效率方面,由表1、表2的验证计算过程数据可以 看出,随着迭代次数的增多,误差平方和差值减小,输出图片 损伤区域的像素点逐渐增多。当迭代符合终止条件时,误差 平方和比值接近0.001,像素点数目几乎不变,输出的图像面

迭代	聚类	聚类	误差	误差平方和	误差平方和	像素点	像素点数目	变化值
次数	中心1	中心2	平方和	差值	比值	数目	变化值	占总数比值
1	135.1	90.51	10070416	2168915	0.215	6972	3580	0.11
2	139.4	96.84	7901501	1368457	0.173	10552	2541	0.08
3	142.5	100.47	6533043	529106	0.081	13093	1301	0.04
4	144.1	102.23	6003937	142276	0.024	14394	799	0.025
5	145.2	103.3	5861661	40995	0.007	15193	326	0.01
6	145.6	103.73	5820666	5802	0.001	15519	0	0
7	145.6	103.73	5814864			15519	_	_

表1 实例a验证计算过程数据

Table 1 Validation data of instance a

表2 实例b验证计算过程数据

able 2	Validati	on data d	of inst	tance	c

24-71	Hy 시스	TH AL	·0-24		ND 쓰 코 스 Te	協主上	協主上北口	主 /1./左
迭代	衆尖	衆尖		误差平方和差值	误差半力和	像系点	像紊点数日	受化阻
次数	中心1	中心2	平方和		比值	数目	变化值	占总数比值
1	71.11	132	6512305	335249	0.051	2726	624	0.037
2	76.42	133.5	6177056	258772	0.042	3350	600	0.036
3	80.46	134.9	5918284	199388.5	0.034	3950	572	0.034
4	83.69	136.34	5718895	135909	0.024	4522	609	0.036
5	86.69	137.768	5582987	99527.5	0.018	5131	399	0.024
6	88.48	138.6	5483459	39133.5	0.007	5530	197	0.012
7	89.33	139.15	5444326	10700	0.002	5727	178	0.011
8	90.07	139.56	5433626	5116.5	0.001	5905	0	0
9	90.07	139.56	5428509	—	_	5905	—	_

积不再变化,迭代运算结束。聚类迭代运算可通过误差平方 和以及像素点数目的变化率动态优化迭代次数,达到最优迭 代次数,不需人工设定迭代次数。在保证损伤区域辨识效果 的基础上,减少运算复杂度,提升计算效率。

4 结论

本文通过分析不同迭代数目输出图像的区别,发现原 算法存在不足,随着迭代次数的增多,图像效果并不是越来 越清晰,所以对原始K均值聚类算法进行改进设计,原始算 法的迭代终止条件是聚类中心不变或者达到最大迭代次 数,由于初始聚类中心的随机选取对结果影响很大,所以在 保证初始聚类中心不变的情况下,改进算法的迭代终止条 件,一种改进方法是根据误差平方和差值进行改进,另一种 方法是根据像素点变化的数目来进行改进。结果证明,改 进算法在保证输出图像效果的同时,减少了迭代计算次数, 提高了执行效率。

参考文献

[1] 李彬.具有全局优化能力的K均值聚类算法[J].西南师范大

学学报(自然科学版),2014,39(7):36-40.

Li Bin. On K-means clustering algorithm with global optimization ability[J]. Journal of Southwest China Normal University(Natural Science Edition),2014,39(7):36-40.(in Chinese)

- [2] 周宏元,张广才,王小娟,等.基于互相关函数的聚类蝴蝶优化 算法的结构损伤识别[J].振动与冲击,2021,40(17):189-196.
 Zhou Hongyuan, Zhang Guangcai, Wang Xiaojuan, et al. Structural damage identification based on clustering butterfly optimization algorithm of cross-correlation function[J]. Journal of Vibration and Shock, 2021,40(17): 189-196.(in Chinese)
- [3] 赵杨超,李艳军,曹愈远,等.基于K-means聚类的航空复合材料敲击检测研究[J].航空科学技术,2019,30(10):45-49.
 Zhao Yangchao, Li Yanjun, Cao Yuyuan, et al. Study on percussion detection of aerospace composite materials based on K-means clustering[J]. Aeronautical Science & Technology, 2019,30(10):45-49.(in Chinese)
- [4] 刘立坤,闫轲.随机子空间颤振模态参数自动辨识方法研究
 [J].航空科学技术,2020,31(3):53-58.
 Liu Likun, Yan Ke. Research on automated stochastic subspace

flutter modal identification method[J]. Aeronautical Science & Technology, 2020,31(3):53-58.(in Chinese)

- [5] 贲志伟,赵勋杰.基于改进的K均值聚类算法提取彩色图像 有意义区域[J].计算机应用与软件,2010,27(9):11-13.
 Ben Zhiwei, Zhao Xunjie. Extracting meaningful region in colour image based on improved K-means clustering algorithm[J]. Computer Application and Software,2010,27(9):11-13.(in Chinese)
- [6] 邢博邯,蔡建,周智权,等. 航空结构中Lamb 波小损伤监测能 力研究[J]. 航空科学技术,2021,32(3):40-48.
 Xing Bohan, Cai Jian, Zhou Zhiquan, et al. Research on smalldamage monitoring abilities of Lamb waves in aircraft structures[J]. Aeronautical Science & Technology, 2021, 32(3):40-48.(in Chinese)
- [7] 李忠良,魏麟,刘晔璐,等.飞机外表面地检设备图像处理技术 初探[J].航空维修与工程,2018(5):56-58.

Li Zhongliang, Wei Lin, Liu Yelu, et al. Preliminary study on image processing technology of ground inspection equipment for aircraft[J]. Aviation Maintenance and Engineering, 2018(5): 56-58. (in Chinese)

 [8] 丁明月,庄晓东.基于数据融合的K均值聚类彩色图像分割 方法[J].青岛大学学报(工程技术版),2018,33(2):42-46,58.
 Ding Mingyue, Zhuang Xiaodong. *K*-means clustering color image segmentation method based on data fusion[J]. Journal of Qingdao University (Engineering and Technology), 2018, 33 (2): 42-46,58.(in Chinese)

- [9] Park H S, Jun C H. A simple and fast algorithm for K-means clustering[J]. Expert Systems with Applications, 2009, 36 (2): 3336-3341.
- [10] 刘玉洁.改进的K均值聚类算法彩色图像分割的研究[J].工 业控制计算机,2012,25(4):76-77.
 Liu Yujie. Color image segmentation based on an improved K-means clustering algorithm[J]. Industrial Control Computer, 2012,25(4):76-77.(in Chinese)
- [11] 裴晨,徐国彬,于艺铭,等.基于改进K均值聚类的图像修复方法[J].包装工程,2020,41(23):255-262.
 Pei Chen, Xu Guobin, Yu Yiming, et al. Image restoration method based on improved *K*-means clustering[J]. Packaging Engineering,2020,41(23):255-262.(in Chinese)
- [12] 谢娟英,蒋帅,王春霞,等.一种改进的全局K-均值聚类算法
 [J].陕西师范大学学报(自然科学版),2010,38(2):18-22.
 Xie Juanying, Jiang Shuai, Wang Chunxia, et al. An improved global K-means clustering algorithm[J]. Journal of Shanxi Normal University (Natural Science Edition), 2010, 38(2):18-22.(in Chinese)

Research on Airframe Damage Identification Method Based on Improved Clustering

Fan Jie, Si Weisen, Ba Xiang

China Southern Airlines Henan Branch, Zhengzhou 450000, China

Abstract: In order to identify the airframe damage region clearly and quickly, the *K*-means clustering algorithm is introduced to process the damage image. By analyzing the limitations of the *K*-means clustering algorithm, an improved method of clustering iteration termination condition based on error square sum difference and pixel change is proposed. Through the verification of aircraft body damage images, it is verified from two aspects: damage recognition effect and operation efficiency. The verification results show that the airframe damage identification method based on improved clustering not only ensures the identification effect of image damage regions, but also significantly reduces the number of iterations, improves the operation efficiency of the clustering algorithm, and can meet the requirements of efficient processing of airframe damage identification.

Key Words: damage region division; image recognition; K-means clustering algorithm