北航学报
赠 阅

ISSN 1001-5965
CODEN BHHDE8

北京航空航天大學学报

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

2015-04
Vol.41 No.4

北京航空航天大学学报

第 41 卷　第 4 期　（总第 266 期）　2015 年 4 月

北航学报
赠阅

目　次

曲线片型加筋壁板的稳定性优化设计 …… 徐元铭，王东（567）
基于石墨烯膜的光纤 Fabry-Perot 腔干涉特性分析 …… 李成，郭婷婷，肖俊，樊尚春，靳伟（574）
具有方向特性的 X 角点的亚像素检测定位 …… 孟偲，薛姣（580）
超高频脉冲电弧焊接熔池金属流动行为…… 杨明轩，杨舟，从保强，王强，齐铂金（589）
高超声速气动热数值计算壁面网格准则 …… 张智超，高振勋，蒋崇文，李椿萱（594）
空间科学任务协同设计论证平台 …… 邓丽，韩潮，曹晋滨，宋友（601）
基于脉冲激光定位的 SRAM 单粒子闩锁事件率预估 …… 余永涛，韩建伟，封国强，蔡明辉（609）
基于大涡模拟的平流层浮空器气动特性分析 …… 王裕夫，陶国权，刘东旭，武哲（616）
基于压缩感知的 DSSS 信号双阶段捕获方法 …… 程艳合，杨文革（624）
过渡状态下材料断裂韧性的计算方法 …… 高潮，何宇廷，崔荣洪，伍黎明（632）
面向机身柔性装配的在线编程技术 …… 武锋锋，李东升，王亮（641）
基于单幅立式标靶图像的单目深度信息提取…… 黄小云，高峰，徐国艳，丁能根，邢龙龙（649）
基于流动控制的水下航行体流体动力特性分析 …… 张孝石，王聪，曹伟，赵成功（656）
Loop 细分曲面精加工刀具轨迹生成 …… 陈甜甜，赵罡（663）
基于显著特性的保持边缘滤波算法 …… 冯策，戴树岭，赵帅和（669）
基于修正 M-K 模型的铝合金板材成形极限图预测 … 杨希英，郎利辉，刘康宁，蔡高参，郭禅（675）
基于负载特征聚类的节能资源调度算法 …… 夏庆新，兰雨晴，唐甜，肖利民（680）
力约束管材自由胀形试验研究与材料性能测试 ……
…… 程鹏志，郎利辉，葛宇龙，阮尚文，王韬，武海（686）
氯盐对碱激发矿渣净浆强度影响试验 …… 程寅，黄新（693）
基于涡方法生成大涡模拟进口条件的数值计算 …… 谢宝林，邵亮（701）
偏置弹簧对 SMA 驱动器输出性能影响规律 …… 杨宝锋，张小勇，闫晓军（707）
基于时空关联图模型的视频监控目标跟踪 …… 张诚，马华东，傅慧源（713）
抗代间污染攻击的网络编码签名方案 …… 彭天丽，尚涛，刘建伟（721）
矩的显式积分算法研究与应用 …… 付子豪，龚光红（727）
基于局部径向二值模式的三维人脸识别 …… 刘青，孙军华（732）
电磁航天器编队动力学建模与运动规划方法 …… 邵龙飞，师鹏，赵育善（737）
基于分段非线性 Arrhenius 的贮存寿命评估方法 …… 周洁，姚军，宋燕（744）
基于多尺度梯度及深度神经网络的汉字识别 …… 潘炜深，金连文，冯子勇（751）

期刊基本参数：CN11-2625/V * 1956 * m * A4 * 190 * zh * P * ￥20.00 * 900 * 28 * 2015-04

（**编　辑**　娄　嘉　张　嵘　赵海容　李　晶）

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Vol. 41 No. 4 (Sum 266) April 2015

北航学报 赠阅

CONTENTS

Design and optimization of improving stability of curvilinear blade-stiffened panels
…… XU Yuanming, WANG Dong (567)
Interference characteristics analysis of optical fiber Fabry-Perot cavity with graphene diaphragm
…… LI Cheng, GUO Tingting, XIAO Jun, FAN Shangchun, JIN Wei (574)
Sub-pixel X-corner detection and location with directional characteristics
…… MENG Cai, XUE Jiao (580)
Metal molten behavior with ultra high frequency pulsed arc welding
…… YANG Mingxuan, YANG Zhou, CONG Baoqiang, WANG Qiang, QI Bojin (589)
Grid generation criterions in hypersonic aeroheating computations
…… ZHANG Zhichao, GAO Zhenxun, JIANG Chongwen, LEE Chunhian (594)
Demonstration platform for collaborative design of space science missions
…… DENG Li, HAN Chao, CAO Jinbin, SONG You (601)
SEL rate prediction for SRAM using pulsed laser sensitivity mapping
…… YU Yongtao, HAN Jianwei, FENG Guoqiang, CAI Minghui (609)
Analysis of stratospheric lighter-than-air vehicle's aerodynamic characteristics based on large eddy simulation
…… WANG Yufu, TAO Guoquan, LIU Dongxu, WU Zhe (616)
Two-stage acquisition algorithm for DSSS signal based on compressive sensing
…… CHENG Yanhe, YANG Wenge (624)
Calculation method of fracture toughness of materials at transition state
…… GAO Chao, HE Yuting, CUI Ronghong, WU Liming (632)
Online programming technique for flexible assembly of fuselage
…… WU Fengfeng, LI Dongsheng, WANG Liang (641)
Depth information extraction of on-board monocular vision based on a single vertical target image
…… HUANG Xiaoyun, GAO Feng, XU Guoyan, DING Nenggen, XING Longlong (649)
Hydrodynamic features of underwater vehicles based on flow control
…… ZHANG Xiaoshi, WANG Cong, CAO Wei, ZHAO Chenggong (656)
Tool path generation for Loop subdivision surface based finish machining
…… CHEN Tiantian, ZHAO Gang (663)
Edge-preserving filtering based on saliency map
…… FENG Ce, DAI Shuling, ZHAO Shuaihe (669)
Prediction of forming limit diagram of AA7075-O aluminum alloy sheet based on modified M-K model
…… YANG Xiying, LANG Lihui, LIU Kangning, CAI Gaocan, GUO Chan (675)
Energy-saving resource scheduling algorithm based on workload characteristic clustering
…… XIA Qingxin, LAN Yuqing, TANG Tian, XIAO Limin (680)
Tube free bulging experiment with force-end and material properties testing
…… CHENG Pengzhi, LANG Lihui, GE Yulong, RUAN Shangwen, WANG Tao, WU Hai (686)
Experiment of chloride effect on strength of alkali-activated slag paste
…… CHENG Yin, HUANG Xin (693)
Generation of unsteady inflow conditions for compressible LES with a modified vortex method
…… XIE Baolin, SHAO Liang (701)
Relationship between bias spring parameters and output performances of SMA actuators
…… YANG Baofeng, ZHANG Xiaoyong, YAN Xiaojun (707)
Object tracking in surveillance videos using spatial-temporal correlation graph model
…… ZHANG Cheng, MA Huadong, FU Huiyuan (713)
Signature scheme for network coding against inter-generation pollution attacks
…… PENG Tianli, SHANG Tao, LIU Jianwei (721)
Explicit moment integration algorithm and its application
…… FU Zihao, GONG Guanghong (727)
3D face recognition based on local radial binary pattern
…… LIU Qing, SUN Junhua (732)
Dynamics modeling and motion programming for electromagnetic formation flight
…… SHAO Longfei, SHI Peng, ZHAO Yushan (737)
Storage life evaluation method based on segmented nonlinear Arrhenius model
…… ZHOU Jie, YAO Jun, SONG Yan (744)
Recognition of Chinese characters based on multi-scale gradient and deep neural network
…… PAN Weishen, JIN Lianwen, FENG Ziyong (751)

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0257

曲线片型加筋壁板的稳定性优化设计

徐元铭*, 王东

(北京航空航天大学 航空科学与工程学院, 北京 100191)

摘 要: 加筋板作为一种典型结构被广泛应用于航空航天领域,为进一步拓展直筋加筋板的设计空间,提出了一套基于 ModelCenter 设计优化环境、Catia 和 Abaqus 的曲线片型加筋板优化设计框架. 一方面,筋条曲线采用参数数目可变的等距剖分三次 B 样条插值函数,为曲筋加筋板提供了足够大的优化空间. 另一方面,采用筋条加载端相聚和筋条垂直于板边 2 种方式来解决曲线片型加筋板的筋条加载问题,从而产生了一些新的曲筋加筋板布局形式. 优化结果表明,对于单向载荷主导的加筋板,曲筋加筋板相对于直筋加筋板稳定性改善不大. 但对于复杂组合载荷,曲筋加筋板获得了比直筋加筋板更好的稳定性.

关 键 词: 优化; 曲线片型加筋板; 稳定性; B 样条; 筋条加载
中图分类号: V214.4
文献标识码: A **文章编号**: 1001-5965(2015)04-0567-07

加筋板作为一种典型的薄壁结构形式,被广泛应用于飞机机身和机翼的盒段[1]、运载火箭的舱段[2],提高其在不同工况下的稳定性是工程设计的迫切需求. 曲筋加筋板作为直筋加筋板的一种结构扩展,已经开始受到研究工作者的重视. 目前,一些创新性的制造技术,如搅拌摩擦焊和电子束自由曲面成型技术已经可以使用铝合金与钛合金,从而使曲筋片型加筋板的整体加工成为了可能[3].

文献[3-7]曾对曲筋加筋板的优化设计做过研究并提出了优化系统 EBF3PanelOpt. 研究表明,对于 T 型筋,直筋加筋板是最优方案,而对于曲线片型筋条,曲筋加筋板相比直筋加筋板获得了更好的稳定性. 但由于 EBF3PanelOpt 框架中加筋板筋条设计参数数目固定,无法为曲筋加筋板提供更广阔的优化空间. 另外,EBF3PanelOpt 框架没有施加载荷于筋条上,这样便无法保证筋条的稳定性[7].

因此,本文提出了一套基于 ModelCenter、Catia和 Abaqus 的曲线片型加筋板优化设计框架. 一方面,筋条设计曲线采用设计参数数目可变的等距剖分三次 B 样条插值函数,为曲筋加筋板提供了足够大的优化空间. 另一方面,由于筋条只能承受其本身面内的载荷,采用了 2 种方式来考虑曲线片型加筋板的筋条加载问题. 第 1 种是 2 条或 2 条以上筋条加载端相聚的情况,并由此提出了一种新的曲筋布局形式;第 2 种是让筋条垂直于板边.

1 优化框架

优化平台基于 PHX ModelCenter 9.0 设计优化环境、Catia 模型设计工具和 Abaqus 有限元分析系统. 该平台能对一般的矩形板和筒状板进行稳定性或质量的优化设计.

利用 ModelCenter 的二次开发功能,自主编写 2 个数据模块:优化算法模块 Optimization 和建模

收稿日期: 2014-05-09; **录用日期**: 2014-08-05; **网络出版时间**: 2014-09-05 15:58
网络出版地址: www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0257.html
* **通讯作者**: 徐元铭(1965—),男,江苏武进人,教授,xuymg@sina.com,主要研究方向为飞行器结构设计、结构优化、神经网络专家系统.

引用格式: 徐元铭, 王东. 曲线片型加筋壁板的稳定性优化设计[J]. 北京航空航天大学学报, 2015, 41(4): 567-573.
Xu Y M, Wang D. Design and optimization of improving stability of curvilinear blade-stiffened panels[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(4): 567-573 (in Chinese).

北航学报 赠阅

与分析模块 CatiaAbaqusCOM. 优化算法模块内部封装了粒子群算法(PSO)并提供了目标函数、约束条件和设计变量接口. 建模与分析模块完成了从 Catia 建模到生成 stp 文件,再进行 Abaqus 有限元分析并获取数据的完整过程. 曲线片型加筋板优化框架如图 1 所示.

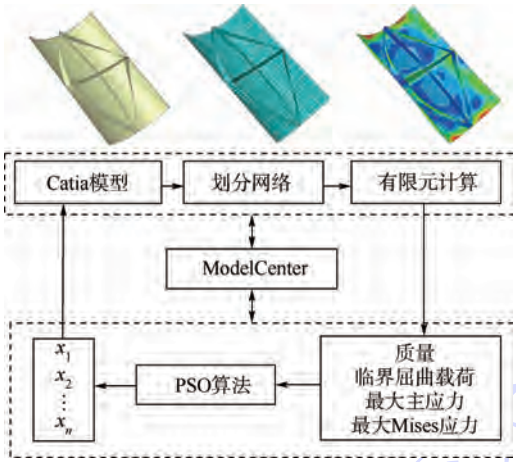


图 1 曲线片型加筋板优化框架

Fig. 1 Optimization framework of curvilinear blade-stiffened panels

1.1 曲筋曲线生成方法

采用三次等距剖分 B 样条插值方法生成曲线[8].

设 $f(x)$ 为定义在区间 $[a,b]$ 上的函数,$f(x_i)=f_i(i=0,1,\cdots,n)$,求三次样条函数 $s(x)$ 使得:

$$\begin{cases} s(x_i)=f_i \quad i=0,1,\cdots \\ s'(x_0)=f'_0 \\ s'(x_n)=f'(x_n) \end{cases} \tag{1}$$

在 $[a,b]$ 上给定一个等距剖分:

$$\Delta_h: a=x_0<x_1<\cdots<x_n=b$$

$$h=x_i-x_{i-1}=\frac{b-a}{n} \tag{2}$$

并记外节点 $x_{-1}=a-h, x_{n+1}=b+h$.

基于等距剖分三次 B 样条理论,可将插值问题解表示为

$$f(x)\approx s(x)=\sum_{i=-1}^{n+1} c_i B_3\left(\frac{x-x_i}{h}\right) \tag{3}$$

式中

$$B_3=\begin{cases} 0 & |x|\geqslant 2 \\ -\frac{1}{6}x^3+x^2-2x+\frac{4}{3} & 1\leqslant x<2 \\ \frac{1}{2}x^3-x^2+\frac{2}{3} & 0\leqslant x<1 \\ B_3(-x) & -2<x<0 \end{cases} \tag{4}$$

c_i 由式(5)求得.

$$\boldsymbol{Ac}=\boldsymbol{f} \tag{5}$$

式中

$$\boldsymbol{A}=\begin{bmatrix} -1 & 0 & 1 & & \\ 1 & 4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 4 & 1 \\ & & -1 & 0 & 1 \end{bmatrix}$$

为 $n+3$ 阶矩阵.

$$\boldsymbol{f}=(2hf'_0,6f_0,\cdots,6f_n,2hf'_n)^{\mathrm{T}}$$

$$\boldsymbol{c}=(c_{-1},c_0,\cdots,c_{n+1})^{\mathrm{T}}$$

1.2 优化数学模型

根据等距剖分下的三次 B 样条插值函数,每一条样条曲线由若干个等距剖分点和两端点处斜率唯一确定. 等距剖分三次 B 样条曲线如图 2 所示.

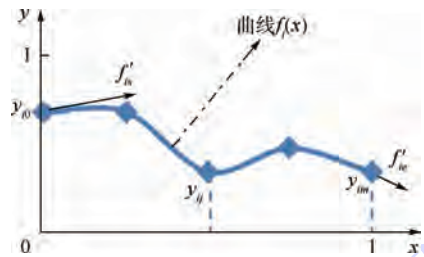


图 2 等距剖分三次 B 样条曲线

Fig. 2 Equally-split cubic B-spline curve

实际的筋条曲线的两端可以在板的 4 条边自由移动. 根据端点位于不同的板边,数学上可以通过定义坐标映射函数 map(S_i, y_{i0}, y_{im}),产生 6 种基本布局形式(约定曲线两端点不位于同一板边),如图 3 所示. 这里 S_i 是布局形式的离散变量,取 1 ~ 6. 其中,1 表示筋条在板宽对边连接,2 表示板长对边连接,3 ~ 6 分别表示板相邻边的 4 种连接方式.

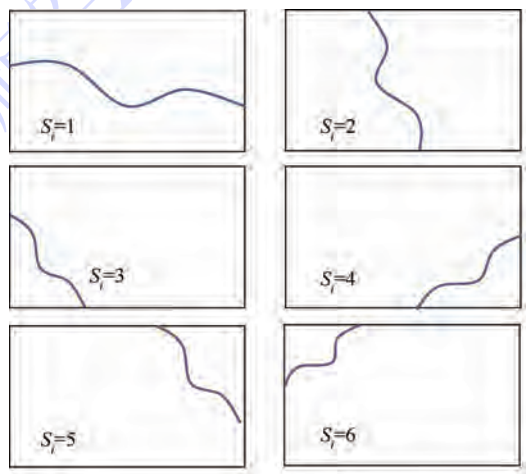


图 3 曲筋布局形式(横向为板长、纵向为板宽)

Fig. 3 Configurations of curvilinear blade-stiffened panels (horizontal edges represent plate length, longitudinal edges represent plate width)

北航学报 赠阅

1.2.1　设计变量

在等质量条件下,将加筋板临界屈曲载荷作为优化目标. 记第 i 条曲线第 j 个数据点坐标为 (x_{ij},y_{ij}),将 y_{ij} 及第 i 条曲线两端点的切线导数 f'_{is} 和 f'_{ie} 作为设计变量. 另外,将第 i 条曲线的布局形式 S_i、筋条高度 h_{si}、筋条厚度 t_{si} 作为设计变量,则筋条数目为 n、曲线数据点数目为 m 的加筋板共有 $(m+5)n$ 个参数.

若将加筋板质量作为优化目标,必须将板厚 t_p 也作为设计变量.

1.2.2　约束条件

1) 几何约束条件:

$$\begin{cases} 0 \leqslant y_{ij} \leqslant 1 \\ S_i = 1,2,\cdots,6 \end{cases} \tag{6}$$

其中 y_{ij} 将根据 S_i 的不同进行坐标映射,变换为真实的坐标值.

2) 考虑到工程需求:

$$\begin{cases} h_s^{\min} \leqslant h_{si} \leqslant h_s^{\max} \\ t_{si} \geqslant 0.2 \\ t_{si} \geqslant t_p \\ h_{si} \geqslant 2t_{si} \\ t_p \geqslant 0.2 \end{cases} \tag{7}$$

另外,若将临界屈曲载荷作为目标函数,需满足等质量约束:

$$m = \rho abt_{p0} \tag{8}$$

式中,a 为板长;b 为板宽;ρ 为密度;t_{p0} 为初始板厚,优化过程中大小不变.

3) 加筋板需要满足稳定性和强度要求.

① 第三强度准则:von Mises 等效应力最大值 σ_{vm} 小于材料屈服载荷 σ_y,即

$$\sigma_{vm} < \sigma_y \tag{9}$$

或者筋条的强度约束可采用式(10):

$$\sigma_s < F_{cc} \tag{10}$$

式中,σ_s 为筋条的最大主应力;F_{cc} 为片型筋条的最大允许主应力. F_{cc} 由式(11)算出.

$$F_{cc} = \begin{cases} \sigma_y & 0.615\,25\left(\dfrac{h\sqrt{\sigma_y/E}}{t}\right)^{-0.783\,87} > 1 \\ 0.615\,25\,\sigma_y\left(\dfrac{h\sqrt{\sigma_y/E}}{t}\right)^{-0.783\,87} & \\ & 0.615\,25\left(\dfrac{h\sqrt{\sigma_y/E}}{t}\right)^{-0.783\,87} \leqslant 1 \end{cases} \tag{11}$$

式中,E 为弹性模量;h 为筋条高度;t 为筋条厚度. 式(11)由文献[9]中的曲线获得,适用于任何材料一般的挤压部件.

② 稳定性条件:要求实际载荷小于临界屈曲载荷,在有限元特征分析里面,可表示为

$$\lambda_0 \geqslant 1 \tag{12}$$

式中,λ_0 为加筋板一阶模态的特征值.

1.2.3　目标函数

若以临界屈曲载荷为目标函数,其值由 Abaqus odb 结果文件中获取. 若以加筋板质量为目标函数,由式(13)求出.

$$m = \rho abt_p + \rho\sum h_{si}t_{si}\int\sqrt{R'^2(x)+1}\,\mathrm{d}x \tag{13}$$

式中,$R'(x)$ 为样条插值函数一阶导.

2　计算实例

2.1　筋端相聚形式的曲筋优化

所谓筋端相聚形式,是指 2 条或多条曲筋的加载边聚于一点的形式. 由于筋条只能承受面内载荷,若 2 条筋条的加载边聚到一起,外载荷可以自动分配到 2 条曲筋面内. 基于这种想法,提出了一种由 2 条曲筋构成的端点闭合的新构型. 基本想法是用 2 条端点闭合的曲筋代替一条直筋,以此来获得更好的稳定性. 端点闭合的曲筋构型如图 4 所示.

图 4　端点闭合的曲筋构型

Fig. 4　Configurations of collecting loaded edges of stiffeners

以图 5 给定的矩形加筋板作为优化实例. 优化目标选择临界屈曲载荷,初始板厚设为 2.6 mm,材料选用 2039 铝合金. 假设筋条中单位长度上的分布载荷与板边分布载荷相同. 加筋板边界条件采用简支,如图 5 所示,其中 v 为 y 向位移,w 为 z 向位移.

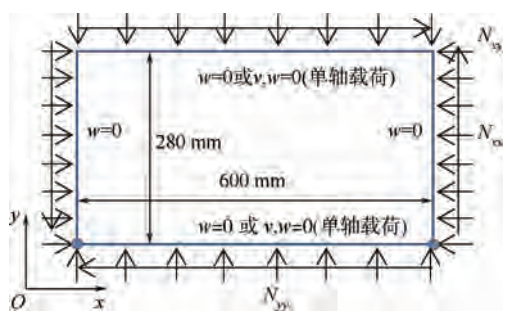


图 5　加筋板几何尺寸与边界条件

Fig. 5　Geometries and boundary conditions

2.1.1 单轴载荷下的优化结果

设定 x 向载荷 $N_{xx}=1$ N，y 向载荷 N_{yy} 和剪切载荷 N_{xy} 为 0. 假定筋条端点给定，取筋条高度 h_{si}，筋条厚度 t_{si}，曲筋斜率 f'_{1s} 和 f'_{1e} 为设计变量，f'_{2s} 和 f'_{2e} 根据对称性由 f'_{1s} 和 f'_{1e} 算出. 2139 铝合金材料属性如表 1 所示，设计变量约束范围如表 2 所示，单轴载荷下的优化结果如表 3 所示. 其中，ν 为泊松比，σ 为屈服强度，n 为迭代次数.

表 1 2139 铝合金材料属性

Table 1 Material properties of 2139 aluminum

参数	数值
E/GPa	73.085
ν	0.33
σ/MPa	427.47
ρ/(kg·m^{-3})	2 700

表 2 设计变量约束范围

Table 2 Range of design variables

设计变量	下界	上界
f'_{is}	−3	3
f'_{ie}	−3	3
h_{si}/mm	5	50
t_{si}/mm	0.2	10

表 3 单轴载荷下的优化结果

Table 3 Optimization results for single-axial load

变量	光板	1 条直筋	2 条曲筋	2 条直筋	4 条曲筋
m/kg	1.179 36	1.179 36	1.179 36	1.179 36	1.179 36
σ_y/N	11 466	37 165	40 503	56 169	60 388
σ_{vm}/kPa	13 900	52 167	84 152	69 755	126 532
σ_s/kPa	0	16 828	43 502	21 357	99 348
n		435	2 004	328	1 693
h_{si}/mm		30.76	30.82	36.26	35.29
t_{si}/mm		2.343 2	2.122 2	2.221 4	1.718 1
t_p/mm	2.6	2.342 6	2.112 5	2.024 7	1.710 4
f'_{1s}		0	−0.603 8	0	0.425 5
f'_{1e}		0	0.426 9	0	−0.390 4

单轴载荷下光板和单直筋的一阶模态如图 6 所示，单轴载荷下双曲筋和双直筋的一阶模态如图 7 所示. 图 8 所示为单轴荷下 4 条曲筋的一阶模态. 由优化结果可看出，2 条端点闭合的曲筋比单直筋板获得了 8.98% 的临界屈曲载荷提高；4 条端点闭合的曲筋比双直筋板获得了 7.51% 的临界屈曲载荷提高.

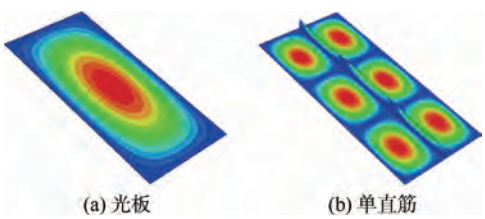


图 6 单轴载荷下光板和单直筋的一阶模态

Fig. 6 First step mode of bare board and one straight stiffener board for single-axial load

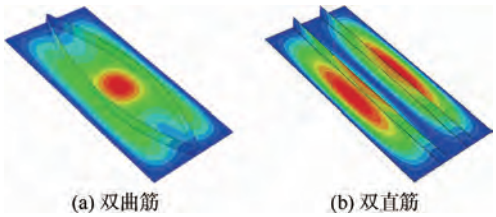


图 7 单轴载荷下双曲筋和双直筋的一阶模态

Fig. 7 First step mode of two curvilinear stiffeners board and two straight stiffeners board for single-axial load

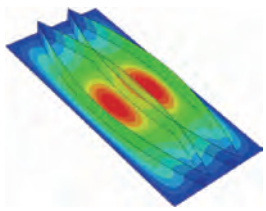


图 8 单轴载荷下 4 条曲筋的一阶模态

Fig. 8 First step mode of four curvilinear stiffeners board for single-axial load

2.1.2 组合载荷下的优化结果

对于组合载荷的情况，设定 x 向载荷 N_{xx}，y 向载荷 N_{yy} 和剪切载荷 N_{xy} 比值为 1∶1∶1. 这时允许筋条端点在板边自由移动，因此增加了一个设计变量 y_{11}，其他端点坐标值根据对称性由 y_{11} 算出. 组合载荷下的优化结果如表 4 所示.

表 4 组合载荷下的优化结果

Table 4 Optimization results for multi-axial load

变量	光板	1 条直筋	2 条曲筋	2 条直筋	4 条曲筋
m/kg	1.179 36	1.179 36	1.179 36	1.179 36	1.179 36
σ_y/N	5 021	12 995	15 635	17 109	21 935
σ_{vm}/kPa	13 793	41 041	82 583	63 069	114 560
σ_s/kPa	0	2 459	36 799	5 403	61 168
n		407	1 021	997	1 594
h_{si}/mm		32.27	27.91	43.82	26.39
t_{si}/mm		2.332 1	2.148	1.980 4	1.858 6
t_p/mm	2.6	2.331 3	2.148	1.980 1	1.856 4
f'_{1s}		0	−0.593 6	0	−0.621 8
f'_{1e}		0	0.593 6	0	−0.621 8
y_{11}/mm		150	150	205.54	116.67

组合载荷下光板和单直筋的一阶模态如图 9 所示，组合载荷下双曲筋和双直筋的一阶模态如图 10 所示. 组合载荷下 4 条曲筋的一阶模态如图 11所示. 由优化结果可以看出，提出的这种端点闭合的双曲筋板构型比单直筋板临界屈曲载荷增加了 20.3%，四曲筋板构型比双直筋板临界屈曲载荷增加了 28.2%，说明曲筋板更适用于复杂载荷.

2.2 筋条垂直于板边情况下的曲筋板优化

由于片型筋条只能承受面内载荷，为了更方

北京航空航天大学学报

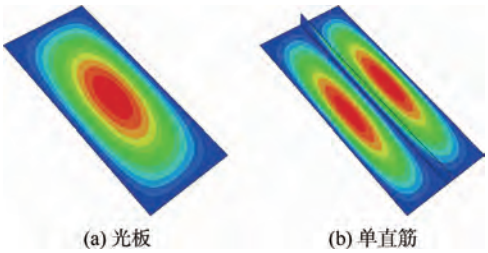


图 9　组合载荷下光板和单直筋的一阶模态

Fig. 9　First step mode of bare board and one straight stiffener board for multi-axial load

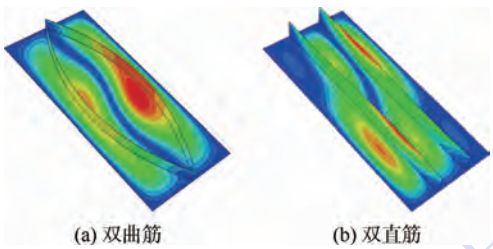


图 10　组合载荷下双曲筋和双直筋的一阶模态

Fig. 10　First step mode of two curvilinear stiffeners board and two straight stiffeners board for multi-axial load

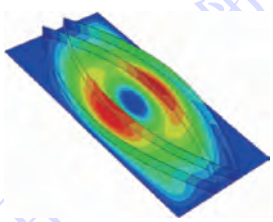


图 11　组合载荷下 4 条曲筋的一阶模态

Fig. 11　First step mode of four curvilinear stiffeners board for multi-axial load

便地施加载荷于筋条上,优化过程约定筋条垂直于板边. 设置筋条布局使外载平衡的基础上,研究筋条数目都为 2 情况下的曲筋加筋板与直筋加筋板的优劣. 板大小、外载及边界条件与 2.1 节的相同.

2.2.1　单轴载荷下的优化结果

令筋条 2 个端点可在板的两加载边移动,即对应于图 3 中 $S_{1,2}=\{1,1\}$ 的布局形式,这里令每条曲筋设 1 个中间插值点,取筋条高度 h_{si},筋条厚度 t_{si},y_{11},y_{12} 为设计变量,其他端点坐标值根据对称性由 y_{11} 和 y_{12} 算出. 单轴载荷下的优化结果如表 5 所示.

单轴载荷下光板和双直筋的一阶模态如图 12 所示,单轴载荷下双曲筋的一阶模态如图 13 所示. 优化结果表明,曲筋加筋板的优化结果趋近于直筋加筋板,且模态相近,说明在单轴载荷条件下直筋是最优的.

表 5　单轴载荷下的优化结果

Table 5　Optimization results for single-axial load

变量	光板	双直筋	双曲筋
m/kg	1.179 36	1.179 36	1.179 36
σ_y/N	11 466	62 327	62 899
σ_{vm}/kPa	13 900	88 130	92 446
σ_s/kPa	0	1 478	16 546
n		1 146	1 854
h_{si}/mm		38.08	38.22
t_{si}/mm		2.044 8	2.043 1
t_p/mm	2.6	2.043 8	2.042 2
f'_{1s}		0	0
f'_{1e}		0	0
y_{11}/mm		93.88	94.49
y_{12}/mm		93.88	97.00

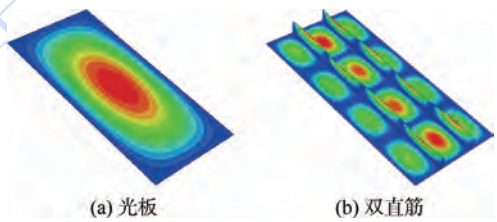


图 12　单轴载荷下光板和双直筋的一阶模态

Fig. 12　First step mode of bare board and two straight stiffeners board for single-axial load

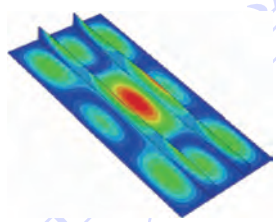


图 13　单轴载荷下双曲筋的一阶模态

Fig. 13　First step mode of two curvilinear stiffeners board for single-axial load

2.2.2　组合载荷下的优化结果

研究组合载荷即双轴载荷、剪切载荷比值为 1∶1 的情况. 共设计了 4 种曲筋布局形式,即“$S_{1,2}=\{1,1\}$ + 中间一个插值点”,“$S_{1,2}=\{3,5\}$”,“$S_{1,2}=\{3,2\}$”,“$S_{1,2}=\{1,1\}$ + 中间无插值点”. 其中,后 2 种曲筋布局由于没有对称性要求,增加设计变量 y_{21}. 优化结果如表 6 所示.

组合载荷下光板和双直筋的一阶模态如图 14 所示,组合载荷下 $S_{1,2}=\{1,1\}$ + 中间一个插值点布局和 $S_{1,2}=\{3,5\}$ 布局的一阶模态如图 15 所示. 图 16 所示为组合载荷下 $S_{1,2}=\{3,2\}$ 布局和 $S_{1,2}=\{1,1\}$ + 中间无插值点布局的一阶模态. 优化结果表明,$S_{1,2}=\{1,1\}$ + 中间一个插值点,即筋条端点在对边双交叉点布局的临界屈曲载荷比双直筋的高出了 15.7%;$S_{1,2}=\{3,2\}$,

即筋条端点在对边和相邻边布局的临界屈曲载荷比双直筋的高出了 5.6%，这说明曲筋加筋板相比直筋加筋板拥有更好的提高板稳定性的能力.

表 6 组合载荷下的优化结果

Table 6 Optimization results for multi-axial load

变量	光板	双直筋	$S_{1,2}=\{1,1\}$ + 中间一个插值点	$S_{1,2}=\{3,5\}$	$S_{1,2}=\{3,2\}$	$S_{1,2}=\{1,1\}$ + 中间无插值点
m/kg	1.179 36	1.179 36	1.179 36	1.179 36	1.179 36	1.179 36
σ_y/N	5 021	17 109	19 795	15 211	18 066	16 716
σ_{vm}/kPa	13 793	63 069	102 650	65 622	301 447	81 850
σ_s/kPa	0	5 403	72 738	32 403	313 266	71 830
n		997	2 391	1 467	2 444	1 343
h_{si}/mm		43.82	20.76	22.71	28.38	35.09
t_{si}/mm		2.332 1	2.476 0	2.354 2	2.492 7	2.070 5
t_p/mm	2.6	1.980 4	2.111 1	2.344 1	2.276 4	2.070 4
f'_{1s}		0	0			0
f'_{1e}		0	0			0
y_{11}/mm		43.73	0	134.24	16.40	150
y_{12}/mm			264.89	251.98	142.62	124.17
y_{21}/mm				192.52	67.35	

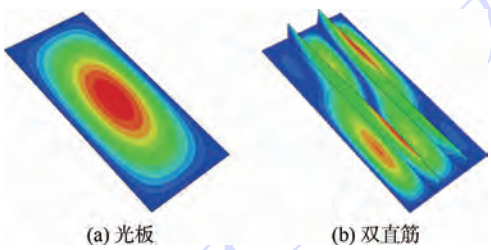


图 14 组合载荷下光板和双直筋的一阶模态

Fig. 14 First step mode of bare board and two straight stiffeners board for multi-axial load

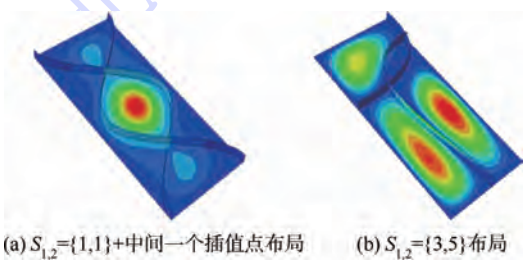


图 15 组合载荷下 $S_{1,2}=\{1,1\}$ + 中间一个插值点布局和 $S_{1,2}=\{3,5\}$ 布局的一阶模态

Fig. 15 First step mode of $S_{1,2}=\{1,1\}$ with one interpolation point and $S_{1,2}=\{3,5\}$ for multi-axial load

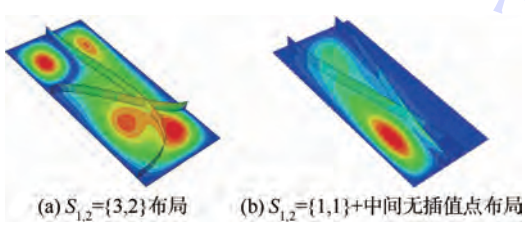


图 16 组合载荷下 $S_{1,2}=\{3,2\}$ 布局和 $S_{1,2}=\{1,1\}$ + 中间无插值点布局的一阶模态

Fig. 16 First step mode of $S_{1,2}=\{3,2\}$ and $S_{1,2}=\{1,1\}$ with no interpolation point for multi-axial load

3 结 论

设计了一种新的曲线片型加筋板优化系统，可以根据需求合理设置设计变量个数，为曲筋加筋板的优化设计提供了广阔的空间. 经过对曲筋加筋板的大量优化计算，得出以下结论：

1）在端点闭合的曲筋布局中，证明了 2 条端点闭合的曲筋比一根直筋更好地提高了板的稳定性，应用这种结构形式，可以将加筋板上的任何一根直筋替换为端点闭合的曲筋.

2）在筋条垂直于板边的布局中，研究了若干种曲筋布局方案，证明了在相同筋条数目情况下，有多种曲筋布局加筋板比直筋加筋板更具优越性. 说明了曲线片型加筋板可以获得比传统直筋加筋板更优的稳定性性能，为曲筋加筋板设计的深入研究提供了基础.

3）除稳定性之外，曲筋加筋板的后屈曲特性、疲劳特性、损伤容限和振动特性等性能也需要优化研究. 这些特性也同样影响着曲筋加筋板的发展.

4）曲筋加筋板存在很大的发展潜力，但如果将直筋加筋板合理布局，也能获得很好的稳定性性能. 由于曲筋加筋板计算量大，加工工艺复杂，所以只有制造工艺和计算机水平发展到一定水平，曲筋加筋板才能体现出它的优势.

参考文献（References）

[1] 张柱国，姚卫星，刘克龙. 基于进化 Kriging 模型的金属加筋板结构布局优化方法[J]. 南京航空航天大学学报，2008，

40(4):497-500.
Zhang Z G, Yao W X, Liu K L. Configuration optimization method for metallic stiffened panel structure based on updated Kriging model[J]. Journal of Nanjing University of Aeronautics and Astronautics, 2008, 40(4): 497-500(in Chinese).

[2] Hao P, Wang B, Li G. Surrogate-based optimum design for stiffened shells with adaptive sampling[J]. AIAA Journal, 2012, 50(11): 2389-2407.

[3] Mulani S B, Slemp W C H, Kapania R K. EBF3PanelOpt: an optimization framework for curvilinear blade-stiffened panels[J]. Thin-Walled Structures, 2013, 63: 13-26.

[4] Mulani S B, Havens D, Norris A, et al. Design, optimization, and evaluation of Al-2139 compression panel with integral T-stiffeners[J]. Journal of Aircraft, 2013, 50(4): 1275-1286.

[5] Kapania R K, Li J, Kapoor H. Optimal design of unitized panels with curvilinear stiffeners[C]//AIAA 5th ATIO and the AIAA 16th Lighter-than-Air Systems Technology Conference and Balloon Systems Conference, 2005, 3: 1708-1737.

[6] Dang T D, Kapania R K, Slemp W C H, et al. Optimization and postbuckling analysis of curvilinear-stiffened panels under multiple-load cases[J]. Journal of Aircraft, 2010, 47(5): 1656-1671.

[7] Mulani S B, Duggirala V, Kapania R K. Curvilinearly T-stiffened panel-optimization framework under multiple load cases using parallel processing[J]. Journal of Aircraft, 2013, 50(5): 1540-1554.

[8] 罗煦琼. 一维搜索问题的三次 B 样条插值法[J]. 浙江科技学院学报, 2008, 20(1): 1-3.
Luo X Q. Cubic B-spline interpolation for one-dimensional search method[J]. Journal of Zhejiang University of Science and Technology, 2008, 20(1): 1-3(in Chinese).

[9] Özakça M, Murphy A, van der Veen S. Buckling and post-buckling of sub-stiffened or locally tailored aluminium panels[C]//25th Congress of the International Council of the Aeronautical Sciences 2006. Red Hook, NY: Curran Associates Inc, 2006, 4: 2367-2382.

Design and optimization of improving stability of curvilinear blade-stiffened panels

XU Yuanming*, WANG Dong

(School of Aeronautic Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: Stiffened panels are widely used in the aerospace field as a typical structure. To further expand the design space of straight stiffened panels, an optimization framework for curvilinear blade-stiffened panels was developed based on ModelCenter optimization environment, Catia and Abaqus. Stiffeners were generated by equally-split cubic B-spline interpolation function, which had no limit to the number of parameters, thus providing a sufficiently large optimization space for curvilinear stiffened panels. On the other hand, two methods were taken to load on stiffeners including collecting loaded edges of stiffeners and making stiffeners perpendicular to panel edges and some new configurations of stiffened panels were proposed. Optimization results show that for single-axle load, curvilinear stiffened panels in comparison with straight stiffened panels make very small stability improvement. But for multi-axial load, curvilinear stiffened panels have better stability than straight stiffened panels.

Key words: optimization; curvilinear blade-stiffened panels; stability; B-spline; load on stiffeners

北京航空航天大学学报

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0279

基于石墨烯膜的光纤 Fabry-Perot 腔干涉特性分析

李成*[1,2], 郭婷婷[1,2], 肖俊[1,2], 樊尚春[1,2], 靳伟[3]

(1. 北京航空航天大学 "新型惯性仪表与导航系统技术"国防重点学科实验室, 北京 100191;
2. 北京航空航天大学 仪器科学与光电工程学院, 北京 100191; 3. 香港理工大学 电机工程学系, 九龙)

摘 要: 针对以新型材料石墨烯为膜片的光纤法珀压力传感器,应用圆薄膜大挠度弹性理论,利用有限元法分析了均布载荷下石墨烯膜的挠度形变;并基于 Fabry-Perot 干涉仪原理,建立了光纤 Fabry-Perot 腔压力传感的数学模型. 根据石墨烯膜折射率特性,分析了层数、入射光角度等参数对石墨烯膜反射率的影响,获取了腔长损耗以及薄膜挠度形变导致腔长变化而引起的干涉光谱变化规律. 仿真结果表明,增加薄膜层数可提高反射率、改善干涉性能;但随着载荷增加,其对挠度形变的影响表现为反向递减效应. 8 层石墨烯薄膜可获得 0.715% 的反射率,且当腔长为 40 μm 时,直径 25 μm 薄膜的理论压力灵敏度约为 10 nm/kPa. 这为基于多层石墨烯的膜片式光纤压力传感器的设计提供了理论依据.

关 键 词: 光纤 Fabry-Perot 腔; 石墨烯膜; 大挠度; 干涉特性; 压力传感器
中图分类号: TB34
文献标识码: A **文章编号**: 1001-5965(2015)04-0574-06

2004 年,英国曼彻斯特大学 Novoselov 和 Geim 等用微机械剥离方法首次获得了单层石墨烯薄膜[1]. 作为自然界中已知的最薄的膜材料,单层石墨烯的厚度为 0.335 nm,其断裂强度为 42 N/m,是不锈钢的 100 倍,弹性模量接近 1 TPa,弹性延展率可达 20%,其不透明度为(2.3 ± 0.1)%,反射率小于 0.1%;但不透明度随薄膜层数的增加而变大,梯度约为 2.3%[2-4]. 随着石墨烯制备及转移方法的不断发展[5-6],石墨烯薄膜优良的机械特性和光学特性受到了研究者的广泛关注. 2008 年,美国康奈尔大学 Bunch 等对石墨烯薄膜的不透气性进行了理论与实验研究[7],这为石墨烯材料用于压力传感器的敏感薄膜提供了可能. 2012 年,瑞典 KTH 皇家理工学院 Smith 等提出了悬浮石墨烯压力传感器的结构与制造工艺,并运用喇曼光谱仪对石墨烯的变形进行了分析,在 13 978 Pa 压力下单层石墨烯的最大变形可达 280 nm[8]. 但所用 SiO_2 与 SiC 基底的性质逊于石墨烯的内在性质,对传感器的结构与功能产生一定局限性. 2013 年,韩国世明大学 Kwon 等利用顶板结构借助压力传递效应,提出了一种改进的悬浮石墨烯压力传感器[9],其在低压范围内具有较好的线性度,但在高压范围内线性度较差,且结构尺寸相对于一般悬浮石墨烯压力传感器较大. 2012 年,香港理工大学 Ma 等以直径 125 μm 的石墨烯膜为感压元件制作了光纤 F-P 腔压力传感器,其灵敏度为 1 100 nm/kPa[10-11].

以上理论与实验研究表明,作为压力敏感薄膜,石墨烯具有比硅材料更高的灵敏度和抗过载能力,在高灵敏度动态压力测量方面具有较高的

收稿日期: 2014-05-19; **录用日期**: 2014-06-12; **网络出版时间**: 2014-07-08 09:53
网络出版地址: www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0279.html
基金项目: 国家自然科学基金资助项目(61273060,61121003); 长江学者和创新团队发展计划资助项目(IRT1203); 中央高校基本科研业务费专项资金资助项目(YWF-13-B03-C11,YWF-14-YQGD-004); 北航研究生创新实践基金资助项目(YCSJ-01-2014-11); 中国空间技术研究院 CAST 创新基金资助项目
* **通讯作者**: 李成(1977—),男,天津人,副教授,licheng@buaa.edu.cn,主要研究方向为先进传感与智能仪器.

引用格式: *李成, 郭婷婷, 肖俊, 等. 基于石墨烯膜的光纤 Fabry-Perot 腔干涉特性分析[J]. 北京航空航天大学学报, 2015, 41(4): 574-579. Li C, Guo T T, Xiao J, et al. Interference characteristics analysis of optical fiber Fabry-Perot cavity with graphene diaphragm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(4): 574-579 (in Chinese).*

北航学报 赠阅

应用前景,但这方面的研究才刚起步. 本文基于圆薄膜大挠度理论,借助有限元力学分析,对石墨烯膜的压力-挠度进行了建模仿真,并针对挠度形变的高灵敏度检测,分析了石墨烯膜对光纤 F-P 腔压力传感特性的影响.

1　F-P 腔压力敏感模型

图 1 为基于石墨烯膜的光纤 F-P 腔结构示意图. 当光纤内入射光垂直入射到光纤端面时,一部分光被光纤端面反射,剩余光透射进入 F-P 腔体,在 F-P 腔体内产生多光束干涉,从而形成光纤F-P 腔传感器. 作为该传感器敏感单元的石墨烯膜感受外界均布载荷 q,由于单层石墨烯膜厚约 0.335 nm,在均布压强下其挠度形变远大于膜厚,挠度特性表现为非线性.

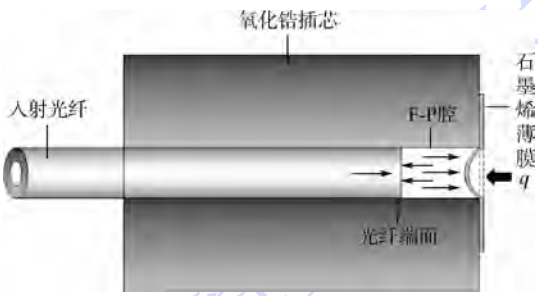


图 1　F-P 腔结构示意图

Fig. 1　Schematic diagram of F-P cavity

考虑到薄膜应力使膜片弯曲,而弯曲曲率半径与薄膜应力、膜片特性、厚度密切相关,1959 年,Beams 通过鼓泡实验获取沉积在基体上的薄膜力学性能,提出了考虑预应力 σ_0 条件下的球壳模型,即当圆形薄膜受均布面载荷时薄膜发生均匀对称的膨胀,其挠度满足的公式为[12]

$$q=\frac{8Ed\omega^3}{3(1-\mu)r^4}+\frac{4\sigma_0 d}{r^2}\omega \tag{1}$$

式中,E 为弹性模量;d 为薄膜厚度;μ 为泊松比;r 为圆薄膜半径;ω 为薄膜中心挠度. 为此,利用 ANSYS 有限元法,构建了压力作用下的石墨烯膜中心挠度的仿真模型.

为此,考虑光纤 F-P 腔压力传感器的微小型设计,结合文献[10]的石墨烯膜参数,以半径为 12.5 μm 的单层石墨烯圆薄膜为分析对象. 该膜片边界条件为周边固支,材料参数选用 Lee 等通过原子力显微镜实测的结果[3],即 E,σ_0,d,μ 分别为 1 TPa,1.2 GPa,0.335 nm,0.17. 在 0 ~ 50 kPa 范围内分别对预应力作用下的 1,3,5,8 层石墨烯圆膜片施加均布载荷,则石墨烯膜均布载荷下中心挠度特性仿真结果如图 2 所示. 可知,Beams 模型的解析解与 ANSYS 仿真解之间的平均相对误差为 0.49%,从而表明式(1)可用于分析均布载荷下石墨烯膜的中心挠度特性. 这样,当探头受到外界均布载荷 q 时,薄膜中心挠度 ω 使腔长由初始值 L_0 缩短至 L,即

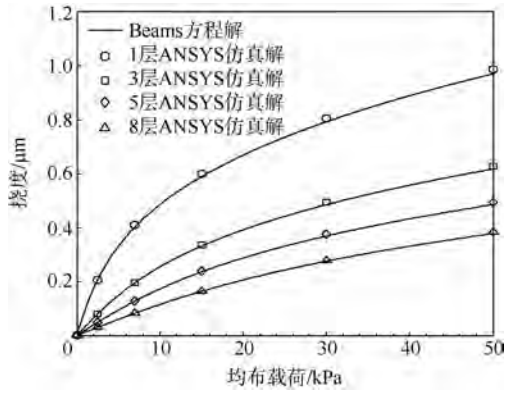


图 2　石墨烯膜均布载荷下中心挠度特性仿真

Fig. 2　Deflection characteristic simulation under uniform load of single-layer graphene film

$$\omega=L_0-L \tag{2}$$

则 F-P 腔输出的多光束干涉光强 I_r 为[13]

$$I_r=\frac{R_1+\xi R_2-2\sqrt{\xi R_1 R_2}\cos\delta}{1+\xi R_1 R_2-2\sqrt{\xi R_1 R_2}\cos\delta}I_i \tag{3}$$

式中,I_i 为入射光强;R_1,R_2 分别为 F-P 腔内光纤端面和石墨烯薄膜的反射率,R_1 取实测值 0.02,R_2 由薄膜特性矩阵确定;δ 为无半波损失时的相邻光束的相位差;ξ 为 F-P 腔长耦合系数. 特别地,当光垂直入射,且腔体内介质为空气时,δ 可表示为

$$\delta=\frac{4\pi L}{\lambda} \tag{4}$$

式中,λ 为入射光波长. 由式(3)可知,干涉光谱为近似等周期的余弦函数,当 $\delta=(2m+1)\pi$,$m=0,1,\cdots$ 时 I_r 取得峰值. 取对应干涉光谱中级次为 m 和 $m+k$ 的峰值,则

$$\begin{cases}\dfrac{4\pi L}{\lambda_m}=(2m+1)\pi\\[2ex]\dfrac{4\pi L}{\lambda_{m+k}}=[2(m+k)+1]\pi\end{cases} \tag{5}$$

$$L=\frac{k}{2}\left(\frac{\lambda_{m+k}\lambda_m}{\lambda_m-\lambda_{m+k}}\right)\qquad k=1,2,\cdots \tag{6}$$

因此,由外界均布载荷 q、薄膜中心挠度 ω、F-P腔腔长 L 等构成的压力传感模型可由式(1)、式(2)和式(6)联立确定.

2　特性仿真

2.1　石墨烯薄膜厚度对反射率的影响

由于入射光射入石墨烯薄膜存在一定的角

北航学报 赠阅

度,分别考虑薄膜对 p-偏振光和 s-偏振光的反射率 R_p,R_s. 根据薄膜光学理论,可得薄膜及其外侧空气的特性矩阵为[14]

$$\begin{cases}\begin{bmatrix}Z_1\\Z_2\end{bmatrix}=\begin{bmatrix}\cos\alpha & \dfrac{\mathrm{i}}{\eta_1}\sin\alpha\\ \mathrm{i}\eta_1\sin\alpha & \cos\alpha\end{bmatrix}\begin{bmatrix}1\\\eta_2\end{bmatrix}\\ \alpha=\dfrac{2\pi}{\lambda}N_1 d\cos\theta_1\end{cases} \tag{7}$$

式中,i 为虚数单位;η_1 为石墨烯薄膜的有效导纳;η_2 为空气的有效导纳;薄膜厚度 d = 薄膜层数 × 单层膜厚;N_1 为石墨烯薄膜复折射率. 对于 p-偏振光,$\eta_1 = N_1/\cos\theta_1$,$\eta_2 = 1/\cos\theta_2$(空气折射率 ≈ 1),而对于 s-偏振光,$\eta_1 = N_1\cos\theta_1$,$\eta_2 = \cos\theta_2$,其中,$\theta_1$ 为空气界面与薄膜的折射角,θ_2 为薄膜与空气界面的出射角.

且石墨烯膜的反射系数可表示为[14]

$$r=\frac{\eta_2 Z_1 - Z_2}{\eta_2 Z_1 + Z_2}=\frac{\mathrm{i}\left(\dfrac{\eta_2^2}{\eta_1}-\eta_1\right)\sin\alpha}{2\eta_2\cos\alpha+\mathrm{i}\left(\dfrac{\eta_2^2}{\eta_1}+\eta_1\right)\sin\alpha} \tag{8}$$

将 p-偏振光和 s-偏振光对应的不同有效导纳代入式(8),可求取相应的反射系数 r_p,r_s 及其共轭复数 r_p^*,r_s^*,则石墨烯膜反射率 R_2 由式(9)确定:

$$R_2=\frac{r_p r_p^* + r_s r_s^*}{2} \tag{9}$$

为简化分析,假设入射光垂直入射到薄膜,则对于 p-偏振光和 s-偏振光,对应介质的有效导纳相等,且有 $\eta_1 = N_1$,$\eta_2 = 1$,$\alpha = 2\pi(N_1 \cdot d)/\lambda$,$r_p = r_s = r$,则式(9)可简化为

$$R_2 = r \cdot r^* \tag{10}$$

式中

$$r=\frac{\mathrm{i}\left(\dfrac{1}{N_1}-N_1\right)\sin\alpha}{2\cos\alpha+\mathrm{i}\left(\dfrac{1}{N_1}+N_1\right)\sin\alpha}$$

由式(10)可知,石墨烯薄膜的反射率 R_2 与其复折射率 N_1 有关. 取 $N_1 = n_1 - \mathrm{j}k_1$,其中 j 为虚数单位,$n_1$ 为折射率,k_1 为消光系数,参数 n_1,k_1 随入射光波长的变化参考基于椭圆偏振法的石墨烯膜折射率测量实验确定[15],则不同入射光波长下薄膜层数对反射率的影响如图 3 所示.

当石墨烯膜为 9 层以下时,其反射率低于 1%. 这表明在 F-P 腔内石墨烯膜的反射光强度很弱,入射光强几乎被透射或损耗. 但随膜厚增加,其反射率随之单调变大. 而且,以 8 层石墨烯膜为例,当波长为 1 600 nm 和 1 550 nm 时薄膜反射率分别为 0.718 9% 和 0.715 4%,两者相差仅为 0.003 5%. 这样,当自由光谱范围 FSR 较小时可忽略入射光波长对反射率的影响.

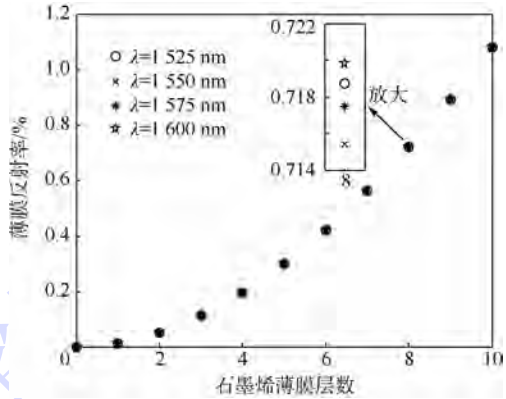


图 3 不同入射光波长下薄膜层数对反射率的影响

Fig. 3 Effects of graphene layer on film reflectivity in different wavelengths of incident light

考虑到石墨烯膜利用范德瓦尔斯力实现石墨烯薄膜与氧化锆陶瓷光纤插芯端面的吸附以制作 F-P 腔. 因此,悬浮于插芯中孔处的石墨烯膜的微小不平整性将会使入射角度发生微小改变,则入射光角度(θ_0)对石墨烯膜反射率的影响如图 4 所示.

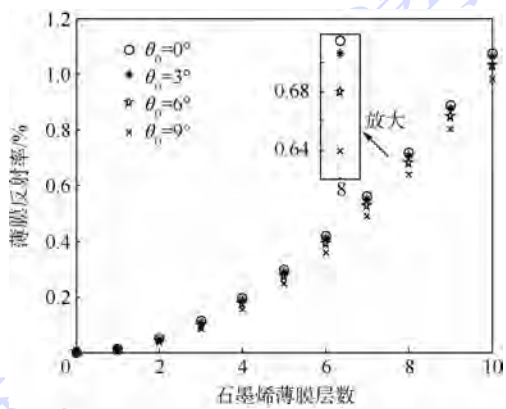


图 4 入射光角度对石墨烯膜反射率的影响

Fig. 4 Effects of incident light angle on graphene film reflectivity

仿真结果表明,随着薄膜变厚,入射光角度对薄膜反射率的影响逐渐显现,但并不明显. 以 8 层石墨烯膜为例,当入射角度(θ_0)为 0°,3°,6°,9° 时,薄膜反射率依次为 0.715 4%,0.706 3%,0.679 8%,0.638 9%. 这表明在 F-P 腔内石墨烯薄膜反射率随入射光角度的增加会发生小幅减小,但减小的趋势有限. 因此,对于在内径 25 μm 的光纤凹槽以及 125 μm 陶瓷插芯内形成的微小 F-P 腔,因薄膜吸附变形而引起的入射光小角度变化对反射率的影响可忽略.

2.2　腔长损耗影响

膜片式 F-P 腔光纤传感器受腔长影响，其干涉光强度损耗可由耦合系数 ξ 确定[16]：

$$\xi=\beta^2(l)=\left[\frac{2w(l)w_0}{w_0^2+w^2(l)}\right]^2 \tag{11}$$

式中

$$w(l)=w_0\sqrt{1+\left(\frac{l}{l_R}\right)^2}$$

$$l_R=\frac{\pi N_0 w_0^2}{\lambda}$$

其中，l_R 为瑞利距离；w_0 为光场在光纤中传播的模场半径；$w(l)$ 为光在 F-P 腔中传播的模场半径；l 为光在 F-P 腔内传播的距离；N_0 为 F-P 腔中的介质折射率. F-P 腔为空气隙且光纤为 SMF-28 单模光纤，则 $N_0=1$，$w_0=4.9$ μm. 即耦合系数是入射光波长 λ 和腔长 $L(L=l/2)$ 的函数，其值越大，光强损耗越小.

图 5 为入射光波长为 1 500，1 550，1 600 nm 时 F-P腔干涉光耦合曲线. 由此可知，对于固定腔长，波长对 F-P 腔光强损耗的影响较弱. 以 $L=40$ μm为例，波长由 1 500 nm 变化至1 600 nm所引起的耦合系数仅差 0.038 1. 但腔长变化对光强损耗的影响更为明显. 以 $\lambda=1\,550$ nm 为例，当腔长大于 50 μm 时，耦合系数小于 0.6，从而使干涉信号光强大幅损耗. 为此，需选用小 F-P 腔以提高干涉光强信号的检测性能.

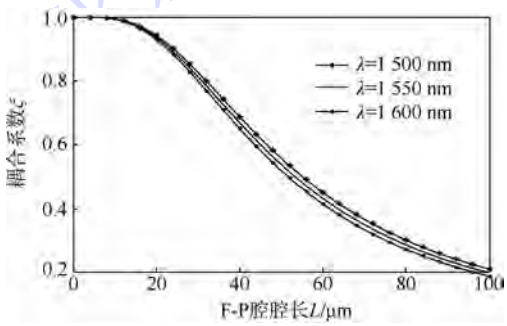


图 5　F-P 腔干涉光耦合曲线

Fig. 5　Interference light coupling curves of Fabry-Perot cavity in different lengths

取无量纲相对光强为 $10\lg(I_r/I_i)$，由式(3)可获取干涉光谱曲线，则当入射波长为 1 550 nm 时，1，3，5，8 层石墨烯薄膜(反射率分别为 0.013%，0.112%，0.297%，0.715%)的干涉光谱如图 6 所示. 图中 n 表示层数；Y，N 表示有无考虑腔长损耗. 仿真结果表明，波长变化引起的腔长损耗主要影响干涉光谱的相对光强峰值，不改变曲线形状，且这种影响与薄膜厚度密切相关. 通过增加膜厚可使薄膜反射率单调正向变化，改善干涉光谱的对比度，提高检测信号的干涉效果.

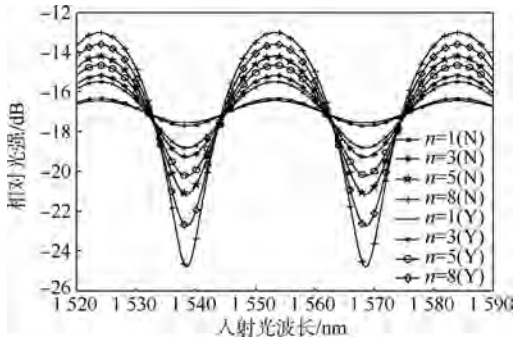


图 6　不同层厚石墨烯薄膜的干涉光谱

Fig. 6　Interference spectra of graphene film with different thicknesses

2.3　均布载荷对 F-P 腔干涉的影响

针对均布载荷对石墨烯膜光纤 F-P 腔信号检测的影响，结合获取的石墨烯膜机械力学特性(第 1 节)和光学反射特性(2.1 ~ 2.2 节)，以单层和 8 层的周边固支圆薄膜为例，取 F-P 初始腔长为 40 μm，图 7 为均布载荷下单层及 8 层石墨烯膜的 F-P 腔干涉光谱. 由此可知，对周边固支的直径为 25 μm 的圆薄膜施加均布载荷，在 1 520 ~ 1 590 nm 范围内光纤信号表现为干涉光谱特征，且随着载荷以 10 kPa 为间隔由 0 递增至50 kPa，其腔长变化量 ΔL 及干涉光谱峰值对应的波长偏移量 $\Delta\lambda$ 如表 1 所示.

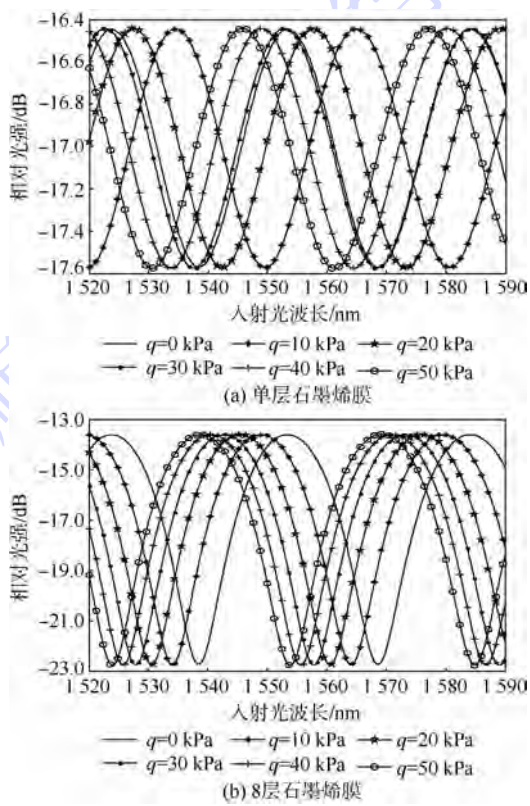


图 7　均布载荷下单层及 8 层石墨烯膜的 F-P 腔干涉光谱

Fig. 7　Fabry-Perot interference spectra for single-layer and 8-layer graphene film under uniformly distributed loads

表 1 不同均布载荷下波长与腔长的变化

Table 1 Change of incident light wavelength and F-P cavity under uniformly distributed loads

载荷/kPa	单层		8 层	
	ΔL/nm	$\Delta\lambda$/nm	ΔL/nm	$\Delta\lambda$/nm
10	485	19.21	115	4.47
20	668	26.46	206	8.00
30	791	31.33	275	10.68
40	888	35.17	331	12.85
50	968	38.34	377	14.64

仿真结果表明,一方面,增加薄膜厚度将降低挠度变形及其干涉光谱峰值处波长的偏移量,但另一方面,随着薄膜层数的增加,石墨烯压力-位移特性曲线的曲率变小,使非线性得到相应改善,不过对压力的敏感程度则相对降低. 特别地,对于所用 40 μm 腔长,单层石墨烯膜 F-P 腔的干涉对比度为 0.128,其明显低于 8 层石墨烯膜的 0.782. 如将图 7 中干涉信号光谱转化至对数域,则相对光强的峰-峰值分别约为 1.12 dB 和 9.13 dB. 考虑到光谱仪 AQ6370C 的 0.1 dB/Div 功率精度,双峰法可分辨出 8 层石墨烯膜的 F-P 干涉光谱,则由直径 25 μm 的石墨烯膜挠度特性,其理论压力灵敏度约为 10 nm/kPa. 而且,通过调整薄膜结构尺寸及薄膜与 F-P 腔基体的吸附性能可有效改善传感器灵敏度与干涉光谱信号提取.

3 结 论

1) 针对超薄石墨烯薄膜在高灵敏度压力传感中的潜在应用,根据薄膜大挠度理论,应用 Beams 球壳模型和 ANSYS 有限元方法,对石墨烯膜大挠度力学特性进行了数值解析与有限元分析,两者间 0.49% 的平均相对误差验证了 Beams 球壳模型用于石墨烯薄膜大挠度形变特性分析的有效性.

2) 考虑光纤 F-P 腔的多光束干涉效应,仿真求取了膜厚、入射光的波长、角度等参数对薄膜反射率以及干涉光谱的影响规律. 结果表明,膜厚对膜片中心挠度形变及其干涉光谱强度起主要影响,但 2 种影响表现为相反的作用效果.

3) 根据石墨烯膜的干涉对比度仿真分析,直径 25 μm 的 8 层石墨烯圆薄膜具有 0.715% 的反射率,当 F-P 腔长设计为 40 μm 时其可实现约为 10 nm/kPa 的理论压力灵敏度.

参考文献 (References)

[1] Novoselov K S, Geim A K, Morozov S V, et al. Electric field effect in atomically thin carbon films[J]. Science, 2004, 306(5696): 666-669.

[2] Liang X, Sperling B A, Calizo I, et al. Toward clean and crackless transfer of graphene[J]. ACS Nano, 2011, 5(11): 9144-9153.

[3] Lee C, Wei X D, Kysar J W, et al. Measurement of the elastic properties and intrinsic strength of monolayer graphene[J]. Science, 2008, 321(5887): 385-388.

[4] Nair R R, Blake P, Grigorenko A N, et al. Fine structure constant defines visual transparency of graphene[J]. Science, 2008, 320(5881): 1308.

[5] 袁小亚. 石墨烯的制备研究进展[J]. 无机材料学报, 2011, 26(6): 561-570.
Yuan X Y. Progress in preparation of graphene[J]. Journal of Inorganic Materials, 2011, 26(6): 561-570(in Chinese).

[6] Martins L G P, Song Y, Zeng T, et al. Direct transfer of graphene onto flexible substrates[J]. Proceedings of the National Academy of Sciences, 2013, 110(44): 17762-17767.

[7] Bunch J S, Verbridge S S, Alden J S, et al. Impermeable atomic membranes from graphene sheets[J]. Nano Letters, 2008, 8(8): 2458-2462.

[8] Smith A D, Vazirs S, Delin A, et al. Strain engineering in suspended graphene devices for pressure sensor application[C]// Proceedings of the 13th International Conference on Ultimate Integration on Silicon. Grenoble, France: IEEE, 2012: 21-24.

[9] Kwon O K, Lee J H, Kim K S, et al. Developing ultrasensitive pressure sensor based on graphene nanoribbon: molecular dynamics simulation[J]. Physica E: Low-Dimensional Systems and Nanostructures, 2013, 47: 6-11.

[10] Ma J, Jin W, Ho H L, et al. High-sensitivity fiber-tip pressure sensor with graphene diaphragm[J]. Optics Letters, 2012, 37(13): 2493-2495.

[11] Ma J, Xuan H F, Ho H L, et al. Fiber-optic Fabry-Perot acoustic sensor with multilayer graphene diaphragm[J]. IEEE Photonics Technology Letters, 2013, 25(10): 932-935.

[12] Beams J W. The structure and properties of thin film[M]. New York: John Wiley and Sons, 1959: 183-190.

[13] Zhou X, Yu Q. Wide-range displacement sensor based on fiber-optic Fabry-Perot interferometer for subnanometer measurement[J]. IEEE Sensors Journal, 2011, 11(7): 1602-1606.

[14] 唐晋发, 顾培夫. 薄膜光学与技术[M]. 北京: 机械工业出版社, 1989: 5-24.
Tang J F, Gu P F. Thin film optics and technology[M]. Beijing: China Machine Press, 1989: 5-24(in Chinese).

[15] Nelson F J, Kamineni V K, Zhang T, et al. Optical properties of large-area polycrystalline chemical vapor deposited graphene by spectroscopic ellipsometry[J]. Applied Physics Letters, 2010, 97(25): 253110.

[16] Ma C, Dong B, Gong J, et al. Decoding the spectra of low-finesse extrinsic optical fiber Fabry-Perot interferometers[J]. Optics Express, 2011, 19(24): 23727-23742.

北航学报 赠阅

Interference characteristics analysis of optical fiber Fabry-Perot cavity with graphene diaphragm

LI Cheng*[1,2], GUO Tingting[1,2], XIAO Jun[1,2], FAN Shangchun[1,2], JIN Wei[3]

(1. Novel Inertial Instrument & Navigation System Technology Key Laboratory of Fundamental Science for National Defense, Beijing University of Aeronautics and Astronautics, Beijing 100191, China;
2. School of Instrumentation Science and Opto-electronics Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China;
3. Department of Electrical Engineering, The Hong Kong Polytechnic University, Jiulong, China)

Abstract: With regard to the design of optical fiber Fabry-Perot cavity using graphene as sensitive diaphragm, the deflection change under uniformly distributed loads in graphene film was analyzed by finite element method based on the large deflection elastic theory of circular film. The pressure-sensing mathematical model of optical fiber Fabry-Perot cavity with graphene diaphragm was established based on the working principle of Fabry-Perot interferometer. The effects of graphene film layer and incident light angle on the film reflectivity were obtained according to the refractive index characteristics of the film. Then the interference spectra change, caused by both the cavity length losses and the film deflection deformations under the pressure loads, were analyzed. The simulation results show that adding the film layer can increase the film reflectivity and further improve the optical interference performance; however, a decreasing effect on the deflection deformation is caused by the film layer with the increase of pressure load. Thus, an 8 layer graphene film can achieve a reflectivity of 0.715% and an approximate theoretical sensitivity of 10 nm/kPa for a 40 μm Fabry-Perot cavity length with a membrane diameter of 25 μm. It provides a theoretical basis for the design and fabrication of high-sensitivity fiber-tip pressure sensor with multiple-layer graphene diaphragm.

Key words: optical fiber Fabry-Perot cavity; graphene diaphragm; large deflection; interference characteristics; pressure sensor

北京航空航天大学学报

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0293

具有方向特性的 X 角点的亚像素检测定位

孟偲*, 薛姣

(北京航空航天大学 宇航学院, 北京 100191)

摘 要: X 角点被广泛应用于相机标定与可见光视觉跟踪系统,不仅具有位置信息,还可以具有方向信息.针对 X 角点的精确位置检测和方向检测,提出了一种利用 X 角点对称性的检测与定位算法——对称性计算算子(SC,Symmetry Calculation),简称 SC 算子.该算法主要由两部分组成.先利用对称性计算值确定 X 角点像素位置,然后基于二次曲线拟合及直线相交求亚像素位置.通过角点定位实验,仿真分析高斯噪声和畸变对定位精度及鲁棒性的影响,并与 Harris 算法及 Micron Tracker 系统进行比较.实验表明,本文方法能有效检测 X 角点,排除其他类型的角点和干扰点,具有较高的定位精度和鲁棒性.

关 键 词: X 角点; 对称性; 亚像素位置; 方向检测; 二次曲线拟合
中图分类号: TP391.4
文献标识码: A **文章编号**: 1001-5965(2015)04-0580-09

在视觉测量中,为了降低对目标检测和识别的难度,出现了各种人工标记点[1-2],其中 X 标记点因对比性强、易检测、易制作等优点,被广泛应用于相机标定与光学跟踪系统.如基于阵列 X 角点的棋盘格,被广泛应用于张正友的共面靶标相机标定法[3]中的平面靶标[4],并被集成到 Matlab 视觉工具包[5]和开源计算机视觉库 OpenCV[6]中;Micron Tracker 光学跟踪系统[7]则将 X 标记组合粘贴于工具上进行工具位姿的测量与跟踪.相比于基于红外光线的 NDI Polaris 跟踪系统,需要使用特制主动红外发光球状标记或被动反射红外小球作为检测目标,基于可见光与 X 角点标记的 Micron Tracker 视觉跟踪系统无论是相机硬件还是标记制作都具有明显的成本优势,因而已被广泛应用于手术室器械跟踪和机器人位姿跟踪.

利用 X 角点进行相机标定或位姿测量的前提是 X 角点的精确检测与定位.目前已有多种方法能实现 X 角点的检测与亚像素定位,如 OpenCV 使用通用的基于灰度分布最大曲率点检测的 Harris 算子进行棋盘格 X 角点检测[8-10].针对相机标定中棋盘格图像上 X 角点的检测,利用 X 角点特征的属性,朱枫等提出基于对称方差的 SV 检测算子[4,11],主要利用关于 X 角点对称的像素灰度值接近且 X 角点邻域内灰度变化显著的特征,该算法原理简单,但在某些情况下会出现误检,且环境的变化会导致阈值选取的不稳定;储珺等提出采用环形模板的角点检测算子[12],该算子利用 X 角点为黑白区域边界直线交点的属性,设计环形遍历模板对棋盘格图像进行遍历,通过利用遍历后图像的属性确定棋盘格角点的位置,该算法需要知道棋盘格格子的边长,然后确定环形模板的半径,对于边长变化的棋盘格或畸变的棋盘格图像会存在漏检的情况;胡海峰和侯晓微综合使用了几种算法[13],首先利用 Radon 变换检测棋盘格直线,然后利用 Harris 和 Forstner 算子对角点进行精确定位,该算法精确度较高但运算量

收稿日期: 2014-05-23; **录用日期**: 2014-07-01; **网络出版时间**: 2014-07-29 11:08
网络出版地址: www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0293.html
基金项目: 国家科技支撑资助项目(2013BAF07B02); 中央高校基本科研业务费专项资金资助项目(YWF-14-YHXY-015)
* **通讯作者**: 孟偲(1977—),男,新疆沙湾人,副教授, Tsai@buaa.edu.cn,主要研究方向为机器视觉与智能系统.

引用格式: *孟偲, 薛姣. 具有方向特性的X角点的亚像素检测定位[J]. 北京航空航天大学学报, 2015, 41(4): 580-588.*
Meng C, Xue J. Sub-pixel X-corner detection and location with directional characteristics[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(4): 580-588 (in Chinese).

大,且只适用于棋盘格角点的检测;张广军等提出利用 Hessian 矩阵进行 X 角点检测[14-15],主要利用平滑后 X 角点为曲面鞍点的属性,该检测方法精度较高、算法简单,但易受非 X 角点的干扰. 针对手术器械跟踪的 X 角点检测,魏军等提出的基于快速筛选的段测试检测方法[16],借鉴了 FAST 检测算子[17]思想和 X 角点的邻域灰度对称变化的属性,并利用直线相交方法进行亚像素定位,该算法运算量和检测结果易受图像纹理和噪声的影响.

上述方法虽都能实现 X 角点的检测与定位,但都只给出了 X 角点的位置信息. 而 X 角点特征不仅具有位置信息,还包含有方向信息. 充分利用 X 角点特征提供的信息,可简化 X 角点立体匹配与识别的难度. 如光学跟踪系统 Micron Tracker,利用 X 角点特征的位置信息及角点连线所在向量实现模板匹配和识别. 但 Micron Tracker 系统封闭,对于算法的普及推广不利,因而有必要对包含方向信息的 X 角点特征的检测定位方法进行进一步研究,以便于能在开放或自行设计的视觉系统中使用.

张广军等基于 Hessian 矩阵的检测算法,除了能给出 X 角点特征的位置信息外,还能给出 X 角点的对称轴方向(此方向与 X 角点 Hessian 矩阵[14-15]的特征方向一致),但其对称轴方向在实物测试中并不稳定,难以有效应用. SIFT 算子[18]、SURF 算子[19]除了能提供特征的位置信息外,也包含了方向信息,但 SIFT 与 SURF 算子并不是专门针对 X 角点特征检测,没有利用 X 角点特征属性,而且其方向信息在图像上比较抽象,难以适用于 X 角点特征的检测定位.

针对 X 角点特征位置与方向同时定位的需要,同时考虑角点定位及定向的精确性和鲁棒性,本文提出了基于 X 标记点对称性计算的检测方法,即 SC(Symmetry Calculation)算子. 首先,分析 X 标记点的图像特点,在此基础上提出对称性计算的数学模型,利用此数学模型计算角点的像素位置;然后通过二次曲线拟合,实现 X 角点的亚像素定位. 仿真试验和实际测试表明该方法在确定角点的位置和方向时,具有较高的精确度和鲁棒性,满足应用要求.

1　X 角点图像特征分析

X 标记点由 4 个相邻的黑白区域组成,X 角点是区域块的公共点,如图 1 所示. 以 X 角点为中心,4 块黑白区域形成两条边界线. 为后续表述方便,在此做如下定义:按逆时针方向,根据黑白变化序列定义这 2 条边界线,从黑到白跳变的边界线定义为 BW(Black-to-White)线,另一条从白到黑跳变的边界线为 WB(White-to-Black)线.

图 1 中 X 标记点具有以下图像特征:

① X 角点是 BW 线和 WB 线的交点.

② X 标记点的方向可以用 BW 线和 WB 线的方向表示.

③ X 标记点的对称性. 对其中的任意像素 p,都能找到其关于 X 角点对称的像素,该像素与 p 同时位于黑块或白块.

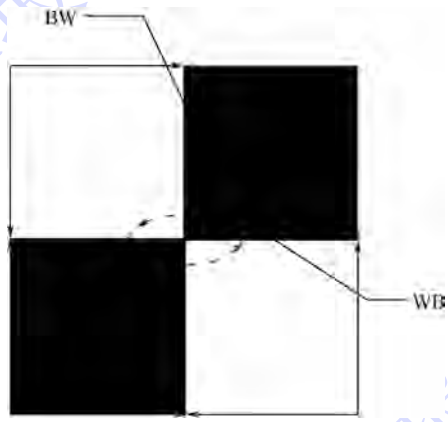


图 1　X 角点

Fig. 1　X point

中心射影变换(或小孔成像)不保证正交性,但保证共线性质. 因而在图像上,X 标记点的图像在局部仍保持对称性. 令 BW 线和 WB 线的像分别为 l_{bw},l_{wb},交点 X 的像为 x.

2　基于对称性计算的 X 角点检测

2.1　对称性计算与像素级位置确定

在位于以候选角点为中心,边长为 $2r$ 的正方形窗口上,提取像素组成首尾相连的像素段,如图 2(a)所示. 将图 2(a)中的 BW 线和 WB 线与像素段的 4 个交点,按逆时针方向分别称为 A,B,C,D 台阶;令 a,b,c,d 分别表示折线 AO_1B,BO_2C,CO_3D,DO_4A 的长度. 本文算法根据窗口的平行移动提取像素段,采样数据与 X 角点的具体方向无关. 在待测试像素段上,像素的亮度在实际情况下如图 2(b)所示. 设像素段由 n 个像素组成,记其中第 i 个像素为 $p_i(i=1,2,\cdots,n)$,$f(p_i)$ 为 p_i 的灰度值. 由于同一区域内的像素在位置和灰度级上具有较强的一致性和相关性,不同区域内的像素存在照度不均匀、对比度不同的情况. 为了保证较强的抗噪性能,并体现同一区域的整体情况,选择像素段的均值作为二值化阈值. 为了降

北航学报 赠阅

低噪声点对二值化阈值选取的影响,需先用滤波器对像素段进行预处理. 通过对像素段内灰度值进行局部二值化,将 p_i 的灰度值化为 $f'(p_i)$. 二值化后像素段内各个像素的灰度值如图 2(c)所示.

$$\text{mean} = \frac{\sum_{i=1}^{n} f(p_i)}{n} \tag{1}$$

$$f'(p_i) = \begin{cases} 0 & f(p_i) \leqslant \text{mean} \\ 255 & f(p_i) > \text{mean} \end{cases} \tag{2}$$

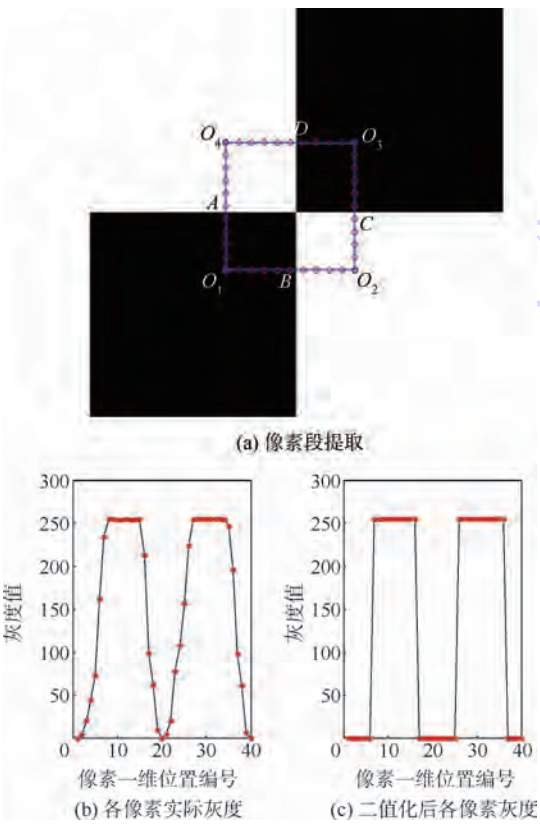


图 2 取像素段并确定台阶

Fig. 2 Taking pixel segment and determining steps

图 2(c)中的曲线反映 X 角点测试像素段内存在 4 处灰度值变化剧烈的点,即图 2(a)中介绍的 A,B,C,D 4 个台阶. 通过对二值化后像素段内的相邻元素取差分 $\Delta = f'(p_i) - f'(p_{i-1})$ $(i = 1,2,\cdots,n)$,判断 Δ 的绝对值是否为 255,从而确定此像素段内是否存在 4 个台阶及它们在像素段内的像素级位置. 然后求出 a,b,c,d.

根据 X 标记点图像特征中的对称性,可知对于当前窗口内的任意像素点 p,若要在当前窗口内找到其关于窗口中心的对称点,4 个台阶之间的距离 a,b,c,d 理论上应该相等. 由此可得对称性计算的数学公式:

$$\text{SC} = (a - c)^2 + (b - d)^2 \tag{3}$$

以图 2(a)中的每个像素为窗口中心计算对应的 SC 值,其结果如图 3 所示. 图中 x,y 坐标表示像素点的二维位置,z 坐标表示对应的对称性计算值. 可知,窗口中心越接近 X 角点,其对称性计算值 SC 越小. 所以当 SC 取得局部最小值时,即可确定当前窗口中心为角点位置.

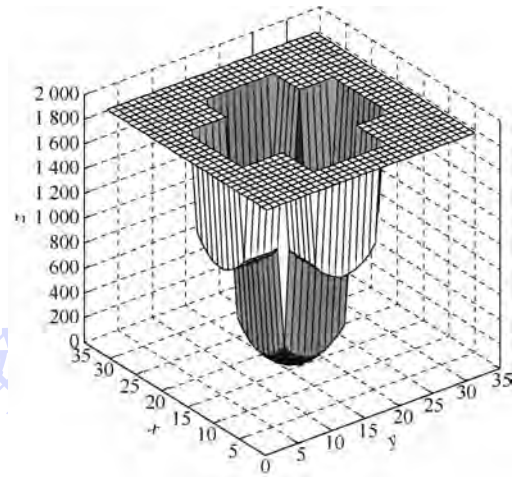


图 3 图 2(a)对称性计算结果

Fig. 3 Result of symmetry calculation of Fig. 2(a)

2.2 角点的亚像素位置计算及方向确定

通过对称性搜索得到 X 角点及其对应台阶的像素级位置,为求解其亚像素级位置,采用二次曲线拟合法[20]先求此 X 角点 4 个台阶的一维亚像素位置 A',B',C',D',然后利用边缘直线相交确定角点的亚像素位置.

在已知台阶像素位置的前提下,只选取台阶邻近点进行二次曲线拟合可以减少计算量,避免无用点影响拟合结果. 选取以台阶(假设为台阶 A)为中心的 5 个点进行二次曲线拟合. 拟合点的亮度及相邻点亮度差(梯度)如图 4(a)和图 4(b)所示. 将离散的梯度值拟合为一条连续的曲线,近似为二次抛物线,如图 4(b)所示. 二次抛物线的极值点即为沿该梯度方向灰度变化最大的地方,即为台阶的准确位置.

一个角点 $q(x,y)$ 周围存在 4 个台阶,所以要进行 4 次曲线拟合,分别得到 4 个台阶在像素段内的一维亚像素位置 A',B',C',D'. 因所得台阶存在于以角点像素位置为中心的边长为 $2r$ 的正方形窗口上,即可利用此对应关系求出其二维的亚像素位置. 假设某台阶的一维位置为 m,根据 m 值所在范围可得其对应二维坐标为

$$\begin{cases} (x - r + 0.5, y + m - r + 1) & 0 \leqslant m < 2r \\ (x + m - 3r + 1, y + r + 0.5) & 2r \leqslant m < 4r \\ (x + r + 0.5, y - m + 5r - 1) & 4r \leqslant m < 6r \\ (x - m + 7r - 1, y - r + 0.5) & 6r \leqslant m < 8r \end{cases} \tag{4}$$

将台阶的一维位置转至与图像相对应的二维

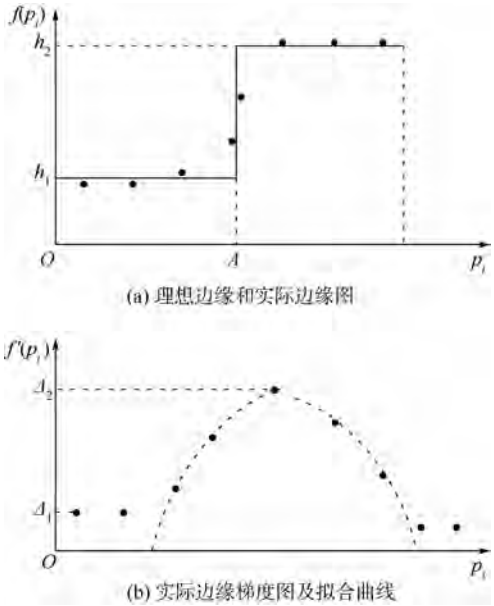


图 4　一维边缘模型

Fig. 4　Model of one-dimensional edge

位置时，由于其在垂直或水平方向上与候选角点的像素距离为 r，所以可以通过增减 r 确定在此方向上的像素位置，但实际上边缘的位置存在于像素内的任何位置，为了减小在此方向上的误差，所以增加 0.5 个像素的偏移. 一维和二维空间对应关系如图 5 所示，图中 $r=5$. 据此可求得 4 个台阶的二维亚像素位置 A''，B''，C''，D''.

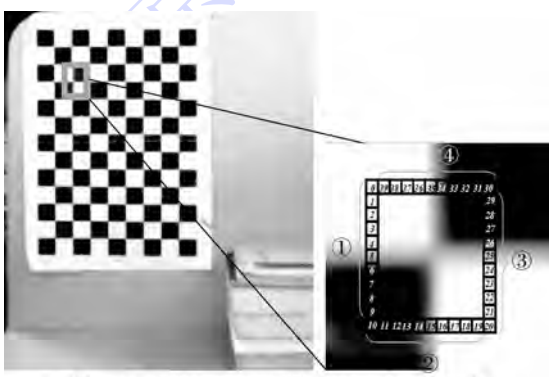


图 5　一维与二维空间对应关系

Fig. 5　Relationship between one- and two-dimensional spaces

由影射几何可知，X 角点的亚像素位置为

$$\tilde{x}_{\text{sub}} = l_{\text{bw}} \cdot l_{\text{wb}} \tag{5}$$

式中

$$l_{\text{wb}} = \tilde{A}'' \cdot \tilde{C}'' \tag{6}$$

$$l_{\text{bw}} = \tilde{B}'' \cdot \tilde{D}'' \tag{7}$$

其中，$\tilde{x}$ 为 x 的齐次坐标表示；l_{bw}，l_{wb} 为直线在图像上的齐次坐标；$\tilde{A}''$，$\tilde{B}''$，$\tilde{C}''$，$\tilde{D}''$ 是 A''，B''，C''，D'' 的图像齐次坐标.

在介绍 X 标记点的图像特征时曾说到 X 标记点的方向信息可以用 BW 线和 WB 线的方向表示，即 l_{bw} 与 l_{wb} 的方向. 通过求 l_{wb}，l_{bw} 与水平方向的夹角 θ_1，θ_2 来存储角点的方向信息，X 角点的方向如图 6 所示.

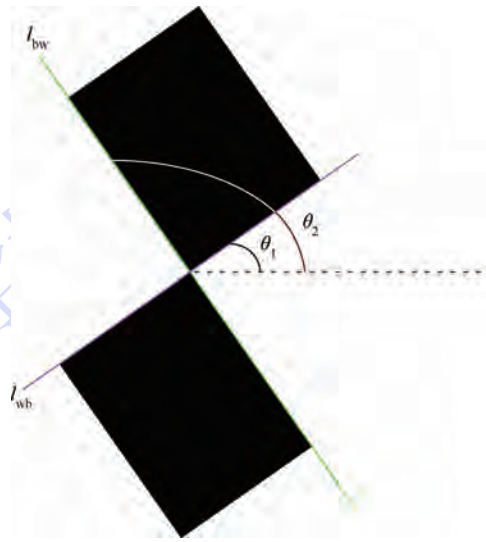


图 6　X 角点的方向

Fig. 6　Direction of X-corner point

$$\cos\theta_1 = \frac{x_C - x_A}{A''C''} \qquad \theta_1 = \arccos\frac{x_C - x_A}{A''C''} \tag{8}$$

$$\cos\theta_2 = \frac{x_D - x_B}{B''D''} \qquad \theta_2 = \arccos\frac{x_D - x_B}{B''D''} \tag{9}$$

3　角点定位实验及分析

为了对第 2 节阐述的基于对称性计算的 X 角点检测方法进行验证，利用 Visual Studio 2008 和 Matlab 7.11 对算法进行了编程实现. 实验分为 2 部分：第 1 部分是已知各个角点理论位置及角点真实方向信息的仿真实验，通过该部分实验能较便捷精确地检验算子的精确度、抗噪性能和抗畸变性能，同时在该部分实验中使用了经典 Harris 算法和本文算法对相同实验图像进行对比分析. 第 2 部分是物理实验，使用双目摄像机 BumbleBee2 BB2-08S2C，在不同环境下采集实验图像，图像尺寸大小为 1 024 × 768 像素. 由于不知道角点在实时采集图像中的理论位置，所以在物理实验中通过与 Micron Tracker 系统及 Harris 算法进行对比来分析 SC 算子的性能. Micron Tracker 系统由 Claron Tech 公司生产，其被广泛应用于众多需要进行精确定位的场合. 文献[7]指出其测量精度达到 0.2 mm. 实验中用圆圈标记 X 角点，分别用短线和长线标识 l_{wb} 与 l_{bw}，如图 7 所示.

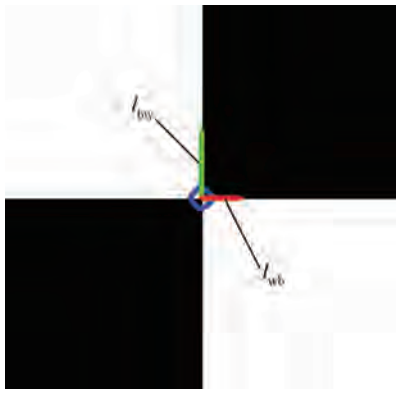


图 7 X 角点及其方向标识

Fig. 7 X-corner point and its direction detected by SC detector

3.1 仿真实验

为验证噪声和畸变对本文算法性能的影响，选择畸变后的仿真图像进行角点定位实验. 首先利用 Matlab 软件对含有 25 个 X 角点的棋盘格图案进行仿射变换，得到图片仿射后各个角点的理论位置 $\hat{x}_i(i=1,2,\cdots,25)$ 及相应 l_{wb}，l_{bw} 与水平方向的夹角 $\hat{\theta}_{1i}$，$\hat{\theta}_{2i}$. 对仿射后的图像加入标准差为 δ 的高斯噪声生成噪声图像，然后采用 SC 算子和集成到 OpenCV 中的 Harris 算子分别进行角点定位，记定位结果为 $x_i(i=1,2,\cdots,25)$，记 SC 算子求得的 l_{wb} 与 l_{bw} 的角度分别为 θ_{1i}，θ_{2i}. 然后将畸变噪声图像的检测结果与理论值进行比较，分析噪声和畸变对算法定位和定向精度的影响，由于 Harris 算法不能提供角点的方向信息，所以只列出 SC 算子的定向结果及分析. 定义

$$\sigma_n=\sqrt{\frac{1}{25}\sum_{i=1}^{25}\|\hat{x}_i-x_i\|^2} \tag{10}$$

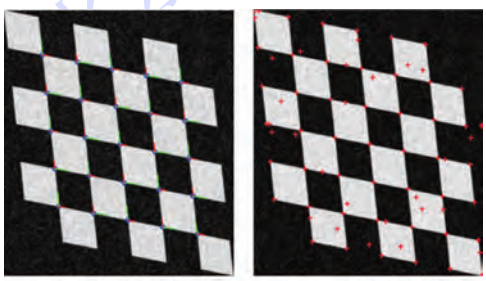
$$E_n=\frac{\sum_{i=1}^{25}|\hat{\theta}_{1i}-\theta_{1i}|+\sum_{i=1}^{25}|\hat{\theta}_{2i}-\theta_{2i}|}{50} \tag{11}$$

式中，σ_n 和 E_n 分别为定位和定向误差. 仿真实验结果如表 1 所示，图 8 为噪声标准差 $\delta=0.24$ 时的畸变图像定位结果.

表 1 仿真实验结果

Table 1 Results of simulation experiments

参数	算子	噪声标准差 δ						
		0.04	0.08	0.12	0.16	0.20	0.24	0.28
σ_n/像素	SC 算子	0.329 3	0.329 3	0.355 3	0.354 4	0.432 0	0.420 0	0.421 8
	Harris	0.305 4	0.321 0	0.338 7	0.362 2	0.378 0	0.413 2	
E_n/(°)	SC 算子	1.538 4	1.468 8	1.481 2	1.424 4	1.481 2	1.764 0	1.876 4



(a) SC算子检测结果 (b) Harris算子检测结果

图 8 噪声标准差 $\delta=0.24$ 时的畸变图像定位结果

Fig. 8 Result of distortion image when noise standard deviation is 0.24

SC 算子和 Harris 算子的定位精度都会随图像噪声的加强而有所下降，当噪声标准差小于 0.28时，SC 算子和 Harris 算子的定位误差都在 0.5个像素以内. 两者在定位精度方面不相上下，但是 Harris 易受噪声的影响，误检点较多，在实际应用中会产生不必要的干扰，如图 8(b)所示. 在噪声标准差为 0.28 时，因 Harris 算子误检点过多，以致无法计算精确度. 而 SC 算子则表现出良好的抗噪性和抗畸变性. 同时，由表 1 可以看出，当噪声标准差小于 0.28 时，SC 算子的定向精度也很稳定，平均误差在 2°以内，稳定精确的定向功能也是 SC 算子的一个优势. 总体来说，本文提出的算法能在保证基本不受噪声和畸变影响的同时进行精确定位及定向，满足应用中对鲁棒性能及精度的要求.

3.2 物理实验

物理实验分为 3 部分：第 1 部分是与 Harris 算法及 Micron Tracker 系统的定性比较实验；第 2 部分是检验光照亮度对算法性能影响的实验；第 3 部分是与 Micron Tracker 系统的定量比较实验.

第 1 部分的对比实验分为 4 组，统计结果在表 2 中给出. 第 1 组实验里放置的棋盘图案有 96 个X 角点，黑白块的边长为 20 mm，放置在距相机 1 m 远的地方. 第 2 组实验里增加了一黑白块边长为 30 mm，角点个数为 40 的棋盘格. 第 3 组实验时，改变第 2 个棋盘格平面与相机平面之间的夹角. 第 4 组实验中将第 2 个棋盘格进行旋转并改变 2 个棋盘格与相机平面之间的夹角. 图 9 给出了第 4 组棋盘格实验的结果图. 由实验结果看出，Harris算法的检测结果中存在很多误检点，而Micron Tracker系统不太稳定，容易出现漏检的

情况. 与它们相比,本文提出的 SC 算子在保证基本排除非 X 角点的前提下全部检出了 X 角点.

图 10 显示了在不同光照亮度下 SC 算子的检测情况. 光照亮度从高亮度到正常再到低亮度,都没有影响 SC 算子对 X 角点及其方向的检测. 图中分别用短线和长线标出了 WB 线和 BW 线的方向. 实验表明本文提出的算法对亮度改变具有一定的鲁棒性.

表 2　SC 算子、Harris 算子与 Micron Tracker 系统检测结果对比

Table 2　Comparison between experimental results of SC detector, Harris detector and Micron Tracker system

	实际角点数	SC 算子		Harris		Micron Tracker	
		X 角点	非 X 角点	X 角点	非 X 角点	X 角点	非 X 角点
第 1 组	96	96	0	96	310	44	0
第 2 组	136	136	0	136	371	71	0
第 3 组	136	136	0	136	363	85	0
第 4 组	136	136	0	136	357	69	0

图 9　棋盘格实验

Fig. 9　Experiments of X-corner points

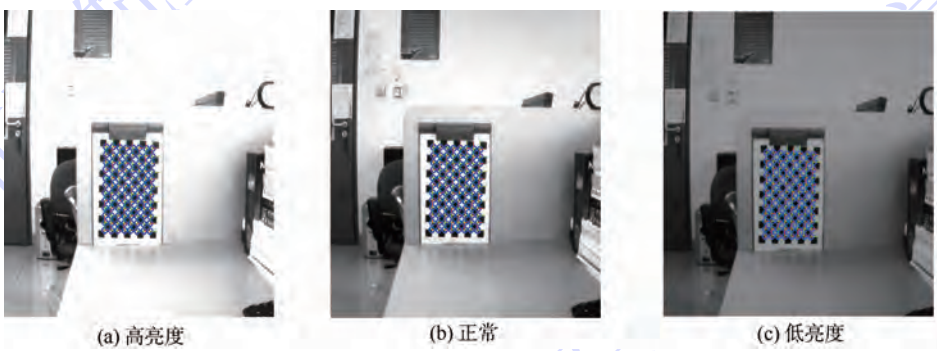


图 10　不同光照亮度下 SC 算子的检测结果

Fig. 10　Results of SC detector under different light conditions

上述 2 部分物理实验对角点检测算法的抗噪性、抗畸变性以及对光照变化的鲁棒性做了一个定性的分析. 下面通过与 Micron Tracker 系统的对比实验定量地分析角点检测算法的误检率、漏检率和检测精确度. 实验所用图像中包括 96 个 X 角点的棋盘格图案,黑白块边长为 20 mm;距离相机 1 m 远.

为比较角点定位精度,在利用生产该摄像机的 Claron Tech 公司提供的源代码对其提取的左视图进行角点检测后,将导出的左视图置于 SC 算子中进行检测. 为比较 2 个算法在不同条件下的检测结果,还改变了棋盘格平面和摄像机平面之间的角度. 图 11 所示为 30° 时,与 Micron Tracker系统的对比实验的情形. 表 3 中的数据是 20 组实验结果的平均值. 其中正检率 = 正确检测的角点数/实际总角点数,漏检率 = 漏检的角点数/

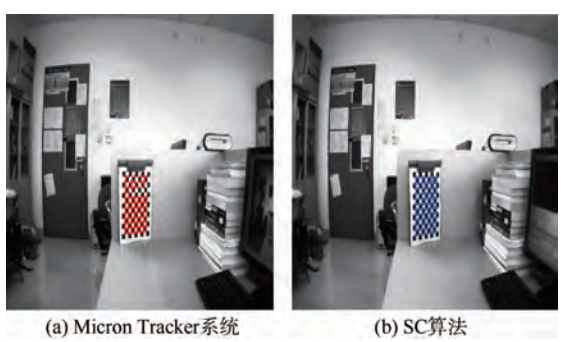


图 11　与 Micron Tracker 系统的对比实验

Fig. 11　Comparative experiments with Micron Tracker

实际总角点数,误检率 = 误检的角点数/实际总角点数. 数据显示当棋盘格平面和摄像机平面之间的夹角较小时,Micron Tracker 系统的漏检率较高,SC 算子则几乎不存在漏检,随着角度的增加,Micron Tracker 系统漏检率逐渐降低,正检率逐渐升高. SC 算子则相反,当被测平面过于倾斜时,由于检测窗口的大小 r 不适合该情况,正检率会有所下降,出现漏检. 2 个算法的误检率则一直都很低,总的来说都具有较好的鲁棒性. 对于更普遍的接近平行的情况,SC 算法则在稳定性方面更有优势.

表 3 不同角度下 Micron Tracker 系统与 SC 算子的检测结果

Table 3 Results of Micron Tracker system and SC detector under different conditions

夹角/(°)	Micron Tracker			SC 算子		
	正检率/%	漏检率/%	误检率/%	正检率/%	漏检率/%	误检率/%
0	31.25	68.75	0	100	0	0
30	98.05	1.95	0	100	0	0
60	99.22	0.78	0.52	69.53	30.47	0.65

通过对 20 组实验的检测结果进行统计分析,获得 2 个算法的检测结果分别在 x,y 方向上测量差的标准差(σ_s)、均方根(RMS,σ_r)及 2 个结果间距离的标准差和 RMS. Micron Tracker 系统与 SC 算子检测结果间的关系如表 4 所示. 可以看出 2 个算法在精度上是不相上下的.

$$\sigma_s = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_i - \bar{z})^2} \tag{12}$$

$$\sigma_r = \sqrt{\frac{1}{N}\sum_{N} z^2} \tag{13}$$

当计算 Δx,Δy 或距离 Dis 的标准差和 RMS 时,分别用它们代替式(12)和式(13)中的 z 进行计算.

表 4 Micron Tracker 系统与 SC 算子检测结果间的关系

Table 4 Relations between the two detectors' test results

像素

$\Delta x = x' - x$		$\Delta y = y' - y$		$\text{Dis} = \sqrt{(x'-x)^2 + (y'-y)^2}$	
标准差 (σ_s)	RMS (σ_r)	标准差 (σ_s)	RMS (σ_r)	标准差 (σ_s)	RMS (σ_r)
0.1743	0.1917	0.1647	0.1802	0.1109	0.2631

注:x'—Micron Tracker 系统的检测结果;x—SC 算子的检测结果.

与 Micron Tracker 系统的对比实验表明 SC 算子与其有相同的检测精度,满足精确定位的要求.

在 Intel Core i3 CPU 3.3 GHz RAM 2.0 GB 的处理器上,检测一帧大小为 1 024 × 768 包含96 个角点的棋盘格图像,本文算法一般需要0.187 s,Harris 算法与 Micron Tracker 系统平均花费时间则分别为 0.053 s,0.069 s. 传统的 Harris 算法并不满足实时性要求[9,21],因本文使用的是 OpenCV 中已优化的 Harris,所以速度快. 而Micron Tracker 系统封闭,不做讨论. 本文 SC 算子因需要利用式(3)对图像中的每个像素点计算对应的 SC 值,所以耗时相对较大,但在实际工程项目中,若采用适当的优化加速算法,还能大大降低本文算法的处理时间.

4 讨论及分析

实验发现,本文算法会因窗口大小 r 选取不当导致漏检,通过实验总结和分析,利用自适应法根据图片的大小及实验条件自动调节窗口的大小,以求降低漏检率. 如果出现误检,则主要是因为窗口内检测到如下 2 种干扰情况,它们与 X 角点一样存在 4 个台阶,但是窗口内并无交点. 通过对 X 角点和干扰情况进行分析,可以用下列条件对干扰情况进行排除:

1) 若当前窗口中心的邻域内像素灰度值为黑白混合,则为 X 角点,若为纯黑或纯白,则为干扰情况.

2) 根据前面对 A,B,C,D 的定义,标出各干扰情况中的 A,B,C,D,如图 12 所示. 取 AC 或 BD 上的像素点,若其邻域内像素灰度值为黑白混合,则为 X 角点,否则为干扰.

对这些干扰情况进行排除,能大大降低误检率.

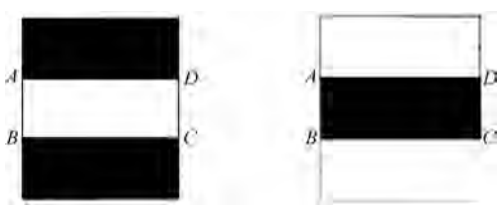


图 12 干扰图形

Fig. 12 Interferences

5 结 论

分析了 X 角点的图像特征,提出了基于对称性计算的 X 角点检测算法. 该方法具有以下特点:

1）仅提取 X 角点,排除了其他类型的角点,为特征识别提供亚像素级的定位信息和准确的定向信息.

2）能有效排除非 X 角点的干扰.

本文算法虽然在亚像素定位精度方面与其他 2 种经典算法相当,但在图像发生畸变并有较高噪声的情况下,本文算法仍能在基本排除非 X 角点的同时准确提取 X 角点的位置信息,并且还能提供较高精度的方向信息,在位姿测量及模板匹配中具有较高的应用价值.

参考文献（References）

[1] Sánchez-Margallo J A, Sánchez-Margallo F M, Pagador J B, et al. Technical evaluation of a third generation optical pose tracker for motion analysis and image-guided surgery [M]//Clinical Image-Based Procedures. From Planning to Intervention. Heidelberg: Springer, 2013: 75-82.

[2] 王志衡,吴福朝,王旭光. 基于局部方向分布的角点检测及亚像素定位[J]. 软件学报,2008,19(11):2932-2942.
Wang Z H, Wu F C, Wang X G. Corner detection and sub-pixel localization based on local orientation distribution[J]. Journal of Software, 2008, 19(11): 2932-2942(in Chinese).

[3] Zhang Z Y. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330-1334.

[4] 刘阳成,朱枫. 一种新的棋盘格图像角点检测算法[J]. 中国图象图形学报,2006,11(5):656-660.
Liu Y C, Zhu F. A new algorithm for X-corner detection[J]. Journal of Image and Graphics, 2006, 11(5): 656-660(in Chinese)

[5] Bouguet J Y. Camera calibration toolbox for matlab[EB/OL]. California, Tex: California Institute of Technology, 2009 (2013-12-02). http://www.vision.caltech.edu/bouguetj/calib_doc/.

[6] Bradski G, Kaehler A. Learning OpenCV: computer vision with the OpenCV library[M]. Sebastopol: O'Reilly Media, Inc., 2008: 422-430.

[7] Claron Technology Inc. Micron tracker developer's manual MTC 3.6[M]. Toronto: Claron Technology Inc., 2011: 25-35.

[8] Harris C, Stephens M. A combined corner and edge detector[C]//Proceedings of Alvey Vision Conference. Manchester: Organising Committee AVC 88, 1988, 15: 147-152.

[9] 王崴,唐一平,任娟莉,等. 一种改进的 Harris 角点提取算法[J]. 光学精密工程,2008,16(10):1995-2001.
Wang W, Tang Y P, Ren J L, et al. An improved algorithm for Harris corner detection[J]. Optics and Precision Engineering, 2008, 16(10): 1995-2001(in Chinese).

[10] 侯建辉,林意. 自适应的 Harris 棋盘格角点检测算法[J]. 计算机工程与设计,2009,30(20):4741-4743.
Hou J H, Lin Y. Adaptive Harris X-corner detection algorithm[J]. Computer Engineering and Design, 2009, 30(20): 4741-4743(in Chinese).

[11] 郝颖明,朱枫. 摄像机在线标定中的棋盘格角点自动检测方法[J]. 计算机工程,2007,33(17):213-215.
Hao Y M, Zhu F. Automatic X-corners detection for online camera calibration[J]. Computer Engineering, 2007, 33(17): 213-215(in Chinese).

[12] 储珺,郭卢安政,赵贵花. 采用环形模板的棋盘格角点检测[J]. 光学精密工程,2013,21(1):189-196.
Chu J, Guo L A Z, Zhao G H. Chessboard corner detection based on circular template[J]. Optics and Precision Engineering, 2013, 21(1): 189-196(in Chinese).

[13] 胡海峰,侯晓微. 一种自动检测棋盘角点的新算法[J]. 计算机工程,2004,30(14):19-21.
Hu H F, Hou X W. A new algorithm for automatically detecting chessboard corners[J]. Computer Engineering, 2004, 30(14): 19-21(in Chinese).

[14] Chen D, Zhang G J. A new sub-pixel detector for X-corners in camera calibration targets[J]. WSCG (Short Papers), 2005, 5: 97-100.

[15] 张广军. 视觉测量[M]. 北京:科学出版社,2008:57-61.
Zhang G J. Vision measurement[M]. Beijing: Science Press, 2008: 57-61(in Chinese).

[16] 魏军,刘达,王田苗,等. 基于段测试的 X 角点检测与亚像素定位[J]. 机器人,2011,33(1):97-101.
Wei J, Liu D, Wang T M, et al. Segment test based X comer detection and sub-pixel localization[J]. Robot, 2011, 33(1): 97-101(in Chinese).

[17] Rosten E, Porter R, Drummond T. Faster and better: a machine learning approach to corner detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(1): 105-119.

[18] Lowe D G. Distinctive image features from scale-invariant key points[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[19] Bay H, Tuytelaars T, Van Gool L. Surf: speeded up robust features[C]//Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2006: 404-417.

[20] 尚雅层,陈静,田军委. 高斯拟合亚像素边缘检测算法[J]. 计算机应用,2011,31(1):179-181.
Shan Y C, Chen J, Tian J W. Sub-pixel edge detection algorithm based on Gauss[J]. Journal of Computer Applications, 2011, 31(1): 179-181(in Chinese).

[21] 郭永芳,于明,孙以材. 一种改进的快速角点检测方法研究[J]. 计算机工程与应用,2011,47(12):159-161.
Guo Y F, Yu M, Sun Y C. Study on advanced rapid corner detection method[J]. Computer Engineering and Application, 2011, 47(12): 159-161(in Chinese).

北航学报 赠阅

Sub-pixel X-corner detection and location with directional characteristics

MENG Cai*, XUE Jiao

(School of Astronautics, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: X-corner points are widely used in camera calibration and visible-spectrum visual tracking system as they can provide not only location information, but also direction information. To accurately detect the location and direction of X-corner points, a new algorithm based on the symmetry calculation of X-corner points was proposed. The algorithm mainly consists of two parts. First, the pixel position of the X-corner is determined by using the value of symmetry calculation. Then the sub-pixel location for the X-corner was obtained based on the quadratic curve fitting and inter-section of lines. Through a series of experiments and a comparison with Harris algorithm and Micron Tracker system, the impacts of Gauss noise and distortion on the positioning accuracy and robustness were analyzed. Experiments show that the method can effectively detect X-corners and exclude other types of corner points and disturbance, and thus has higher localization accuracy and robustness.

Key words: X-corner point; symmetry; sub-pixel location; direction detection; quadratic curve fitting

北京航空航天大学学报

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0246

超高频脉冲电弧焊接熔池金属流动行为

杨明轩*，杨舟，从保强，王强，齐铂金

（北京航空航天大学 机械工程及自动化学院，北京 100191）

摘 要：熔池金属流动一方面受电弧行为影响，另一方面与凝固组织形核、结晶过程联系紧密，是电弧焊接基础研究的关键，针对电弧力作用研究了超高频脉冲电弧焊接（UHFP-AW，Ultra High Frequency Pulsed Arc Welding）熔池金属流动特点，采用理论研究与试验分析相结合的方式对熔池流动方式、强度、温度分布等要素进行了分析，并完成了液态金属表面运动特征监测. 结果表明，超高频脉冲电弧对熔池液态金属的热、力作用显著，表面流体流速大于30 cm/s，UHFP-AW 电弧力造成熔池表面沉降变形，引起双环流，熔池温度扩散范围相应减小，有效降低了电弧热量对母材金属的影响，尤其在钛合金焊接中对控制焊缝晶粒尺寸具有重要意义.

关 键 词：超高频脉冲电弧；熔池；流动行为；高速摄像；表面变形
中图分类号：TG403
文献标识码：A **文章编号**：1001-5965(2015)04-0589-05

超高频脉冲电弧收缩显著[1]，与常规电弧相比，其能量密度、挺度及电弧力均大幅增长，研究表明超高频脉冲电弧焊（UHFP-AW，Ultra High Frequency Pulsed Arc Welding）可改善焊缝组织形貌[2]，提高接头力学性能[3-4]. 熔池流动受电弧行为影响，且与凝固组织结晶联系紧密，是电弧焊基础研究的关键.

焊接熔池是复杂的流体系统，浮力、电磁力、表面张力和等离子流力是熔池流动的主要驱动力，它们在熔池内部产生不同的流体运动形式，直接决定了熔池的尺寸形貌. 研究表明，熔池形貌及焊缝成型是各驱动力共同作用的结果，其中电磁力和表面张力是熔池流动的主要驱动力[5-7]. 计算表明，电磁力将引起熔池内部自激搅拌作用，同时引发向中心聚拢、向下流动的流体状态，易形成深而窄的焊缝；表面张力是熔池流动性的决定性因素，受熔池液态金属表面性质影响较大[8]；此外，还有研究验证了电弧力对熔池几何形貌的作用[9-10]. 从 20 世纪 80 年代开始，学者们相继提出并接受熔池中 2 个环流并存的双环流模式，分别由电磁力和表面张力驱动，各自驱动的环流流向相反，并共同影响熔池流动状态[11-12]. 若其中熔池内部电磁搅拌增强[13]或表面张力增大，则熔池金属流动性提高.

在前期电弧行为研究的基础上[14]，本文拟围绕超高频脉冲焊接电弧力增大的试验事实，分析其对熔化金属流体流动方式、速度、热流分布等流动行为的影响，完成超高频脉冲电弧焊熔池流动行为机理研究，其结果将为控制焊缝质量、优化工艺参数提供理论支撑.

1 试验设备及方法

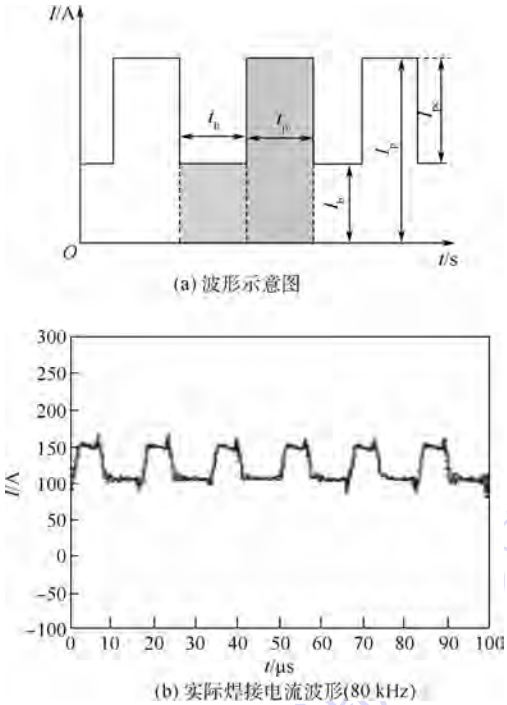
超高频脉冲电弧焊接电流波形如图 1 所示. 电弧力可通过精密测力传感器（0.01 g）获得. 熔化金

收稿日期：2014-05-05；**录用日期**：2014-06-24；**网络出版时间**：2014-07-16 10:23
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0246.html
基金项目：中央高校基本科研业务费专项资金资助项目（YWF-14-JXXY-008）；国家自然科学基金资助项目（51405007）
* **通讯作者**：杨明轩（1985—），男，四川南充人，讲师，yangmingxuan@buaa.edu.cn，主要研究方向为高频脉冲焊接电弧/熔池行为.

引用格式：*杨明轩，杨舟，从保强，等. 超高频脉冲电弧焊接熔池金属流动行为[J]. 北京航空航天大学学报，2015，41(4)：589-593. Yang M X，Yang Z，Cong B Q，et al. Metal molten behavior with ultra high frequency pulsed arc welding[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：589-593 (in Chinese).*

北航学报 赠阅

属表面流动状态监测将基于 Mega Speed 高速摄像机(MS75k,满幅最大采样频率 6 kHz) + 滤光镜 + 广角可调减光镜完成,熔池流动监测如图 2 所示.



(a) 波形示意图

(b) 实际焊接电流波形(80 kHz)

I_b—脉冲基值电流;I_{pc}—脉冲电流幅值;I_p—脉冲电流;t_b—基值电流作用时间;t_p—脉冲电流作用时间.

图 1 超高频脉冲电弧焊接电流波形

Fig. 1 Current waveform of ultra high frequency pulsed arc welding

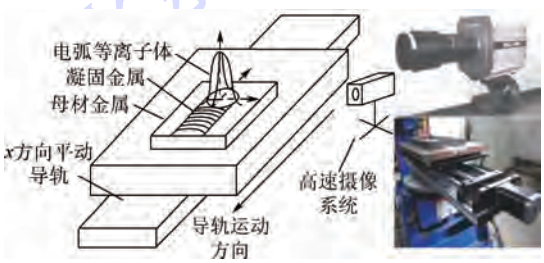


图 2 熔池流动监测

Fig. 2 Surveillance for motel fluid

电极为 ϕ2.4 mm 铈钨 W-2% Ce,平均电流 80 A,电极顶端与工件表面距离 7 mm,保护气为 99.99% 普通氩气,流量 12 L/min. 在监测过程中,工件运动状态为:①随工作台在 x 方向静止;②随工作台在 x 方向恒速(v = 50 mm/min)平动. 瞬态流动行为监测使用 5 mm 厚 0Cr18Ni9Ti 奥氏体不锈钢为母材金属,理论分析涉及同样规格的不锈钢及 2.5 mm 厚 Ti-6Al-4V 钛合金平板.

2 试验结果

2.1 熔化金属流动行为——定点

图 3 为熔化金属表面流动行为——定点,即 50 幅/s采样频率下熔化金属流动瞬态行为,采样时间间隔为 0.02 s. 通过连续拍摄的动态行为可观测到液体表面流体从中心向边缘呈辐射状流动. 监测发现,这一结果在常规电弧焊及 UHFP-AW 中均存在,这与表面张力的作用规律相符,即在常规条件下(表面张力温度梯度 $d\gamma/dT < 0$),金属流体从表面张力较小的熔池中心(温度较高)流向表面张力较大的边缘(温度较低).

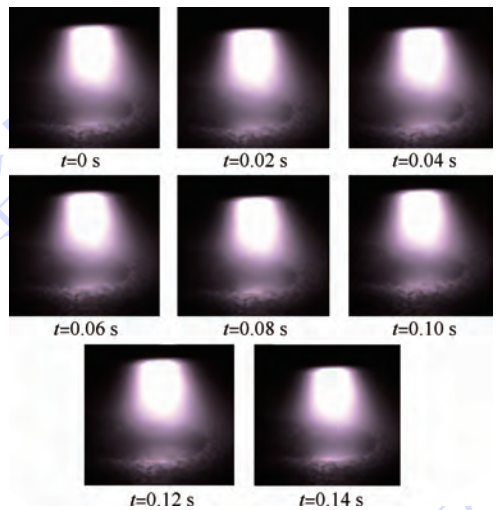


图 3 熔化金属表面流动行为——定点

Fig. 3 Flow behavior on surface of molten metal—fixed point

进一步通过熔池表面氧化膜杂质运动可估算表面流体流动速度,熔化金属表面流动速度估算如图 4 所示. 由于推力矩作用,长条形物体在金属流体推动下的受迫运动多呈旋转状态,不利于通过位移推算流速,故选择自身体积较小的杂质作为参考目标,不同时刻杂质所在位置可通过图 4 中的标注点标明. 由不同时刻点的位移量可估算

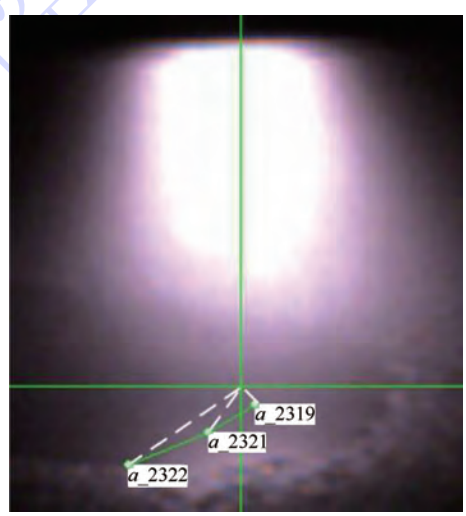


图 4 熔化金属表面流动速度估算

Fig. 4 Calculation for velocity of fluid on surface of molten metal

各个采样时间间隔内物体受迫运动的速度. 结果表明,表面张力推动液态金属表面杂质运动的平均速度约为15.76 cm/s,靠近边缘时运动速度明显大于熔池中心部位处的运动速度,这与表面张力的作用机理相吻合. 但这一流速仅可用于表征密度与母材金属类似的杂质受表面张力环流驱动时所表现出的宏观速度,若考虑杂质自身受迫流动机制及其流动阻力,则表面液态金属流体流动的峰值速度可能成倍增长,并将出现在熔池边缘位置.

2.2　熔化金属流动行为——恒速平动

在常规电弧焊条件下监测工件以恒速(50 mm/min)单一自由度平动时熔池表面金属流动状态,熔化金属表面流动行为——恒速平动结果如图 5 所示. 从初始时刻(t = 0 s)开始,熔池金属自液面中心位置向尾部流动(与工件平动方向一致),至固液界面处流向两侧,一部分流体及氧化物杂质被推送至熔池侧边缘(t = 0.04 s),其他液态金属快速流向熔池前端(t = 0.06 s),受前端固液界面阻挡,向熔池两侧沉积凝固(t = 0.08 s). 在此过程中,液面仍存在由中心向边缘发散的流体流向,与 2.1 节定点熔池流动行为类似,这一流动方式将液面附近的氧化物杂质输送至熔池边缘.

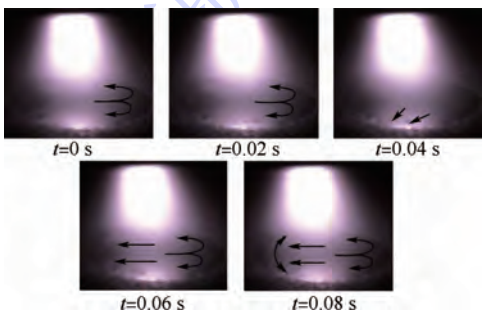


图 5　熔化金属表面流动行为——恒速平动

Fig. 5　Flow behavior on surface of molten metal—constant movement

3　结果讨论及机理研究

3.1　UHFP-AW 熔池流动行为

已有的研究表明,常规电弧焊熔池流体集中于表面,即表面张力驱动下的 Marangoni 对流,熔池流体流动状态模拟如图 6 所示. 其中 f 为脉冲频率. 表面张力系数为负时,流体在熔池表面由中心流向边缘,易形成浅而宽的焊缝,此时由于电弧力较小,熔池表面沉降可忽略不计,电弧行为对液态金属流动状态几乎无影响;受 UHFP-AW 电弧力影响,熔池金属呈现显著的双环流流动模式,如图 7 所示,在较大电弧力作用下,电弧在熔池表面的作用点下移,熔池内部轴向驱动力作用加强,由中心向下的环流增强.

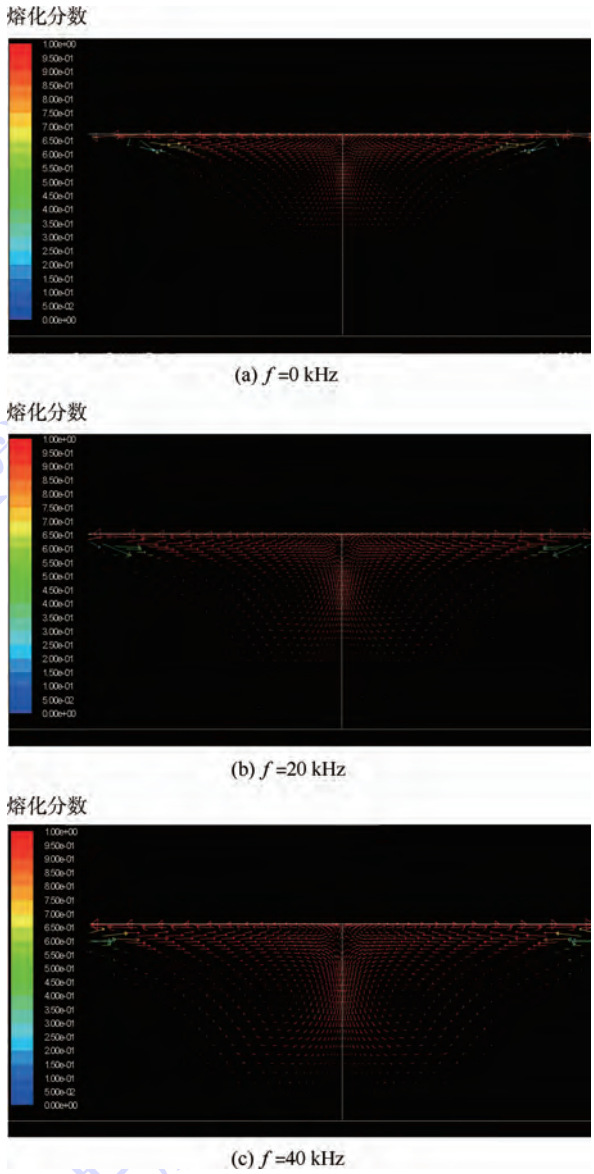


图 6　熔池流体流动状态模拟

Fig. 6　Simulation for fluid status in molten pool

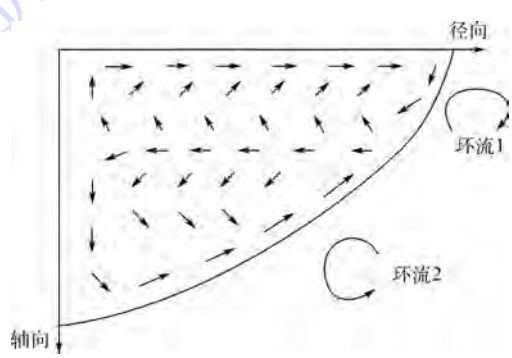


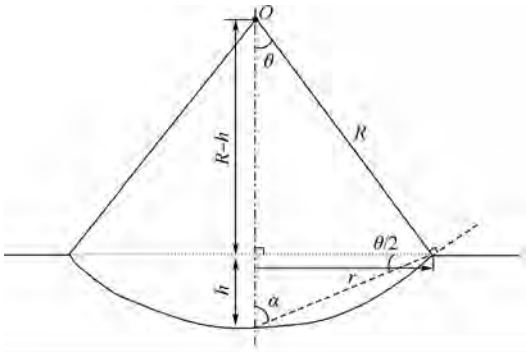
图 7　UHFP-AW 双环流流动模式

Fig. 7　UHFP-AW double circulations mode

3.2　电弧力对熔池流动的作用

与常规电弧焊接相比,UHFP-AW 电弧力增

长显著,该结果在奥氏体不锈钢及 Ti-6Al-4V 钛合金试验中均得到验证. 针对不锈钢的研究成果表明,较大的电弧力将在熔池液面造成沉降变形,可能对熔池金属流动产生重要作用,类似地,本节将以 2.5 mm 厚 Ti-6Al-4V 钛合金为对象分析电弧力、表面张力、变形流体重力、液面压力差等因素的平衡条件,从而研究 UHFP-AW 熔池流动行为与电弧行为之间的关系. 分析过程将采用圆弧变形假设,受迫变形面在空间存在虚拟原点 O,由此产生虚拟半径 R,熔池表面凹陷变形几何关系如图 8 及式(1)所示.



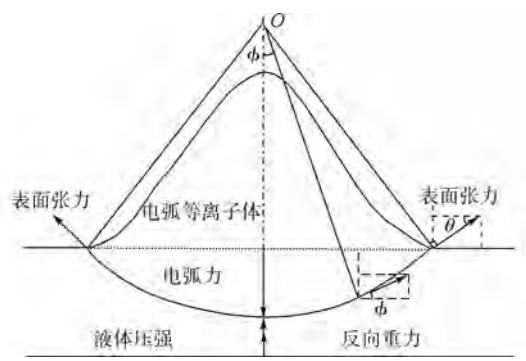
h—表面凹陷深度;r—凹陷最大处半径(假设为焊缝熔宽的1/2);θ—球冠中心半角;α—球冠顶半角.

图 8 熔池表面凹陷变形几何关系

Fig. 8 Geometry of surface depression on surface of molten pool

$$R=\frac{h^2+r^2}{2h} \tag{1}$$

在焊接过程中,熔池表面若出现沉降变形,则电弧与液态金属的交界面将处于平衡状态,如图 9所示,此处涉及的力包括电弧力、排开液体的重力、表面张力、凹陷变形所造成的液面压力差等.



ϕ—球冠面任意点切线方向与水平方向的夹角.

图 9 凹陷液面交界处力平衡

Fig. 9 Force balance of surface depression

当液态金属流体表面在电弧力作用下出现凹陷变形时,球冠部分(如图 8 所示)排开液态金属体积可由式(2)计算,进而可据此计算排开液体的重力,在焊接过程中,该重力将驱使被排开的液态金属复原,从而使得熔池表面恢复为平面. 因此,这一重力在液面交界处将沿法线方向竖直向上,与电弧力方向相反.

$$V=\pi h^2\left(R-\frac{h}{3}\right) \tag{2}$$

表面张力沿交界面切向对液态金属产生向边缘的拉力,并以电弧中心轴线为对称轴,促使液面恢复平面状态;根据经典物理理论,液面压力差可通过排开部分液体的压强及作用面积计算,其作用趋势与重力的类似. 由此,将式(1)、式(2)代入力平衡方程($f(h)=0$)可得式(3),通过求解即可计算不同电弧力条件下,Ti-6Al-4V 钛合金熔池表面凹陷深度 h[15].

$$f(h)=F-G-PS-\sigma_{\mathrm{sum}}=$$

$$F-\rho gV-\rho ghS-2\int_{\phi}\sigma R\phi\sin\phi=$$

$$F-\rho g\pi h^2\left(R-\frac{h}{3}\right)-\rho gh\pi r^2-2\sigma R\cdot 2\theta\sin\theta=$$

$$-\frac{\pi}{6}\rho gh^3-\frac{3\pi}{2}\rho gr^2h-8\sigma r\arctan\frac{h}{r}+F \tag{3}$$

式中,F 为试验测得电弧力;G 为凹陷部分排开液态金属的重力;P 为液面压强差;S 为作用面积;σ_{sum}为气液交界面上表面张力;σ 为表面张力系数;ρ 为液体密度.

分析式(3)可知,方程 $f(h)$ 的零解即为液态金属液面沉降的深度,表面凹陷变形近似计算结果如表 1 所示. 在钛合金母材条件下,常规电弧焊熔池表面凹陷深度为 0.12 mm,这与焊缝熔透率较低的试验结果相一致;20,40 kHz 条件下液面沉降深度至少增加了 2.5 倍,这将使得熔池流体呈现双环流流动状态,有利于电弧热量向熔池底部输运.

表 1 表面凹陷变形近似计算结果

Table 1 Approaching results of surface depression

脉冲频率 f/kHz	电弧力 F/mN	熔宽 W/mm	熔深 D/mm	表面凹陷深度 h/mm
0	1.79	4.16	2.29	0.12
20	4.19	3.10	1.75	0.31
40	7.00	2.77	1.85	0.53

4 结 论

1) 熔池表面流体在表面张力驱动下由熔池中心流向边缘,最大流动速度出现在熔池边缘,监测结果至少大于 30 cm/s,UHFP-AW 可能具有更大的流动速度.

2) 以钛合金为母材金属,分析并计算了电弧

力造成的熔池表面凹陷,20 kHz,40 kHz 条件下 UHFP-AW 所得液面沉降深度为常规电弧焊的 2.5倍以上.

参考文献(References)

[1] Ghosh P K, Dorn L, Hübner M, et al. Arc characteristics and behavior of metal transfer in pulsed current GMA welding of aluminum alloy[J]. Journal of Materials Processing Technology, 2007, 194(1): 163-175.

[2] Yang M X, Qi B J, Cong B Q, et al. Effect of pulse frequency on microstructure and properties of Ti-6Al-4V by ultrahigh-frequency pulse gas tungsten arc welding[J]. The International Journal of Advanced Manufacturing Technology, 2013, 68(1-4): 19-31.

[3] 从保强,齐铂金,周兴国. 超快变换高频变极性方波 TIG 电弧行为[J]. 焊接学报,2009,30(6):87-90.
Cong B Q, Qi B J, Zhou X G. TIG arc behavior of ultra fast-convert high-frequency variable polarity square wave[J]. Transactions of the China Welding Institution, 2009, 30(6): 87-90(in Chinese).

[4] Onuki J, Anazawa Y, Nihei M, et al. Development of a new high-frequency, high-peak current power source for high constricted arc formation[J]. Japanese Journal of Applied Physics, 2002, 41(9R): 5821-5826.

[5] Kim W H, Na S J. Heat and fluid flow in pulsed current GTA weld pool[J]. International Journal of Heat and Mass Transfer, 1998, 41(21): 3213-3227.

[6] Kumar A, DebRoy T. Calculation of three-dimensional electromagnetic force field during arc welding[J]. Journal of Applied Physics, 2003, 94(2): 1267-1277.

[7] Tsai M C, Kou S. Electromagnetic force induced convection in weld pool with a free surface: a model is developed to predict heat transfer, fluid flow and pool shape in GTAW[J]. Welding Journal, 1990, 69(6): 241-246.

[8] 李冬杰,陆善平,李殿中,等. TIG 焊接过程中熔池对流形式的示踪分析[J]. 焊接学报,2011,32(8):45-48.
Li D J, Lu S P, Li D Z, et al. Tracer investigation of convection in weld pool under TIG welding process[J]. Transactions of the China Welding Institution, 2011, 32(8): 45-48(in Chinese).

[9] Ko S H, Yoo C D, Choi S K. Effects of surface depression on pool convection and geometry in stationary GTAW[J]. Welding Journal, 2001, 80(2): 39-45.

[10] 赵明,武传松,胡庆贤. TIG 焊接熔透熔池形状和表面变形的数值模拟[J]. 机械工程学报,2006,42(10):203-213.
Zhao M, Wu C S, Hu Q X. Numerical simulation of penetrated weld pool geometry and surface deformation in TIG welding[J]. Journal of Mechanical Engineering, 2006, 42(10): 203-213(in Chinese).

[11] Kou S, Sun D K. Fluid flow and weld penetration in stationary arc welds[J]. Metallurgical Transactions A, 1985, 16(2): 203-213.

[12] Oreper G M, Szekely J. A comprehensive representation of transient[J]. Metallurgical Transactions A, 1987, 18(7): 1325-1332.

[13] Atthey D R. A mathematical model for fluid flow in a weld pool at high currents[J]. Journal of Fluid Mechanics, 1980, 98(4): 787-801.

[14] Yang M X, Qi B J, Cong B Q, et al. Study on electromagnetic force in arc plasma with UHFP-GTAW of Ti-6Al-4V[J]. IEEE Transactions on Plasma Science, 2013, 41(9): 2561-2568.

[15] Yang M, Yang Z, Cong B, et al. A study on the surface depression of the molten pool with pulsed welding[J]. Welding Journal, 2014, 93(8): 312-319.

Metal molten behavior with ultra high frequency pulsed arc welding

YANG Mingxuan*, YANG Zhou, CONG Baoqiang, WANG Qiang, QI Bojin

(School of Mechanical Engineering and Automation, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: Molten liquid metal is the key factor of the fundamental research during welding process, as it is the transfer element between arc behavior and solidification. Researches of the characteristics of molten pool were carried out by ultra high frequency pulsed arc welding (UHFP-AW), especially with the focus on the effect of arc force. The methods combining experiments and theoretical study were produced for the analyses of important factors, such as fluid status, velocity and temperature distribution of molten pool. And the fluid in molten surface was monitored as the supplement for further study. The results indicate the significant effect of the arc plasma on heat and force of molten pool with UHFP-AW. The monitor data demonstrate that the velocity of surface fluid is more than 30 cm/s, and the double-circulation exists with surface depression caused by larger velocity during UHFP-AW. As a result, the temperature diffusion of molten pool would reduce, which can limit the impact of arc heat on base metal effectively. It is the key for controlling the grain size of titanium alloys with arc welding process.

Key words: ultra high frequency pulsed arc; molten pool; molten behavior; high speed camera; surface deformation

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0288

高超声速气动热数值计算壁面网格准则

张智超，高振勋*，蒋崇文，李椿萱

（北京航空航天大学 航空科学与工程学院，北京 100191）

摘 要：首先对高超声速气动热数值模拟中的3种壁面法向网格准则（平均自由程（MFP）、自由来流参数网格雷诺数和壁面参数网格雷诺数准则）进行了分析. 随后，提出了壁面参数预估方法，使得壁面参数网格雷诺数和MFP两种准则无需进行试算即可直接预估壁面法向网格尺度. 其次，运用提出的壁面参数预估方法结合不同网格准则，在同一实验条件下确定多个壁面法向网格尺度，通过与实验数据对比研究各网格准则确定的壁面法向网格尺度是否满足热环境模拟精度要求，并对提出的壁面参数预估方法进行验证. 最后，开展了高超声速二维完全气体及真实气体效应算例、三维钝双锥算例的数值模拟. 研究表明，壁面参数预估方法是可靠的；自由来流参数网格雷诺数准则随来流静温变化所确定的网格尺度变化趋势与物理推理相悖；基于壁面参数网格雷诺数和MFP两种准则主要受壁面温度影响，其在同一条件下所确定的网格尺度基本一致，并满足热环境模拟的精度要求.

关 键 词：网格准则；气动热；高超声速；数值计算；网格雷诺数
中图分类号：V211.3
文献标识码：A **文章编号**：1001-5965(2015)04-0594-07

近年来，高超声速临近空间飞行器迅速发展，随之带来的飞行器热防护问题日益突出，而气动热环境的准确预测对飞行器热防护系统的设计至关重要. 随着数值方法和计算机硬件的迅速发展，计算流体力学（CFD）方法逐渐成为气动热环境预测的重要手段. 但运用CFD方法模拟气动热环境的难点在于其精度受多种因素影响，如离散方法、边界条件、数值格式和计算网格等[1-2]，其中壁面法向网格尺度特别是壁面法向第1层网格高度被认为是较重要的影响因素[3].

临近空间飞行器进行高空高马赫数飞行时，由于高空大气密度较小，外部流动一般处于层流状态. 目前，层流状态下气动热环境预测的壁面法向网格尺度通常由基于壁面法向第1层网格高度的网格雷诺数确定. 网格雷诺数确定方法有两种：一种为基于壁面参数的网格雷诺数[4-5]，另一种为基于自由来流参数的网格雷诺数[6-7]. 除上述方法外，2010年程晓丽等提出了基于气体分子平均自由程（MFP）的壁面网格尺度确定准则[8]. 以上3种方法中，基于壁面参数的网格雷诺数准则在计算前无法完全确定壁面参数，只能通过多套网格进行测试，导致该方法不具有实用性，因此，很多学者采用基于自由来流参数的网格雷诺数准则确定壁面法向网格尺度[6-7]. 然而，直接运用自由来流参数计算网格雷诺数是否可行，与壁面参数网格雷诺数准则有何差异目前并无确切的研

收稿日期：2014-05-19；**录用日期**：2014-08-25；**网络出版时间**：2014-09-05 16:00
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0288.html
基金项目：中央高校基本科研业务费专项资金资助项目（YWF-14-HKXY-002）
作者简介：张智超（1988—），男，河北唐山人，博士生，zhichaozhang@ase.buaa.edu.cn
*通讯作者：高振勋（1982—），男，河北廊坊人，讲师，gaozhenxun@buaa.edu.cn，主要研究方向为高超声速空气动力学、计算流体力学.

引用格式：张智超，高振勋，蒋崇文，等. 高超声速气动热数值计算壁面网格准则[J]. 北京航空航天大学学报，2015，41(4)：594-600. Zhang Z C，Gao Z X，Jiang C W，et al. Grid generation criterions in hypersonic aeroheating computations[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：594-600 (in Chinese).

究. MFP 准则虽然有较好的物理意义,但其仍需要先在粗网格上进行计算得到壁面密度才能确定壁面法向网格尺度.

此外,飞行器以高超声速飞行时,由于高马赫数导致激波后温度迅速升高,当气体温度足够高时,振动能激发,气体发生离解甚至电离,出现真实气体效应现象,该现象会强烈影响气动热环境. 对于有真实气体效应现象的流动进行气动热数值模拟时,上述各壁面法向网格尺度确定准则是否适用仍需要进行深入研究.

本文将开展高超声速气动热数值计算壁面网格准则研究,首先针对壁面参数网格雷诺数和 MFP 准则提出壁面参数快速估算方法,并对上述估算方法确定的壁面参数进行验证. 其次比较分析了 3 种准则在完全气体及真实气体效应下确定的壁面法向网格尺度的适用性,同时对自由来流变化对网格准则的影响进行了研究. 最后对上述不同网格准则的分析结果进行了三维复杂外形的验证研究.

1　壁面法向网格雷诺数准则

1.1　基于壁面参数的网格雷诺数准则

Papadopoulos 等通过研究[4]认为,热流的准确预测必须对壁面附近温度梯度分辨能力进行控制,并提出了以网格雷诺数作为控制参数来确定壁面法向第 1 层网格高度的方法.

$$Re_{\text{cellw}} = \frac{\rho_w c_w \Delta n_w}{\mu_w} \tag{1}$$

式中,Re_{cellw}为基于壁面参数的网格雷诺数;ρ_w 为壁面密度;c_w 为壁面声速;μ_w 为壁面流体黏性系数;Δn_w 为壁面法向第 1 层网格尺度.

Men'shov 等通过研究[5]发现,网格雷诺数的大小对高超声速气动热环境预测尤为重要,网格雷诺数即使为 10,计算结果也可能产生 100% 的误差. 其研究还认为网格雷诺数应不大于 3 才能保证热流计算结果的准确性. 由于在计算前无法完全确定壁面参数,因此只能通过多套网格进行测试. 为了解决上述问题,本文提出了针对基于壁面参数网格雷诺数的网格尺度预估方法.

由于热流峰值一般位于驻点附近,驻点附近网格尺度一般最小,所以只需得到驻点附近壁面参数即可用上述准则确定壁面网格法向尺度. 首先确定驻点附近壁面密度,对于完全气体壁面密度可采用状态方程计算:

$$\rho_w = \frac{p_s}{RT_w} \tag{2}$$

式中,p_s 为驻点壁面压强;T_w 为等温壁面条件中设置的壁面温度;R 为气体常数. 临近空间飞行器高马赫数飞行时头部往往产生脱体激波,可近似认为驻点壁面压强等于驻点线上激波波后总压:

$$p_s \approx p_{02} \tag{3}$$

式中,p_{02}为驻点线上波后总压. 驻点线上波后总压可根据正激波前后总压关系式获得:

$$\frac{p_{02}}{p_{01}} = \left(\frac{2\gamma}{\gamma+1}Ma_\infty^2 - \frac{\gamma-1}{\gamma+1}\right)^{-\frac{1}{\gamma-1}} \cdot \left[\frac{(\gamma+1)Ma_\infty^2}{(\gamma-1)Ma_\infty^2+2}\right]^{\frac{\gamma}{\gamma-1}} \tag{4}$$

式中,p_{01}为自由来流总压;Ma_∞ 为自由来流马赫数;γ 为比热比. 自由来流总压 p_{01}可根据一维定常绝热流驻点总参数关系式获得:

$$p_{01} = p_1\left(1 + \frac{\gamma-1}{2}Ma_\infty^2\right)^{\frac{\gamma}{\gamma-1}} \tag{5}$$

式中,p_1 为自由来流压强. 在已知自由来流马赫数、静压或总压以及壁面温度的情况下可根据驻点壁面估算公式快速估算出驻点壁面密度值. 对于存在真实气体效应的流动,若给定完全催化冷壁条件,壁面附近原子完全复合成自由来流气体分子,气体常数 R 可认为与自由来流值相同,因此上述壁面密度估算方法同样可适用于此种情况. 而完全非催化壁面条件下,上述公式是否适用将在随后进行数值验证.

式(1)中的壁面声速 c_w 和壁面流体黏性系数 μ_w 采用声速公式及 Sutherland 公式给出:

$$c_w = \sqrt{\gamma R T_w} \tag{6}$$

$$\mu_w = \mu_0\left(\frac{T_w}{T_0}\right)^{\frac{3}{2}}\frac{T_0+C}{T_w+C} \tag{7}$$

式中,T_0 为参考温度;μ_0 为参考温度所对应的黏性系数;C 为常数.

由式(2) ~ 式(7)及式(1)可估算出基于壁面参数的网格雷诺数,进而可根据需要控制网格雷诺数大小确定出壁面法向第 1 层网格高度 Δn_w. 由上述分析可以看出,基于壁面参数网格雷诺数确定的壁面法向网格尺度与自由来流马赫数、压强和壁面温度有关,在这些条件给定的情况下,与来流静温无关.

1.2　基于自由来流参数的网格雷诺数准则

为了避免使用壁面参数计算网格雷诺数,很多学者[6-7]采用了自由来流参数计算网格雷诺数,并通过研究发现某些条件下当网格雷诺数接近于 1 时,热流计算结果与实验值吻合得较好. 自由来流网格雷诺数由式(8)给出:

$$Re_{\text{cellf}}=\frac{\rho_\infty u_\infty \Delta n_w}{\mu_\infty} \tag{8}$$

式中,ρ_∞ 为自由来流密度;u_∞ 为自由来流速度;μ_∞ 为自由来流黏性系数;Re_{cellf} 为基于自由来流参数的网格雷诺数. 使用该准则时,先根据自由来流参数计算出单位长度雷诺数:

$$Re_\infty=\frac{\rho_\infty u_\infty}{\mu_\infty} \tag{9}$$

之后基于自由来流参数的网格雷诺数为 1 来获得壁面法向第 1 层网格高度:

$$\Delta n_w=\frac{1}{Re_\infty} \tag{10}$$

由式(8)~式(10)可以看出,基于自由来流参数网格雷诺数确定的壁面法向网格尺度与自由来流马赫数、温度和密度相关,在这些条件给定的情况下,与壁面温度无关.

2 壁面法向网格尺度 MFP 准则

程晓丽等摒弃了网格雷诺数的概念,引入气体分子平均自由程分析网格尺度,提出了一种新的壁面网格法向尺度确定方法[8]. 该方法按照气体动力学的观点,认为壁面传热是大量微观气体分子通过热运动与壁面碰撞进行能量交换的宏观反映,只有当壁面法向第 1 层网格高度不小于一个分子平均自由程时,才能包含所有与壁面碰撞的气体分子. 基于上述分析,建立了基于平均自由程(MFP)的壁面网格法向尺度确定准则.

MFP 准则壁面法向第 1 层网格高度 Δn_w 根据式(11)确定:

$$\Delta n_w=c_w\lambda_w \tag{11}$$

式中,c_w 为修正参数;λ_w 为壁面附近气体分子平均自由程,可由式(12)进行计算:

$$\lambda_w=\frac{m}{\sqrt{2}\pi d^2}\cdot\frac{1}{\rho_w} \tag{12}$$

式中,m 为气体分子质量;d 为气体分子有效直径.

取式(11)中 c_w 等于 1,即认为壁面法向网格尺度为一个分子平均自由程,式(11)和式(12)可进一步简化为

$$\Delta n_w=C_{\text{MFP}}\cdot\frac{1}{\rho_w} \tag{13}$$

对于空气,分子量 $M=28.9$ kg/kmol,分子有效直径 $d=3.5\times10^{-10}$ m,因此可近似计算出式(13)中的系数 C_{MFP},近似为

$$C_{\text{MFP}}\approx 0.88\times10^{-7}\ \text{kg/m}^2 \tag{14}$$

壁面法向网格尺度 MFP 准则形式简单,但在计算前无法准确获得壁面密度,程晓丽等采用的策略为先进行粗网格计算,得到壁面密度,再重新在由 MFP 准则确定的密网格上进行数值计算. 这种方式针对复杂外形网格量较大时实用性较差. 本文采用式(2)~式(5)提出的壁面密度估算方法,与上述 MFP 准则结合能够在计算之前预估高超声速热环境模拟中壁面网格法向尺度. 此外,MFP 准则确定的壁面法向网格尺度与自由来流压强、马赫数以及壁面温度相关,在这些条件给定的情况下,与来流静温无关,这一点与基于壁面参数的网格雷诺数有相同的特性.

3 壁面网格准则研究的数值实验

本节将针对上述 3 种壁面网格准则分别在完全气体以及考虑真实气体效应的条件下开展数值实验,并对本文提出的壁面密度估算公式进行数值验证. 根据阎超等的研究[9],AUSM(Advection Upstream Splitting Method)格式在热流计算精确性方面具有优势,所以验证算例均采用 AUSM 格式进行计算. 此外,本文是基于国家计算流体力学实验室 ACANS(Aerodynamic Combustion and Aerothermodynamics Numerical Simulation)有限差分程序开展数值计算研究,该程序已经得到相关研究者[10-11]的验证.

3.1 完全气体壁面网格准则数值实验研究

3.1.1 完全气体壁面参数估算公式验证

为验证本文提出的壁面密度估算公式的准确性,本文选取了 Holden 等[12]的高超声速圆柱绕流实验作为验证算例,圆柱直径 $D=0.076$ m. 实验条件为:自由来流马赫数 $Ma_\infty=16.34$,静温 $T_\infty=52$ K,静压 $p_\infty=82.95$ Pa,壁面温度 $T_w=294.4$ K,雷诺数 $Re=3.94\times10^6$/m. 由上述条件根据自由来流参数网格雷诺数准则确定壁面法向第 1 层网格高度约为 2.57×10^{-7} m;根据驻点壁面密度估算公式得到的驻点壁面密度为 0.338 kg/m^3,由 MFP 准则确定壁面法向第 1 层网格高度约为 2.6×10^{-7} m;在该壁面法向网格尺度条件下,根据式(1)得到基于壁面参数的网格雷诺数约为 1.66,满足小于 3 的要求. 在此算例条件下,3 种网格准则给出的壁面法向第 1 层网格高度恰好一致. 此外,为比较不同壁面法向网格尺度对热流计算结果的影响,另取其他 5 种不同的壁面法向网格尺度,各网格尺度及基于壁面参数的网格雷诺数如表 1 所示.

图 1 给出了不同壁面法向网格尺度得到的热

流 q_w 计算结果,其中 0°位置为驻点. 由图 1 可以看到,粗网格计算所得热流与实验值存在较大的偏差,Grid 5 和 Grid 6 计算结果与实验值吻合较好,Grid 5 壁面参数网格雷诺数略大于 3. 上述 3 个壁面网格准则在此算例条件下确定的壁面法向网格尺度均满足了热流计算精度要求.

表 1　算例 1:壁面法向网格尺度及网格雷诺数

Table 1　Wall normal grid scale and cell Reynolds number for Case 1

网格编号	壁面法向第 1 层网格高度/m	网格雷诺数
Grid 1	1.0×10^{-5}	63.90
Grid 2	6.0×10^{-6}	38.34
Grid 3	3.0×10^{-6}	19.17
Grid 4	1.2×10^{-6}	9.58
Grid 5	6.0×10^{-7}	3.83
Grid 6	2.6×10^{-7}	1.66

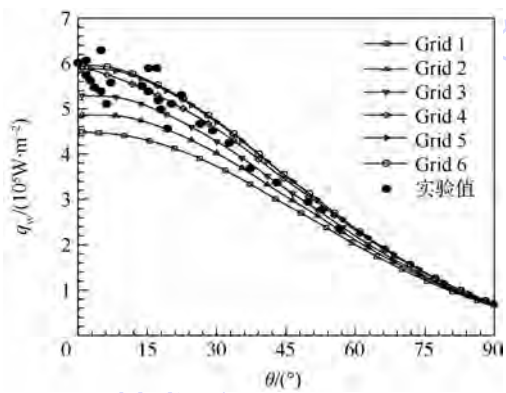


图 1　不同壁面法向网格尺度热流计算结果

Fig. 1　Comparison between experimental and computational heat flux values with different wall normal grid scales

根据图 1 计算结果,Grid 6 计算结果与实验值最为吻合,表 2 给出了 Grid 6 驻点壁面密度计算值与估算值的对比. 可以看出本文提出的驻点密度估算值与数值计算误差约为1.8%,表明本文提出的由驻点壁面密度估算公式结合 MFP 准则的策略在完全气体条件下是适用的.

表 2　完全气体驻点壁面密度估算与计算值对比

Table 2　Comparison between estimated and computational density value of stagnation point in ideal gas condition　kg/m³

参数	估算值	计算值
数值	0.338	0.332

3.1.2　来流静温变化对网格准则影响

由基于自由来流及壁面参数的网格雷诺数准则和 MFP 准则确定的壁面法向网格尺度在 3.1.1 节Holden 等实验条件下基本一致. 然而,是否在所有情况下 3 种网格准则给出的壁面法向网格尺度均一致则需要进一步分析,本节将在 3.1.1 节实验条件下,改变来流静温,研究分析来流静温变化对网格准则的影响.

将上述实验条件提高来流静温至 145 K,其他条件不变. 基于自由来流参数网格雷诺数准则确定的壁面法向第 1 层网格高度约为 1.27×10^{-6} m,与静温变化前相比明显增大,该尺度下基于壁面参数的网格雷诺数约为 8.10;由 MFP 准则确定的壁面法向第 1 层网格高度仍为 2.6×10^{-7} m,该网格尺度在此条件下的壁面法向网格雷诺数仍为 1.66. 此外,由于本节主要针对来流静温和壁面温度变化对网格准则的影响开展研究,虽然该条件下可能发生真实气体效应,计算模型中并未进行考虑,真实气体效应影响将在后续内容进行研究.

图 2 给出了表 1 中 6 种不同壁面法向网格尺度在来流静温提高至 145 K 条件下的热流计算结果. 与图 1 相比,来流静温提高后,热流显著提高,Grid 5 与 Grid 6 计算结果也出现较大差异,根据 Men'shov 等的研究[5],壁面参数网格雷诺数应小于 3 才可能保证热流计算精度要求,此时Grid 5 可能不再满足热流计算精度要求. 在等温壁面条件下,自由来流温度提高后,物理推理认为由于壁面附近温度梯度增加,应对网格进行适当加密才能保证获得足够精确的热流计算结果.

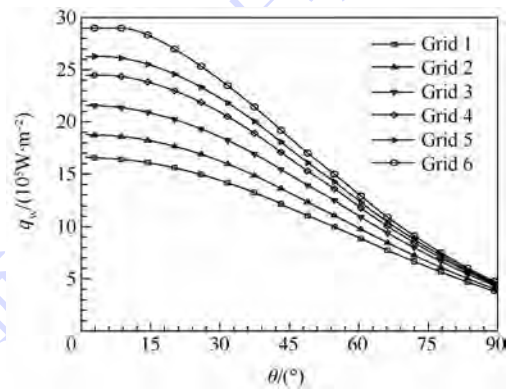


图 2　提高来流静温后不同壁面法向网格尺度热流计算结果

Fig. 2　Result of heat flux values with different wall normal grid scales when the wall temperature increases

基于自由来流参数的网格雷诺数准则在自由来流温度提高后,其确定的壁面法向第 1 层网格高度反而由 2.57×10^{-7} m 稀疏至 1.27×10^{-6} m,其壁面参数网格雷诺数由 1.66 增大至 8.10. 鉴于该准则确定的网格尺度随来流静温的变化与物理推理趋势相反,不建议将其作为高超声速热环境模拟壁面法向网格尺度确定的准则.

3.2 真实气体效应下壁面网格准则数值实验研究

真实气体效应对热环境预测模拟有重要影响,本节开展上述 3 个壁面网格准则在真实气体效应条件下适用性的研究. 算例选取了 Klaus 等的高超声速圆柱绕流实验[13],圆柱直径 D = 0.09 m. 实验条件为:自由来流马赫数 Ma_∞ = 8.78,自由来流静温 T_∞ = 694 K,静压 p_∞ = 687 Pa,密度 ρ_∞ = 0.003 26 kg/m^3,自由来流总温约为 11 393.8 K,壁面温度 T_w = 300 K. 自由来流空气由 N_2,O_2,NO,N 和 O 这 5 种组分组成,其质量分数分别为 ξ_{N_2} = 0.735 55,ξ_{O_2} = 0.134,ξ_{NO} = 0.050 9,ξ_N = 1.0 × 10^{-9},ξ_O = 0.079 55. 各反应组分及混合气体的输运系数、热力学函数等均由文献[14]给出,化学反应动力学模型采用其中的 5 组分 6 化学反应模型. 由上述条件根据自由来流参数网格雷诺数准则确定的壁面法向网格尺度约为 2 × 10^{-6} m,基于壁面参数的网格雷诺数约为 30.99;根据驻点壁面密度估算公式得到的估算值为 0.796 kg/m^3,由 MFP 准则确定的壁面法向网格尺度约为 1.1 × 10^{-7} m,基于壁面参数的网格雷诺数约为 1.65. 取介于二者之间的一个壁面法向网格尺度,以对比真实气体效应下不同壁面法向网格尺度对热流计算结果影响(Grid 2),基于壁面参数网格的雷诺数如表 3 所示.

表 3 算例 2:壁面法向网格尺度及网格雷诺数

Table 3 Wall normal grid scale and cell Reynolds number for Case 2

网格编号	壁面法向第 1 层网格高度/m	网格雷诺数
Grid 1	2.0 × 10^{-6}	30.99
Grid 2	8.0 × 10^{-7}	11.97
Grid 3	1.1 × 10^{-7}	1.65

根据文献[13]可知,此算例数值计算中应采用完全催化壁面条件. 为研究驻点壁面估算公式是否适用于完全非催化壁面条件下驻点壁面密度估算,算例中分别给定了完全催化和完全非催化壁面条件进行计算,得到驻点壁面密度计算值.

图 3 给出了表 3 中不同壁面法向网格尺度的网格热流计算结果,其中 0°位置为驻点. 由 MFP 准则确定的网格(Grid 3)也满足基于壁面参数网格雷诺数小于 3 的要求,其热流计算结果与实验值吻合较好,说明 MFP 准则和壁面参数网格雷诺数准则对发生真实气体效应现象的流动有较好的适应性. 而自由来流参数网格雷诺数准则确定的网格(Grid 1)热流计算结果与实验值出现较大的偏差,其壁面参数网格雷诺数也远大于 3. 由 3.1 节分析可知,基于自由来流参数的网格雷诺数准则不再适用,主要是由于本算例来流静温较高造成的.

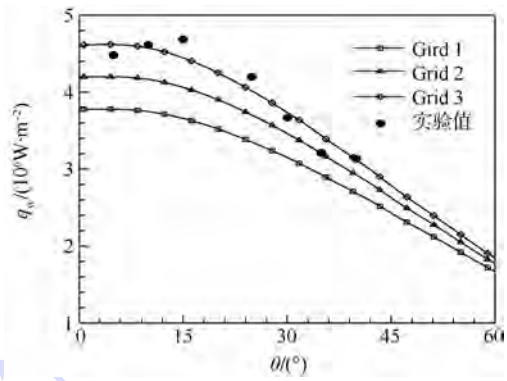


图 3 不同壁面法向网格尺度下考虑化学反应的热流对比

Fig. 3 Comparison between experimental and computational heat flux values with different wall normal grid scales considering chemical reaction

图 4 和图 5 分别给出了 Grid 3 网格驻点线上温度 T 与氧气质量分数变化曲线. 其中,自由来流方向为 x 轴正方向,驻点位置为 x = −0.045 m. 可以看出,自由来流空气经过激波后温度迅速升

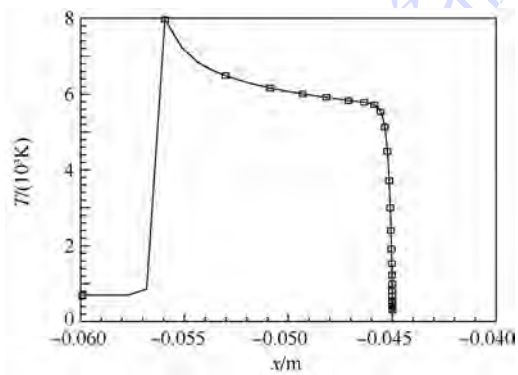


图 4 驻点线上温度变化曲线

Fig. 4 Temperature variation along the stagnation point line

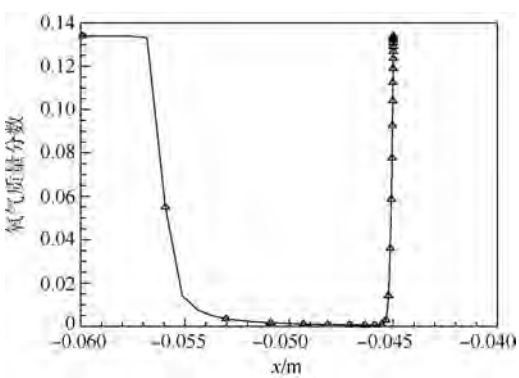


图 5 驻点线上氧气质量分数变化曲线

Fig. 5 Mass fraction variation of oxygen molecules along the stagnation point line

北航学报 赠阅

高,氧气分子发生离解,其质量分数迅速降低至 0 左右,由于壁面为完全催化条件,氧原子在壁面附近复合成气体分子,所以氧气分子质量分数在壁面附近升高,说明化学反应模型对真实气体效应进行了正确的模拟.

表 4 给出了 Grid 3 网格分别在完全催化和完全非催化壁面条件下驻点壁面密度计算值与估算值对比. 由驻点壁面密度估算公式所估算的值与完全催化壁面条件计算得到的结果误差约为 3.9%,与完全非催化壁条件计算结果误差约为 12.7%. 驻点壁面密度估算公式在产生真实气体效应并给定完全壁面催化条件下可获得较准确的驻点壁面密度,在完全非催化壁面条件下,误差略大于 10%. 因此,真实气体效应条件下采用本文提出的驻点壁面密度估算公式并进一步由相应准则确定壁面法向网格尺度是一种可行的方法.

表 4　驻点壁面密度估算值与完全催化和非催化壁面条件下的计算值对比

Table 4　Comparison between estimated and computational density values of stagnation point in fully catalytic and equilibrium catalytic wall condition

kg/m^3

参数	估算值	计算所得密度值	
		完全催化壁面	完全非催化壁面
数值	0.796	0.766	0.706

3.3　复杂外形适用性

由前述算例验证可知,MFP 准则确定的壁面法向网格尺度也满足壁面参数网格雷诺数小于 3 的要求,这两种准则在二维问题中有较好的表现,接下来选取了高超声速钝双锥实验[15]研究其三维复杂外形问题适用性. 该钝锥体前缘半径为 3.835 mm,前锥半锥角为 12.84°,后锥半锥角为 7°,前锥距头部距离 L_f = 69.55 mm,后锥距头部距离即锥体全长 L = 122.24 mm. 实验条件为:自由来流马赫数 Ma_∞ = 9.86,来流静温 T_∞ = 48.88 K,静压 p_∞ = 89.92 Pa,密度 ρ_∞ = 0.004 27 kg/m^3,壁面温度 T_w = 300 K,攻角为 0°. 根据驻点密度估算公式以及 MFP 准则确定的壁面法向第 1 层网格高度约为 1.0×10^{-6} m,壁面参数网格雷诺数约为 1.65,MFP 准则确定的壁面法向网格尺度满足壁面参数网格雷诺数小于 3 的要求.

图 6 给出了钝双锥轴向热流分布. 其横纵坐标进行了无量纲化处理,由 Fay-Riddell 公式估算得到参考热流 q_{ws} = 3.69×10^5 W/m^2,特征长度为锥体总长 L = 122.24 mm. 计算结果表明该壁面法向网格尺度能够满足热环境模拟中热流计算精度的要求,说明 MFP 准则和壁面参数网格雷诺数准则对复杂外形流动仍然具有良好的适用性.

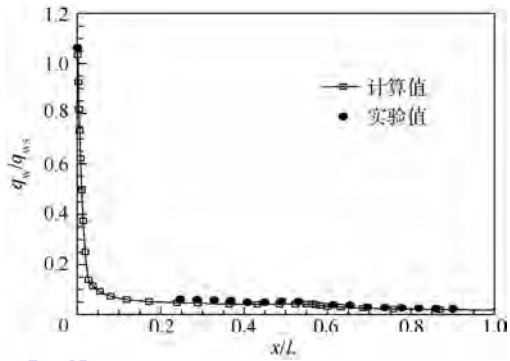


图 6　钝双锥表面热流轴向分布

Fig. 6　Heat flux distribution along axial direction of blunt bicone

4　结　论

1) 针对基于壁面参数网格雷诺数和 MFP 两种准则无法在计算前确定壁面密度的问题,提出了驻点壁面密度估算方法,数值模拟结果显示其在不同流动条件下均可给出可接受的驻点壁面密度估算值,从而使得壁面参数网格雷诺数和 MFP 两种准则无需进行多套网格或粗网格试算而直接预估出壁面法向网格尺度,增加了此两种准则的实用性.

2) 数值实验及理论分析表明:基于自由来流参数的网格雷诺数准则确定的壁面法向网格尺度与壁面温度无关;在给定条件下,提高来流静温而增大壁面附近温度梯度时,其确定的壁面法向网格尺度反而增大,这与物理推理相悖. 数值模拟与实验数据的对比也证明在来流静温较高时该准则所确定的壁面法向网格尺度无法给出准确的气动热结果. 因此,不建议使用该准则来确定高超声速气动热环境数值计算中壁面法向网格尺度.

3) 基于壁面参数网格雷诺数和 MFP 两种准则确定的壁面法向网格尺度与自由来流温度无关,主要受壁面温度影响. 并且不同条件下这两种准则所确定的壁面法向网格尺度基本一致. 本文开展的高超声速二维完全气体及真实气体效应算例、三维钝双锥算例的数值模拟结果均显示,这两种准则确定的壁面法向网格尺度满足热环境模拟的精度要求.

4) 最后,结合本文提出的壁面参数预估方法,可将壁面参数网格雷诺数和 MFP 两种准则作为高超声速气动热数值计算中壁面法向网格尺度的确定方法.

参考文献（References）

[1] Olynick D R, Henline W D. Numerical benchmarks for Navier-Stokes heating calculations on access to space vehicles, AIAA-1995-2078[R]. Reston: AIAA, 1995.

[2] Siddiqui M S, Hoffman K A, Chiang S T, et al. A comparative study of the Navier-Stokes solvers with emphasis on the heat transfer computations of high speed flows, AIAA-1992-0835[R]. Restion: AIAA, 1992.

[3] Hoffman K A, Siddiqui M S, Chiang S T. Difficulties associated with the heat flux computations of high speed flows by the Navier-Stokes equations, AIAA-1991-0457[R]. Reston: AIAA, 1991.

[4] Papadopoulos P, Venkatapathy E, Prabhu D, et al. Current grid-generation strategies and future requirements in hypersonic vehicle design, analysis and testing[J]. Applied Mathematical Modelling, 1999, 23(9): 705-735.

[5] Men'shov I S, Nakamura Y. Numerical simulations and experimental comparison for high-speed nonequilibrium air flows[J]. Fluid Dynamics Research, 2000, 27(5): 305-334.

[6] Klopfer G H, Yee H C. Viscous hypersonic shock-on-shock interaction on blunt cowl lips, AIAA-1988-0233[R]. Reston: AIAA, 1988.

[7] 潘沙, 冯定华, 丁国昊, 等. 气动热数值模拟中的网格相关性及收敛[J]. 航空学报, 2010, 31(3): 493-499.
Pan S, Feng D H, Ding G H, et al. Grid dependency and convergence of hypersonic aerothermal simulation[J]. Acta Aeronautica et Astronautica Sinica, 2010, 31(3): 493-499(in Chinese).

[8] 程晓丽, 艾邦成, 王强. 基于分子平均自由程的热流计算壁面网格准则[J]. 力学学报, 2010, 42(6): 1083-1089.
Cheng X L, Ai B C, Wang Q. A wall grid scale criterion based on the molecule mean free path for the wall heat flux computations by the Navier-Stokes equations[J]. Chinese Journal of Theoretical and Applied Mechanics, 2010, 42(6): 1083-1089(in Chinese).

[9] 阎超, 禹建军, 李君哲. 热流 CFD 计算中格式和壁面法向网格尺度若干问题研究[J]. 空气动力学学报, 2006, 24(1): 125-130.
Yan C, Yu J J, Li J Z. Scheme effect and grid dependency in CFD computations of heat transfer[J]. Acta Aerodynamica Sinica, 2006, 24(1): 125-130(in Chinese).

[10] Gao Z X, Jiang C W, Lee C H. Improvement and application of wall function boundary condition for high-speed compressible flows[J]. Science China Technological Sciences, 2013, 56(10): 2501-2515.

[11] Gao Z X, Lee C H. A numerical study of turbulent combustion characteristics in a combustion chamber of a scramjet engine[J]. Science China Technological Sciences, 2010, 53(8): 2111-2121.

[12] Holden M S, Wieting A R, Moselle J R, et al. Studies of aerothermal loads generated in regions of shock/shock interaction in hypersonic flow, AIAA-1988-0477[R]. Reston: AIAA, 1988.

[13] Klaus H, Jan M S, Sebastian K, et al. Cylinder shock layer density profiles measured in high enthalpy flows in HEG, AIAA-2002-2913[R]. Reston: AIAA, 2002.

[14] Gupta R N, Yos J M, Thompson R A, et al. A review of reaction rates and thermodynamic and transport properties for the 11 species air model for chemical and thermal nonequilibrium calculations to 30 000 K, NASA-TM-101528[R]. Washington, D. C.: NASA, 1989.

[15] Miler C G III. Experimental and predicted heating distributions for biconics at incidence in air at Mach 10, NASA-TP-2334[R]. Washington, D. C.: NASA, 1984.

Grid generation criterions in hypersonic aeroheating computations

ZHANG Zhichao, GAO Zhenxun*, JIANG Chongwen, LEE Chunhian

(School of Aeronautic Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: Three wall normal grid generation criterions in hypersonic aeroheating computations are analyzed, including mean free path (MFP), and cell Reynolds number based on freestream parameters and wall parameters respectively. A wall parameter prediction method is proposed to avoid trial computations when MFP and cell Reynolds number based on wall parameters criterions are used. The proposed method is utilized and combined with different grid generation criterions to generate several grids. A comparison with the experimental data is conducted to test if the wall normal mesh size determined by the grid criterions can satisfy the precision requirement of thermal environment simulation, and also to verify the wall parameter prediction method proposed. Numerical simulations of two-dimensional cylinder with perfect and real gas effects, and three-dimensional blunt bicone are carried out. It can be concluded that the wall parameter prediction method is reliable, the criterion of the cell Reynolds number based on freestream parameters would amplify the wall grid scale with the increase of the freestream temperature, both MFP and cell Reynolds number based on wall parameters are influenced by wall temperature rather than by freestream temperature, and the grid scale determined by these two criterions are consistent and can satisfy the precision requirements of hypersonic aeroheating computations.

Key words: grid criterion; aeroheating; hypersonic; computation; cell Reynolds number

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0284

空间科学任务协同设计论证平台

邓丽[*1], 韩潮[1], 曹晋滨[1], 宋友[2]
(1. 北京航空航天大学 宇航学院, 北京 100191; 2. 北京航空航天大学 软件学院, 北京 100191)

摘 要: 空间科学任务协同设计论证存在设计方案耦合强,数据一致性差,数据变更难,数据与流程脱节等问题.为了解决上述问题,建立了典型空间科学任务的多层数字化模型;采用图形化方式对论证流程进行建模,并建立流程与数据的映射关系;通过共享数据池的方式为多岗位用户提供多方案数据协同机制;采用消息总线对数据变更进行及时提醒;利用方案依赖关系矩阵来判别方案耦合关系,最终自动合并任务总体设计方案.作为一个分布式平台,空间科学任务协同设计平台采用 Eclipse RCP 和 Spring 技术架构,整合了 Hibernate、工作流 Activiti5 等中间件,提供统一门户,支持多岗位、多任务、多方案、多版本的管理能力,提供论证流程监控、数据协同交互等功能.结合某空间科学任务论证验证了该平台的有效性.

关 键 词: 空间科学任务; 协同设计; 数据一致性; 论证流程; 方案合并
中图分类号: TP391.7; V57
文献标识码: A **文章编号**: 1001-5965(2015)04-0601-08

由于空间科学任务立项论证需求越来越多,为了提高概念设计阶段空间科学任务论证的效率,协同设计思想越来越多地应用到空间科学任务的系统论证中.空间科学任务论证是设计岗位依据规范的论证流程开展协同设计的过程.以往的任务论证在数据、流程及协同方式上有如下特点:

1) 基于文档的数据管理模式.该模式导致了文档变更及版本更新时一致性差、数据提取难、方案耦合关系不可管理等问题.

2) 侧重于计划进度的流程管理模式.该模式导致了流程与数据分离,任务完成率不便于控制,论证过程不可追溯等问题.

3) 采用私下沟通和召开会议的协同模式.私下沟通的形式带来数据变更通知不及时、数据不一致等问题;会议上不能立即解决的设计问题需在会议后完成,数据变更只能等到下次会议才能通知其他设计岗位,缺乏及时的沟通和提醒也是造成论证周期长的原因之一.

因此,针对设计方案耦合关系管理难、数据一致性差、数据与流程脱节等问题,如何组织管理多岗位、多方案数据,建立数据与流程映射关系,提供高效的协同机制,一直是众多学者期望解决的关键问题.

针对空间任务的立项论证需求,国外成立了专门的协同设计中心(CDF,Concurrent Design Facility)[1-5],提供专业设计的工具和数据协同管理软件.如美国国家航空航天局(NASA)普遍采用 IceMaker 工具,欧洲航天局(ESA,European Space Agency)协同设计中心采用整合设计模型(IDM,Integrated Design Model),数据协同管理均以 Excel 工作表为基础.Excel 对于结构化数据的统计处理有很强的优势,但数据表的形式不利于复杂数据管理、方案耦合管理、数据与流程的统一.当

收稿日期: 2014-05-19; **录用日期**: 2014-06-23; **网络出版时间**: 2014-06-30 09:49
网络出版地址: www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0284.html
基金项目: 中国科学院空间科学战略性先导科技专项资助项目 (Y129163EHS)
* **通讯作者**: 邓丽(1981—),女,湖北荆州人,副研究员,博士生,dengli@nssc.ac.cn,主要研究方向为空间任务的协同设计和系统仿真.

引用格式: *邓丽, 韩潮, 曹晋滨, 等. 空间科学任务协同设计论证平台[J]. 北京航空航天大学学报, 2015, 41(4): 601-608.*
Deng L, Han C, Cao J B, et al. Demonstration platform for collaborative design of space science missions [J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(4): 601-608 (in Chinese).

前,ESA 正在开展开放协同设计服务(OCDS, Open Concurrent Design Server)的研究,希望升级 IDM,提高数据管理能力.

面向复杂型号任务的工程总体,中国国内也开展了协同设计的相关研究.基于 Web 的企业级协同产品研制管理系统(AVIDM)[6],主要面向型号任务工程大总体;以计划为核心的项目管理平台(AVPLAN)[7],主要面向航天任务复杂型号研制过程;记录设备接口状态的接口数据单(IDS, Interface Data Sheet)系统[8],反映卫星与分系统,以及分系统之间的接口关系;集成平台下的总体数字化设计流程[9];面向复杂工程系统的多学科设计优化集成环境(系统综合设计优化平台,SDOF,System Design Optimization Framework)[10]等诸多的研究[11-16].针对空间科学任务协同设计论证,中科院空间科学与应用研究中心开展了相关的研究并建立了初级系统[13],数据协同管理也是以 Excel 工作表为基础.国内的研究主要服务于复杂型号任务的工程大总体,侧重于计划进度的管理、文档数据的协同,以及多学科建模和优化研究.

因此,本文在空间科学任务协同设计论证初级系统研究的基础上,构建空间科学任务协同设计论证平台,建立任务多层数字化模型、流程与数据映射模型、设计方案合并模型,解决上述关键问题.

1 论证特点

空间科学任务是一个复杂系统,涉及若干分系统,系统层次多、分系统间耦合性强.开展空间科学任务协同设计论证,具有如下特点.

1)论证任务具有特殊性、通用性.

每一个空间科学任务有其不同的任务目标,搭载不同的载荷设备,运行在不同的环境,采集不同的数据.同时,空间科学任务也有一定的通用性,均包含发射段、空间段、地面段.空间段包括多个航天器或单个航天器.任何一个空间科学任务的协同设计论证均需要开展任务分析、方案设计,最终确定可行性的任务总体设计方案.

2)论证过程具有迭代性、试探性、选择性.

空间科学任务论证中各分系统设计方案耦合性强,耦合数据的细微改动可能会引起依赖方案的很大波动,论证过程是一个不断权衡利弊、不断寻优的过程,需要多次的数据迭代、数据共享.方案的设计过程又具有一定的试探性,会产生许多的数据中间版本,这些版本均需要存储供后续借鉴使用.

3)数据海量性、耦合性.

任务论证包含诸多设计活动,每一个设计活动都有输入数据、过程数据和输出数据.输入数据既可来源于本岗位,也可来源于其他岗位,数据之间存在复杂的耦合依赖关系.

4)设计方案多样性、依赖性.

设计岗位针对任务需求开展设计工作,可能提出一套或多套设计方案.例如:设计岗位 1 设计了 A_1,A_2 两套方案,设计岗位 2 设计了 B_1,B_2 两套方案.在协同设计论证过程中,设计岗位的方案之间存在一定的依赖性.又如,A_1 方案依赖了 B_1 方案,A_2 方案依赖了 B_2 方案.从任务角度,需要管理设计方案间的耦合关系,根据方案间的依赖性汇总出任务总体设计方案.

5)数据版本一致性、可追溯性.

设计过程的迭代性、试探性和选择性也导致设计方案的数据版本更新频繁,而数据的耦合性对数据版本一致性提出了高要求.例如:设计岗位设计方案 A_1 数据版本发生变化时,应及时通知设计岗位设计方案 B_1 做相应的调整和修改.为了保证对流程的监管,需要建立数据与流程的映射关系,使得流程具有可追溯性.

6)分布式协同工作环境.

空间科学任务协同设计论证是一个多岗位协作的过程,团队成员分布式地并行开展工作.

从上述特点可以看出,基于空间科学任务协同设计论证中数据、流程的独特性,需要从空间科学任务的数字化建模、流程—数据的映射管理、设计方案依赖性管理、数据协同交互等出发,构建一个友好的空间科学任务协同设计论证平台.

2 论证平台设计

2.1 论证平台框架

空间科学任务协同设计论证平台可分为数据资源层、应用支撑层、业务应用层和访问门户层等,其体系架构如图 1 所示.

1)访问门户层.

访问门户层为管理员、需求提出方、设计岗位和项目负责人 4 类用户提供统一的登录和操作界面,确保正确的人使用正确的权限,并在平台上按照正确的流程执行正确的操作.管理员负责整个项目管理,创建论证任务、创建论证团队,配合项目负责人制定论证流程、确定数据模型和数据接口等.需求提出方提出任务需求,在整个设计过程中参与方案决策.设计岗位是分系统设计人员,如

北航学报 赠阅

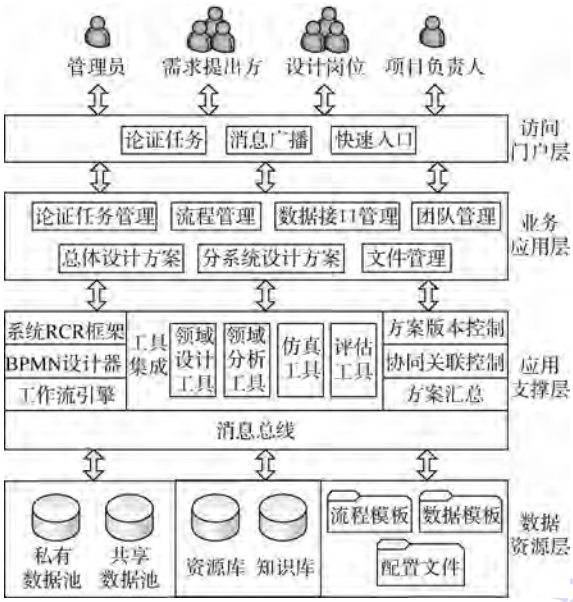


图 1　软件体系架构图

Fig. 1　Software architecture

轨道岗、有效载荷岗、结构岗、姿态岗等，设计岗位负责分系统方案设计与分析. 项目负责人审核其他岗位提交的设计方案，做出正确的决策，汇总所有设计人员的设计方案，形成完整的任务总体设计方案.

2）业务应用层.

业务应用层包括多个业务模块，为不同岗位用户提供不同的应用服务. 该层包含的业务模块有论证任务管理、流程管理、数据接口管理、团队管理、总体设计方案、分系统设计方案以及文件管理等.

其中，论证任务管理包括新建项目、配置方案流程、配置项目表单和配置项目用户等功能. 流程管理具有设计流程、部署流程、启动流程、查看流程和删除流程等功能. 数据接口管理具有设计表单、表单与岗位映射和表单与流程映射等功能. 方案管理具有新建方案、编辑方案和删除方案等功能. 文件管理具有导入文件、导出文件和文件查看/编辑等功能.

3）应用支撑层.

应用支撑层提供各类支撑模块，包括系统 Eclipse RCP 框架、BPMN 设计器、工作流引擎 Activiti5、方案版本控制、协同依赖控制、方案汇总、消息总线等. 同时，应用支撑层具有集成领域设计工具、领域分析工具、仿真工具和评估工具的开放接口.

4）数据资源层.

数据资源层提供任务论证的核心数据池、资源库、知识库和模板库. 私有数据池存储过程数据和结果数据；共享数据池存储审核通过的方案数据，供所有岗位查询和关联. 资源库和知识库包括常用地面站、运载、卫星平台和载荷等信息. 模板库提供了典型任务的论证流程模板、多层数字化模型模板及相关配置文件.

2.2　论证平台使用模式

针对空间科学任务协同设计论证平台的 4 类用户，其使用模式如下：

1）管理员登录平台，创建项目，创建论证团队；然后，基于流程模板创建项目论证流程，映射流程与岗位对应关系；基于数据模板创建任务多层数字化模型，映射流程与数据对应关系；之后，启动项目论证.

2）需求提出方登录平台，提出任务需求，审核设计方案.

3）设计岗位登录平台，创建设计方案，岗位自动领取流程分配的设计活动；开展设计工作，提交设计方案数据，若设计方案通过项目负责人的审核，则设计活动完成，若没有通过审核，则重新开展设计. 各设计岗位并行开展设计工作.

4）项目负责人登录平台，审核提交的设计方案，做正确性决策；并根据所有岗位的分系统设计方案合并出任务总体设计方案，对多套总体设计方案进行系统级的评估优化，确定最终的任务总体设计方案.

2.3　协同交互机制

由于论证过程的试探性，论证过程中许多中间数据需要存储，而对于阶段性数据则需要提交项目负责人审核通过并让所有岗位共享. 因此，将数据库区域划分为独立数据池和共享数据池. 独立数据池存储设计岗位设计方案的多个版本，记录所有设计方案数据的版本变化. 对于阶段性成果，由设计人员将其设计方案数据提交，项目负责人审核批准后将独立数据池的阶段性成果提交至共享数据池，共享数据池同样对提交的设计方案数据进行版本管理. 以数据为核心的思想贯穿于整个论证过程，数据池中记录了数据从论证开始阶段至结束阶段的演化过程，各论证环节的前后数据具有可追溯性.

采用消息总线在岗位之间进行消息的发送和接收，包括方案提交、方案审核、版本更新和数据变更等消息.

3　关键技术

3.1　空间科学任务的多层数字化模型

空间科学任务是一个典型的复杂工程系统，

北航学报 赠阅

由于其结构和功能的复杂性,在协同设计过程中需要对其进行分解,分配给不同的设计岗位完成.结合空间科学任务协同设计论证的实际需求,提出典型空间科学任务的多层数字化模型.

典型的空间科学任务可分为 6 层,如图 2 所示,每一层组成要素称为系统元素.

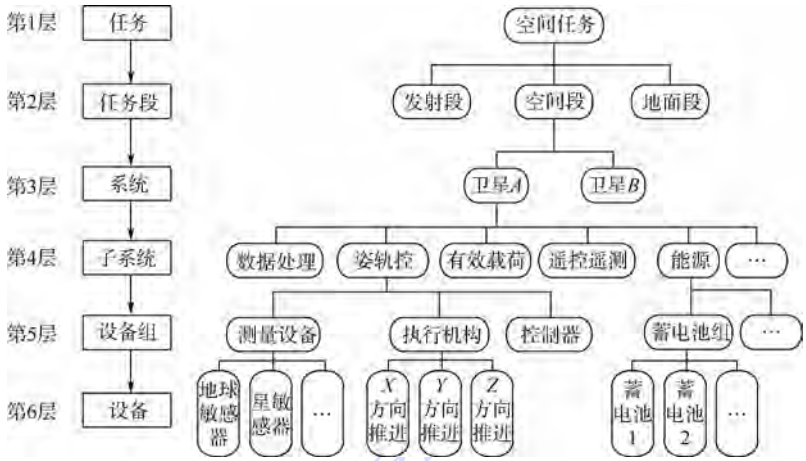


图 2 空间科学任务多层数字化模型示意图

Fig. 2 Multi-layer digital model for space science missions

上一层系统元素包含下一层系统元素.每个系统元素有惟一的名称标识、描述和数据集.这些数据集是整个任务的核心,也是各设计领域开展协同设计的关键;既包含设计方案数据,也包含设计领域间的输入输出接口数据;既包含结构化数据,也包含非结构化数据.数据具有名称、值、版本等属性.

在空间科学任务的多层数字化模型的基础上,建立设计方案与任务多层数字化模型的映射关系,建立设计岗位与设计方案的映射关系.每个设计岗位创建本岗位设计方案,岗位与设计方案为一对一、一对多关系;每个岗位可以创建一套方案,也可创建多套方案.每套设计方案包含分配给本岗位的所有系统元素.

因此,设计方案具有名称、版本、所属岗位等属性;设计方案数据集是任务数字化多层模型的子集,具有数据名称、数据值、数据版本等属性.

通过建立空间科学任务的多层数字模型,对论证数据统一管理,解决了数据一致性差、数据变更难等问题.针对不同的空间科学任务论证任务,可以对多数字化模型进行适应性的配置修改,满足任务要求.

3.2 流程建模及与数据的映射

空间科学任务概念设计阶段的论证可划分为 3 个阶段:论证准备阶段、任务分析阶段和方案设计阶段.各阶段开展的具体工作见文献[13].为解决“数据”与“流程”脱节、协同设计过程追溯等问题,对各阶段进行工作分解,确定每个阶段的所有设计活动,确定设计活动对应的岗位和数据,建立流程与数据的映射关系.

图 3 为姿态岗设计流程与数据映射示意图.论证准备阶段,设计活动“明确任务需求”映射到卫星 *A* 需求数据.任务分析阶段,设计活动“姿轨控系统指标分解”映射到姿轨控系统元素,包括姿态稳定方式、姿态指向精度和姿态稳定度.方案设计阶段,姿轨控分系统方案设计包括“计算干扰力矩”“敏感器配置”“执行机构配置”“控制器配置”和“姿态指向和姿态稳定度分析”等设计活动;“计算干扰力矩”映射到卫星轨道、卫星结构等系统元素;“敏感器配置”映射到测量设备系统元素;“执行机构配置”映射到执行机构系统元素;“控制器配置”映射到控制器系统元素;“姿态指向和姿态稳定度分析”映射到姿轨控系统元素.论证流程中的设计活动可映射到多层数字化模型的每一层系统元素.

采用工作流 Activiti5 中间件技术对任务流程建模,Activiti5 使用标准的 BPMN2.0(业务流程建模标注,Business Process Modeling Notation)流程定义文件来描述工作流.BPMN2.0 流程定义文件包含了流程的流转序列,步骤节点以及各个节点上相关的用户、变量信息等流程元素,以图形化的方式来显示或编辑流程定义文件.

3.3 多岗位多方案的合并算法

空间科学任务论证是一个多岗位协同设计的过程,设计岗位之间必然存在方案的依赖和耦合关系;论证过程也是一个方案寻优的过程,每个岗位可能会设计多套方案,通过分析择优最终确定任务总体设计方案.

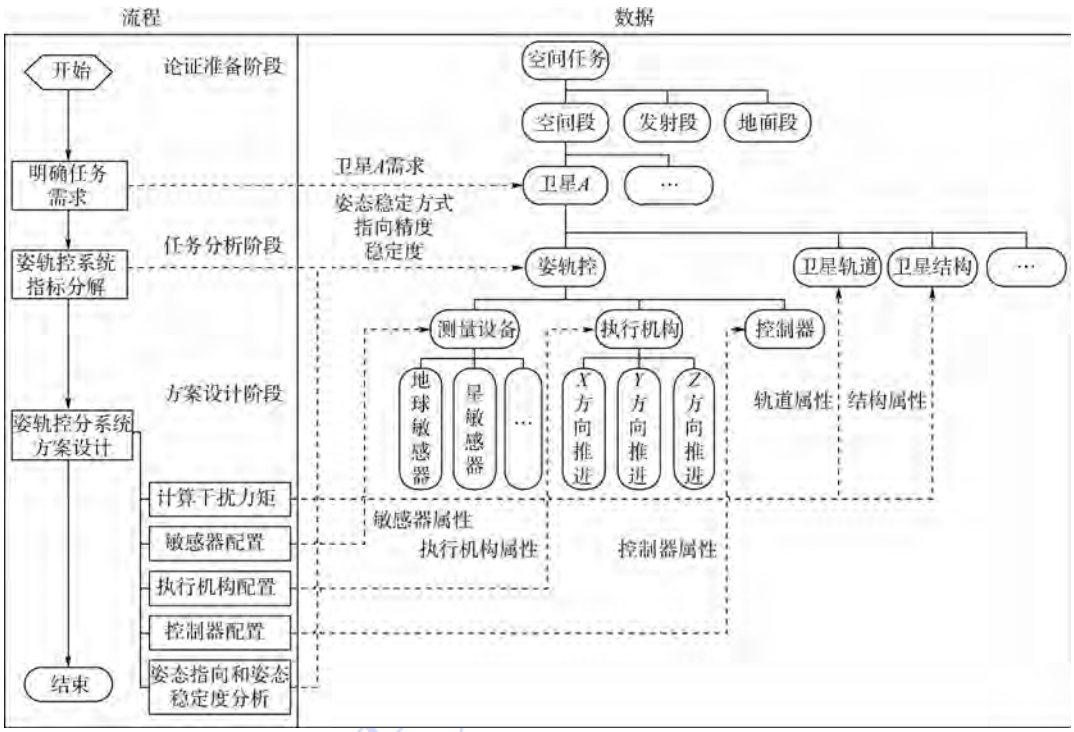


图 3　姿态岗设计流程与数据映射示意图

Fig. 3　Mapping between process and data of attitude post

多岗位多方案依赖关系如图 4 所示. 图 4 中岗位 2 有两套方案分别为 B 和 C,其中方案 B 依赖了岗位 1 的方案 A,同时方案 B 又被岗位 3 的方案 D 依赖;即方案 A 的数据是方案 B 的输入,方案 B 的数据又是方案 D 的输入. 本文提出用方案依赖关系矩阵方法求解任务总体设计方案.

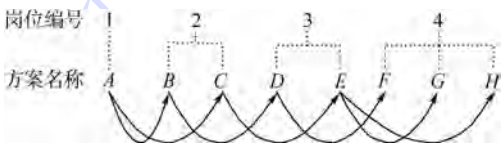


图 4　多岗位多方案依赖示意图

Fig. 4　Dependence relationship of multi-post and multi-solution

假设 1　某论证任务涉及 $m(m\geqslant 1)$ 个岗位,每个岗位依次有 $n_1,n_2,\cdots,n_m$ 套设计方案,设计方案编号依次为 $1,2,\cdots,N$.

$$N=\sum_{k=1}^{m} n_k \tag{1}$$

式中,k 为总体方案套数.

约束 1　某一岗位的某一套设计方案可依赖另一设计岗位的某一套设计方案,但仅可依赖该岗位的一套设计方案,即不能同时依赖同一岗位的多套方案.

定义 1　方案依赖关系矩阵 $\boldsymbol{P}=(p_{ij})_{N\times N}$,为 BOOL 矩阵. 其中:

$$p_{ij}=\begin{cases}1 & i\neq j,\text{方案 } i \text{ 依赖方案 } j\\ 0 & i\neq j,\text{方案 } i \text{ 不依赖方案 } j\\ 1 & i=j\end{cases}$$

定义 2　总体设计方案关系矩阵 $\boldsymbol{Q}=(q_{ij})_{k\times N}$,为 BOOL 矩阵. 其中:

$$q_{ij}=\begin{cases}1 & \text{总体方案 } i \text{ 包含方案 } j\\ 0 & \text{总体方案 } i \text{ 不包含方案 } j\end{cases}$$

通过设计方案间的耦合依赖关系,建立方案依赖关系矩阵 $\boldsymbol{P}$,求解总体设计方案关系矩阵 $\boldsymbol{Q}$,得到数据一致的总体设计方案. 求解算法如下:

```
//追踪方案 i 至方案 j 的依赖关系
for i from 1 to N
    for j from 1 to N and j≠i
        if P(i,j) = 1
            P(i,:) = max(P(i,:),P(j,:))
        end
    end
end
//将依赖关系赋予总体设计方案关系矩阵
for i from 1 to N
    Flag = 0
    for j from 1 to N and j≠i
        if max(P(j,:),P(i,:)) = P(j,:)
            Flag = 1
            break
        end
```

北航学报 赠阅

```
    end
    if Flag = 0
        k = k + 1
        Q(k,:) = P(i,:)
    end
end
```

4 平台实施与实例验证

空间科学任务协同设计平台作为一个分布式平台,采用 Eclipse RCP 和 Spring 技术架构,整合了 Hibernate、工作流 Activiti5 等中间件,提供统一门户,支持多岗位、多任务、多方案和多版本的管理能力,提供论证流程监控和数据协同交互等功能. 该平台是一个通用框架,适应多任务、多专业的总体论证需求. 本文以开展某项空间科学任务协同论证为例,对平台的有效性进行验证.

1) 启动任务论证.

论证开始,管理员配合项目负责人完成论证任务创建、团队创建、论证流程制定、数据部署等操作;配置完毕,启动该项任务论证工作. 管理员通过统一门户登录,图形化的流程制定界面和数据映射如图 5 和图 6 所示.

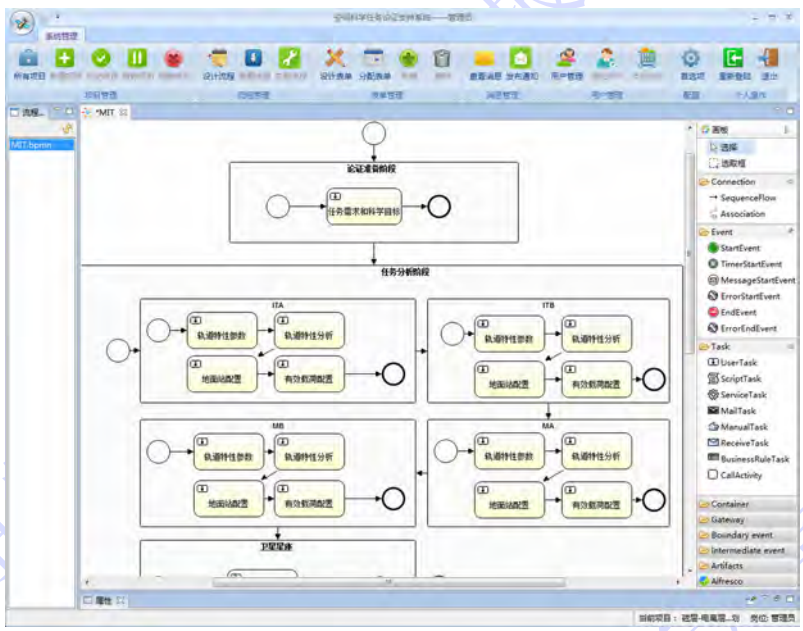


图 5 图形化的流程制定界面

Fig. 5 Interface of graphical model of process

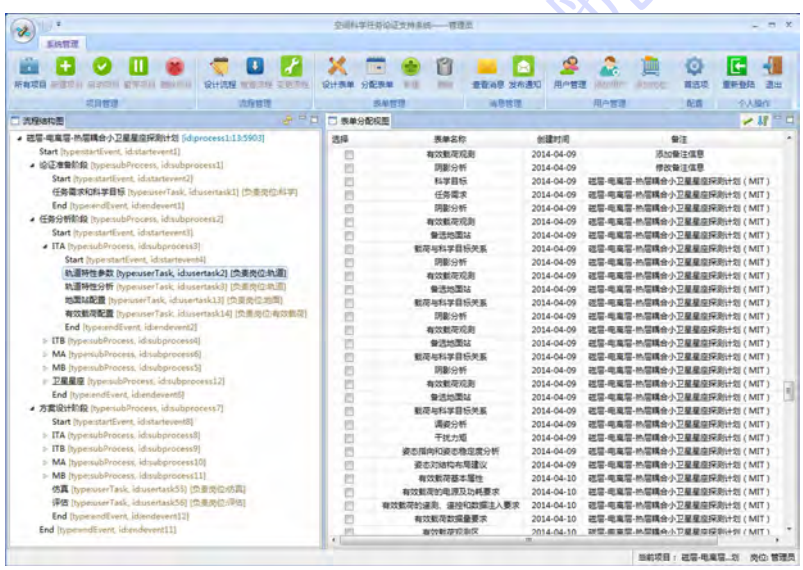


图 6 流程与数据映射界面

Fig. 6 Interface of process and data mapping

北航学报 赠阅

2）协同设计.

各岗位设计人员依据论证流程分配的设计活动开展工作.本项目涉及科学岗(需求提出方)、任务分析岗(项目负责人)、卫星平台岗、有效载荷岗、轨道岗、姿态岗、结构岗和系统管理岗(管理员)等十余个岗位.由科学岗提出科学需求,每个设计岗位创建新方案、保存数据、提交方案;任务分析岗审核提交的设计方案,做出通过、修改或拒绝等决策.各设计岗位并行协同开展设计工作,通过协同设计区查看其他设计岗位的协同数据,如图 7 所示.

3）总体方案形成.

项目论证过程中,轨道岗针对科学需求提出了两套设计方案.针对这两套设计方案,数传岗、地面岗、空间环境岗针对轨道方案各提出了相应的两套设计方案.其他岗位均提交了一套方案.

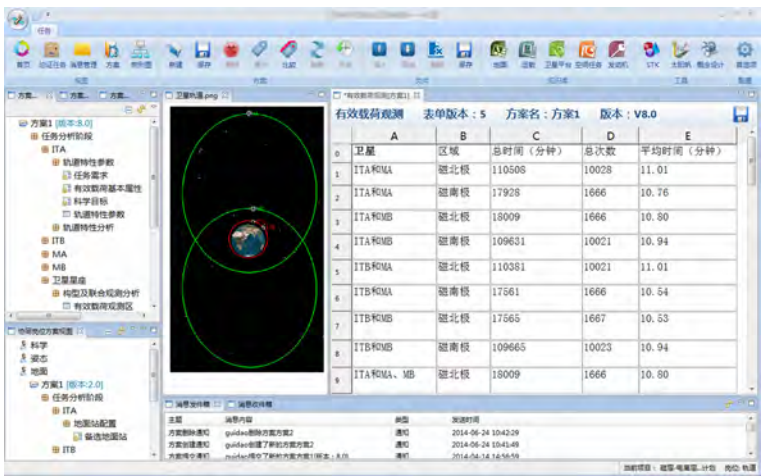


图 7　协同设计工作界面

Fig. 7　Interface of collaborative design

任务岗根据设计方案的依赖性进行方案汇总,形成两套总体设计方案,并开展总体设计方案的评估优化,分系统设计方案评估优化与总体方案评估优化相结合,对方案进行决策,确定最终的任务总体设计方案,并自动形成论证报告.

5　结　论

1）空间科学任务协同设计论证平台提供岗位、数据、流程的统一管理.该平台具有集成学科领域工具的开放接口,具有很好的兼容性和可扩展性.

2）以数据为核心,支持多任务、多岗位、多方案、多种数据类型、多版本的管理;图形化的流程设计,支持流程的监控、历史的追溯;协同交互机制,提供协同数据查询、协同数据关联、数据变更提醒等功能.

3）通过对空间科学任务论证岗位、数据、流程的统一管理,对后续形成论证的标准与规范和开展空间科学任务论证具有重要的意义.

后续应在流程建模的基础上进一步开展基于约束条件的流程规划研究;同时,应拓展数据为核心的管理模式,开展基于模型的论证平台研究.

致谢　衷心感谢中国科学院国家空间科学中心孟新、郑建华、杨震等在研究过程中给予的支持和帮助.

参考文献（References）

[1] 涂歆滢,杨雷. 国外航天器协同设计中心现状分析[J]. 航天器工程,2009,18(1):83-88.
Tu X Y, Yang L. Overview of satellite concurrent design center [J]. Spacecraft Engineering, 2009, 18(1): 83-88(in Chinese).

[2] Donald W M, Willian M P. Aerospace system design in NASA's collaborative engineering environment [J]. Acta Astronautica, 2000, 47(2-9): 255-264.

[3] The ESA Concurrent Design Facility. Concurrent engineering applied to space mission assessments[EB/OL]. [2014-05-10]. http://esa multimedia. esa. int/docs/cdf/CDF-INFOPACK-2011.pdf.

[4] Karpati G, Martin J, Steiner M, et al. The Integrated Mission Design Center (IMDC) at NASA goddard space flight center[C]// Proceedings of the 2003 IEEE Aerospace Conference. Montana: IEEE Publications, 2003: 3657-3667.

[5] Omid F V, Mahmoud H. A collaborative and integrated platform to support distributed manufacturing system using a service-oriented approach based on cloud computing paradigm[J]. Robotics and Computer-Integrated Manufacturing, 2013, 29(1): 110-127.

[6] 章翔峰. 支持多学科协同设计的航天产品仿真数据管理系统[J]. 航天制造技术,2013(6):58-61.
Zhang X F. Simulation data management system of aerospace

product support MDO[J]. Aerospace Manufacturing Technology, 2013(6):58-61(in Chinese).

[7] 李景新,颜丙新,邓小静. 型号计划与生产计划协同管理研究[J],机械设计与制造工程,2013,42(11):39-42.
Li J X,Yan B X,Deng X J. Research on the collaborative management of the product plan and model plan[J]. Machine Design and Manufacturing Engineering,2013,42(11):39-42(in Chinese).

[8] 刘炫,王嘉春,何清法. SOA 在航天器数据共享平台中的应用与研究[J]. 微计算机信息,2010,26(28):192-193.
Liu X,Wang J C,He Q F. Application and research of spacecraft data share platform based on SOA[J],Microcomputer Information,2010,26(28):192-193(in Chinese).

[9] 孙建勋,张立强,陈建江,等. 飞行器总体多学科集成设计平台[J]. 计算机集成制造系统,2012,18(1):1-8.
Sun J X,Zhang L Q,Chen J J,et al. Integrated design platform of conceptual multidisciplinary for aerocrafts[J]. Computer Integrated Manufacturing Systems,2012,18(1):1-8(in Chinese).

[10] 郭爱斌,米洁,董晓琴,等. 概念设计集成系统与星座方案探索[J],北京航空航天大学学报,2007,33(5):613-617.
Guo A B,Mi J,Dong X Q,et al. Concept design integrated system and constellation architecture exploration[J]. Journal of Beijing University of Aeronautics and Astronautics, 2007, 33(5):613-617(in Chinese).

[11] Wu X,Xu L,Jin L,et al. The MDO environment for hypersonic vehicle system design and optimization[C]//AIAA/ASME/SAE/ASEE 42nd Joint Propulsion Conference. Sacramento, CA:AIAA,2006:8327-8334.

[12] 王振华,邓家褆,隋鹏飞,等. 复杂工程系统综合设计优化环境[J]. 北京航空航天大学学报,2007,33(2):192-196.
Wang Z H,Deng J T,Sui P F,et al. Complex engineering system synthesis design optimization framework[J]. Journal of Beijing University of Aeronautics and Astronautics,2007,33(2):192-196(in Chinese).

[13] 孟新,杨震. 空间科学探测任务集同论证平台[J]. 科研信息化技术与应用,2011,2(3):55-65.
Meng X,Yang Z. The space science mission concurrent design platform[J]. E-Science Technology & Application, 2011, 2(3):55-65(in Chinese).

[14] 王书河,何麟书,张玉珠. 飞行器多学科设计优化软件系统[J]. 北京航空航天大学学报,2005,31(1):51-55.
Wang S H,He L S,Zhang Y Z. Flight vehicles multidisciplinary design optimization software system[J]. Journal of Beijing University of Aeronautics and Astronautics,2005,31(1):51-55(in Chinese).

[15] 姚雯,陈小前,赵勇. 基于不确定性 MDO 的卫星总体优化设计研究[J]. 宇航学报,2009,30(5):1808-1815.
Yao W,Chen X Q,Zhao Y. Research on satellite system design based on uncertainty multidisciplinary design optimization[J]. Journal of Astronautics, 2009, 30(5): 1808-1815(in Chinese).

[16] 孙亚东,张旭,宁汝新,等. 面向多学科协同开发领域的集成建模方法[J]. 计算机集成制造系统,2013,19(3):449-460.
Sun Y D,Zhang X,Ning R X,et al. Integrated modeling method oriented to multi-disciplinary collaborative development domain[J]. Computer Integrated Manufacturing Systems, 2013, 19(3):449-460(in Chinese).

Demonstration platform for collaborative design of space science missions

DENG Li*[1], HAN Chao[1], CAO Jinbin[1], SONG You[2]

(1. School of Astronautics, Beijing University of Aeronautics and Astronautics, Beijing 100191, China;
2. School of Software, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: There are several problems during the demonstration process of collaborative design of space science missions, such as strong coupling between sub-system solutions, poor data consistency, difficulty in data modification, and associated relationship between data and demonstration process. In order to solve these problems, a multi-layer digital model for typical space missions was proposed. The demonstration process was graphically modeled, and the mapping between the process and data was established. The shared data pool was provided as the data collaboration mechanism for all users and all sub-system solutions. The message bus was adopted as the timely reminder to notify the changes of data. The design solution coupling matrix was used to describe the coupling relationship between sub-system solutions and to gain the system solution for the missions by the automatically merged method. The demonstration platform, as a distributed platform for the collaborative design of space science missions, uses the technology architecture of Eclipse RCP and Spring and integrates the middlewares of Hibernate and Activiti5. The platform can provide a unified portal and support multi-user, multi-mission, multi-solution and multi-version data management, and also support process monitor and data interactive collaboration. A space mission was used as an example to verify the efficacy of the platform.

Key words: space science mission; collaborative design; data consistency; demonstration process; solution merging

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0301

基于脉冲激光定位的 SRAM 单粒子闩锁事件率预估

余永涛[1,2], 韩建伟*[1], 封国强[1], 蔡明辉[1]

(1. 中国科学院 空间科学与应用研究中心, 北京 100190; 2. 中国科学院大学, 北京 100049)

摘 要: 器件单粒子闩锁效应(SEL)预估方法一般是建立在只有一个敏感体积单元的长方体(RPP)模型上,静态随机存储器(SRAM)单粒子闩锁敏感区的定位试验结果表明敏感体积单元不仅有一个. 利用脉冲激光定位 SRAM K6R4016V1D 单粒子闩锁效应敏感区的试验结果对器件在轨 SEL 事件率进行了修正计算. 首先利用脉冲激光定位 SRAM SEL 敏感区,获得敏感区的分布情况,并确定整个器件敏感体积单元的数量. 然后针对不同的空间轨道、辐射粒子以及敏感体积厚度和敏感体积单元数进行了相应的器件 SEL 事件率计算,并对计算结果进行了分析讨论. 结果表明,重离子引起的 SEL 事件率随敏感体积单元数量的增大而减小;修正敏感体积单元数量对预估质子引起的 SEL 事件率非常必要,否则将会过高评估质子直接电离作用对 SEL 事件率的贡献.

关 键 词: SEL 事件率; 敏感区定位; 敏感体积单元; 质子直接电离; 静态随机存储器; 脉冲激光

中图分类号: V520.6; TN4

文献标识码: A **文章编号**: 1001-5965(2015)04-0609-07

当空间辐射环境中的高能带电粒子入射到航天器电子设备中的 CMOS 器件时,易诱发单粒子闩锁效应(SEL, Single Event Latchup),导致器件失效甚至航天器仪器故障. 单粒子闩锁效应一直以来都是国内外单粒子效应研究的热点,随着半导体工艺尺寸的减小,亚微米器件应用于航天设备后表现出对单粒子闩锁效应的极端敏感性[1-3]. 根据器件地面辐射模拟试验数据和航天器轨道的辐射环境模型,预估器件在轨单粒子效应,对正确选用宇航半导体器件、提高航天器寿命和可靠性具有重大的现实意义,对评价半导体器件的抗单粒子效应能力也具有重要的指导作用和参考价值.

多数单粒子效应预估程序都是建立在 RPP (Rectangular Parallelepiped) 敏感体积模型基础上,模拟入射粒子在敏感体积内的能量沉积. 对于器件单粒子闩锁在轨预估,单粒子闩锁效应没有类似于存储类器件单粒子翻转效应存储位单元的概念,因而在预估计算时多是假设整个器件只存在一个敏感体积单元(SV, Sensitive Volume). 近年来,一系列脉冲激光定位静态随机存储器(SRAM)单粒子闩锁效应敏感区的试验结果[4-7]表明,SRAM 器件的 SEL 敏感区是由许多具有周期性的较小尺寸的 SEL 敏感区构成的,即 SRAM

收稿日期: 2014-05-27; **录用日期**: 2014-09-26; **网络出版时间**: 2014-12-09 15:31
网络出版地址: www.cnki.net/kcms/detail/11.2625.V.20141209.1531.003.html
基金项目: 中国科学院支撑技术资助项目(110161501038)
作者简介: 余永涛(1987—),男,河南宝丰人,博士生,yuyongtao10@mails.ucas.ac.cn
* **通讯作者**: 韩建伟(1970—),男,河南长葛人,研究员,hanjw@nssc.ac.cn,主要研究方向为航天元器件和材料的空间环境效应研究与应用.

引用格式: *余永涛, 韩建伟, 封国强, 等. 基于脉冲激光定位的 SRAM 单粒子闩锁事件率预估[J]. 北京航空航天大学学报, 2015, 41(4): 609-615. Yu Y T, Han J W, Feng G Q, et al. SEL rate prediction for SRAM using pulsed laser sensitivity mapping[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(4): 609-615 (in Chinese).*

SEL 敏感体积单元的数量不是只有一个,这将会对 SRAM 器件空间 SEL 事件率预估产生影响. 如文献[4]和文献[7]的论述,在只有一个敏感体积单元的假设下,在进行 SEL 事件率预估时,根据试验测得的截面数据建立的 RPP 模型的横向尺寸很大,离子倾斜入射时电荷收集路径很长,线性能量传输(LET,Linear Energy Transfer)值低而射程长的离子倾斜入射到器件敏感区时也可沉积足够的电荷而诱发 SEL. 在各向同性辐射环境中,粒子倾斜入射的可能性很大,入射倾角大于等于60°的粒子比例为50%. 实际情形是,SRAM SEL 敏感体积的横向尺寸是有限的,离子倾斜入射时电荷收集路径也是有限的. 与真实情况相比,假设整个器件只存在一个敏感体积单元的传统预估方法最终会得到一个相对保守的事件率预估结果. 对于修正敏感体积单元数量以改进 SEL 事件率预估,以往的工作多是进行定性讨论,没有系统地论述如何修正以及敏感体积单元数量的修正对 SEL 事件率预估有多大程度的影响.

本文工作主要分为 3 部分,一是利用脉冲激光定位 SRAM K6R4016V1D SEL 敏感区,获得不同尺度下敏感区的分布情况,进而确定器件敏感体积单元的数量;二是针对地球同步轨道和极轨道 2 种典型空间轨道辐射环境,利用 OMERE 软件计算不同条件下器件在轨 SEL 事件率;最后对比讨论不同情况下敏感体积单元数对 SEL 事件率预估结果的影响.

1 激光定位试验

1.1 脉冲激光定位 SEL 敏感区试验系统及方法

脉冲激光单粒子效应(SEE)敏感区定位成像系统示意图如图 1 所示. 主要有 4 部分构成:脉冲激光辐照装置、单粒子效应检测系统、电动移动台、同步控制装置. 利用聚焦后的脉冲激光对被测器件进行逐点辐照,同时利用单粒子效应检测系统实时检测被测器件的单粒子效应,利用移动台实现被测器件不同区域的扫描测试,同步控制装置将脉冲激光辐照、单粒子效应检测、电动移动台移动同步控制起来,并将每个辐照位置的激光脉冲注量、单粒子效应数据、物理坐标对应起来传输到计算机,最终将单粒子效应敏感位置显示为二维图像.

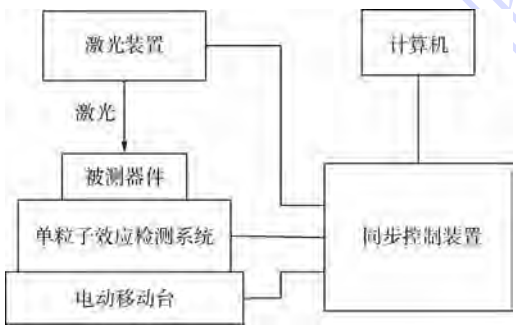


图 1 脉冲激光 SEE 敏感区定位成像系统示意图

Fig. 1 Schematic of pulsed laser facility for SEE sensitivity mapping

SEL 敏感区定位测试流程图如图 2 所示. 首先是初始化被测器件和设定试验测试条件,包括使被测器件正常工作、使激光聚焦定位在测试原点及调节脉冲激光能量等. 然后开始脉冲激光辐照,同时利用单粒子效应检测系统实时监测被测器件的供电电流,一旦电流值超过设定的限值,对被测器件断电. 脉冲激光能量和 SEL 的电流值以及辐照位置对应的物理坐标会被记录下来,这个坐标就是发生 SEL 的物理位置. 断电延时一段时间后,检测系统恢复被测器件供电,接着移动台移动一定的距离到下一个测试位置,重复以上过程,直到完成对所选区域的扫描测试,最终可以获得器件测试区域内的 SEL 敏感位置分布图.

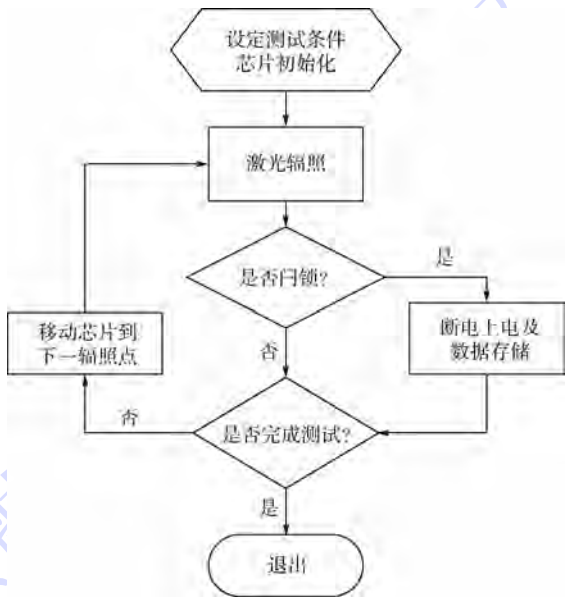


图 2 SEL 敏感区定位测试流程图

Fig. 2 State diagram of SEL sensitivity mapping

1.2 脉冲激光定位 SEL 敏感区试验及结果

试验样品为 SAMSUNG SRAM K6R4016V1D,容量为 256 K × 16 bit,工作电压为(3.3 ± 0.3)V,芯片大小为 0.46 × 0.52 cm^2. 这款器件在重离子试验中表现出对 SEL 非常敏感[8]. 为避开器件正面金属层对激光的阻挡作用,激光辐照试验采用背部辐照方法. 根据不同的测试尺度,移动台的步距分别设定为 10,5,1 μm. 测试过程中,激光光斑直径约为 1.7 μm[9],每个测试位置辐照 1 个激光

北航学报 赠阅

脉冲,各点的脉冲激光能量相同.

不同尺度及扫描步距下 SEL 敏感区位图如图 3 所示. 图中浅色带状区域是 SEL 敏感位置,深色背景是器件的显微图片. 如图 3(a)所示,首先对整个器件的上半部分进行测试,占器件总面积的 1/2,激光辐照扫描的步距是 10 μm,可以看出 SEL 敏感区位于 SRAM 器件存储阵列部分,而周围的电路部分对 SEL 不敏感. 为了更好地表现 SEL 敏感区分布情形,对其中局部区域进行放大,如图 3(b)所示,敏感区表现出明显的带状分布图形. 增加激光辐照扫描定位精度,移动步距减小为 5 μm. 如图 3(c)所示,对其中大小为 200 μm × 200 μm的区域进行测试,结果显示 SEL 敏感区仍具有周期性的重复结构,在该测试区域内约有 4 ×9个周期结构. 最后,进一步减小激光辐照扫描的步距为 1 μm,使用能量稍高于 SEL 阈值能量的脉冲激光对上述周期结构中的一个进行 SEL 敏感区测试. 如图 3(d)所示,在 20 μm ×30 μm 的区域内,SEL 敏感区表现出周期性图形,由 5 个单一 SEL 敏感区组成.

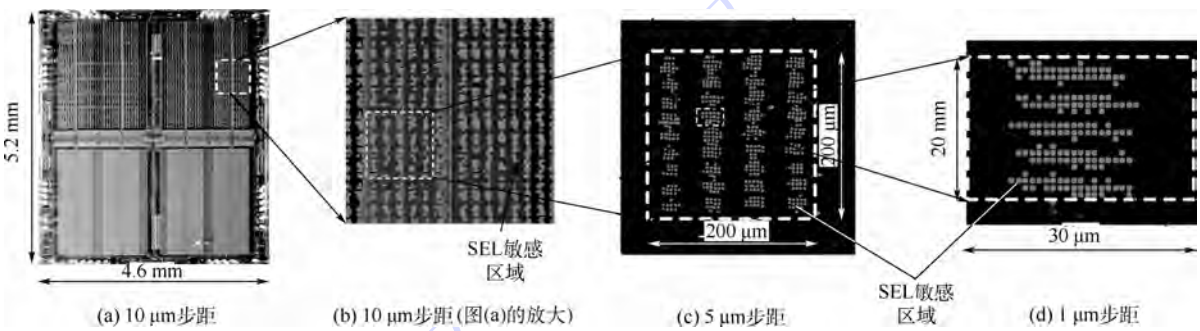


图 3　不同尺度及扫描步距下 SEL 敏感区位图

Fig. 3　SEL sensitivity mappings at different scales with different scanning steps

从图 3(a)中可以看出,整个器件的存储阵列部分对 SEL 的敏感性没有显著差异,而且器件 SEL 的电流均在 250 ~ 300 mA 范围内,因而认为器件所包含的众多 SEL 敏感区对辐照的敏感性是一致的. 根据以上的试验结果和测试器件的相关信息,可计算整个器件包含单一 SEL 敏感区的数量. 在器件的上半部分,存储阵列区域在竖直方向上的尺寸为 2 200 μm,而敏感区在横向上共有 64 条周期结构,而在竖直方向上 200 μm 范围内共有9 个周期结构,而其中一个周期结构包含5 个单一 SEL 敏感区. 则在不同的扫描步距的情况下,整个器件包含 SEL 敏感体积的数量 N 估算如下:

$$10\ \mu m: N = 2 \times 64 = 128$$

$$5\ \mu m: N = 2 \times 64 \times 2\,200 \times 9/200 = 12\,672$$

$$1\ \mu m: N = 2 \times 64 \times 2\,200 \times 9 \times 5/200 = 63\,360$$

2　SEL 事件率修正计算

通过以上试验获得了不同定位尺度下 SRAM K6R4016V1D 单粒子闩锁效应敏感区的二维周期分布图. SEL 敏感区的周期性分布对器件 SEL 在轨事件率预估具有重要的影响. 在预估器件 SEL 事件率时,一般利用试验测得的闩锁阈值和截面数据,根据 RPP 模型进行计算. 经典的预估方法是假设整个器件只有一个敏感体积单元,收集沉积在敏感体积内的电荷,沉积电荷量与离子穿过敏感体积的路径相关. 根据试验测得的截面数据建立的经典 RPP 模型的横向尺寸很大,离子倾斜入射时电荷收集路径很长,因而低 LET 值而长射程的离子倾斜入射到器件敏感区时也可沉积足够的电荷而诱发 SEL. 而激光定位的试验结果表明,SRAM SEL 敏感区具有周期重复性,在相邻的敏感区之间存在不敏感的区域. 离子入射到任意单一 SEL 敏感区都可使整个器件发生 SEL,而单一 SEL 敏感区的横向尺寸远小于经典模型的横向尺寸. 这样传统的 SEL 事件率预估方法将会高估 SEL 截面,最终给出比真实情况更保守的结果,这可能会对正确评估和选用宇航器件产生重大影响.

根据以上试验结果及分析,类似于存储器件"存储单元"的概念,认为 SRAM 器件 SEL 敏感区由 N 个敏感体积单元组成. 为了表现 SEL 事件率随敏感体积单元数量的变化趋势,在以下部分,将通过计算,对比在不同的敏感体积单元数情况下器件在轨 SEL 事件率. 器件在轨 SEL 事件率计算的基础是器件地面模拟试验截面数据、器件所在轨道的空间环境辐射模型和高能粒子与器件相互作用模型. K6R4016V1D SRAM 重离子测试结果及韦伯拟合参数如表 1 所示[8].

表 1 K6R4016V1D SRAM 重离子测试结果及韦伯拟合参数[8]

Table 1 K6R4016V1D SRAM heavy ion test result and Weibull parameters[8]

参数	LET 阈值/(MeV·cm²·mg⁻¹)	饱和截面/cm²	尺度参数 W/(MeV·cm²·mg⁻¹)	形状参数 S
数值	1.0	0.06	9	3.5

所有计算都是采用 OMERE V3.6[10] 软件. 为了研究不同轨道辐射环境下敏感体积单元数对 SEL 事件率的影响,选取地球同步轨道(GEO)(35 870 km, 0°)和极轨低地球轨道(LEO)(800 km,98°)2 种典型空间轨道,重离子辐射环境模型为银河宇宙线,质子辐射环境模型为俘获带质子模型 AP8min,采用 3.0 mm Al 等效屏蔽,不考虑太阳活动异常. 器件在轨 SEE 预估过程中,一般根据器件工艺确定 RPP 模型的敏感体积厚度,对于 CMOS/epi 工艺一般取 2 μm[11]. 但对于单粒子闩锁效应这可能不再适用,因而在计算敏感体积厚度时还取 3 μm[5]. 对应不同的敏感体积单元数量,SEL 敏感体积横向尺寸如表 2 所示.

表 2 SEL 敏感体积横向尺寸

Table 2 Lateral dimensions of SEL sensitive volume

参数	SV 数量			
	1	128	12 672	63 360
横向尺寸/μm²	2 450×2 450	271.6×217.6	21.8×21.8	9.7×9.7

2.1 重离子 SEL 事件率

对应不同的敏感体积单元数和敏感体积厚度,利用 OMERE 计算的地球 GEO 轨道重离子 SEL 事件率如图 4 所示.

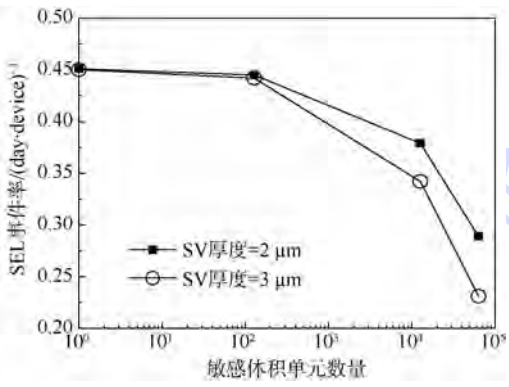


图 4 地球 GEO 轨道重离子 SEL 事件率

Fig. 4 Heavy ion SEL predicted rates of GEO

从图 4 中可以看出,随着敏感体积单元数量的增大,SEL 事件率逐渐减小,SEL 事件率对敏感体积厚度变得敏感. 在敏感体积厚度为 2 μm 而敏感体积单元数量由 1 增加到 63 360 时,SEL 事件率由 0.45 (day·device)⁻¹ 减小到 0.28 (day·device)⁻¹. 另外,当敏感体积单元数较小时,SEL 事件率与敏感体积厚度基本无关;当敏感单元数增大到 12 672 和 63 360 时,敏感体积厚度增大,SEL 事件率明显减小.

图 5 为极轨道重离子 SEL 事件率. 可以看出,极轨道辐射环境下的计算结果与同步轨道的计算结果表现出同样的趋势. 在敏感体积厚度为 2 μm 而敏感体积单元数量增加到 63 360 时,SEL 事件率由 0.14 (day·device)⁻¹ 减小到 0.08 (day·device)⁻¹.

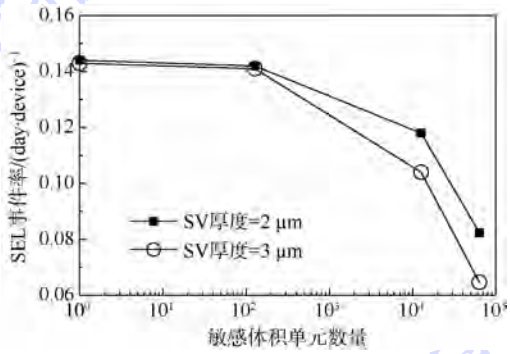


图 5 极轨道重离子 SEL 事件率

Fig. 5 Heavy ion SEL predicted rates of polar LEO

2.2 质子 SEL 事件率

对于空间辐射环境中质子诱发的 SEL,主要针对 LEO 极轨道考虑俘获带质子诱发的 SEL,质子环境模型采用 AP8min,使用 SIMPA 等效模型计算质子 SEL 截面. SIMPA 是一个基于重离子数据计算质子 SEE 的半经验模型,计算质子 SEE 截面的公式为[12]

$$\sigma_{\mathrm{p}}(E_{\mathrm{p}})=\int_{E_{\mathrm{s}}}\sigma_{\mathrm{II}}(E)\Phi(E,E_{\mathrm{p}})\,\mathrm{d}E$$

式中,$\sigma_{\mathrm{p}}(E_{\mathrm{p}})$ 为质子能量为 E_{p} 时的 SEE 截面;$\sigma_{\mathrm{II}}(E)$ 为重离子 SEE 的截面;E_{s} 为重离子试验的单粒子效应阈值能量;$\Phi(E,E_{\mathrm{p}})$ 是入射能量为 E_{p} 的质子在敏感体积内沉积能量为 E 的概率. 公式中各项参数是根据质子与 Si 核反应的 Monte Carlo 模拟以及质子沉积能量测量结果确定的,极轨道质子 SEL 事件率(OMERE SIMPA)如图 6 所示.

从图 6 中可以看出,对应同样的敏感体积厚度,敏感体积单元数对 SEL 事件率计算没有影响. 敏感体积厚度增加,SEL 事件率略有减小.

利用 SRIM 计算了 2 MeV 和 10 MeV 质子 LET 值随粒子在 Si 中穿透深度的变化. 在 Bragg 峰附近质子的 LET 值可达到 0.5 MeV·cm²·mg⁻¹, 10 MeV 质子在 Si 中射程可达 710 μm. 质子直接电

北航学报 赠阅

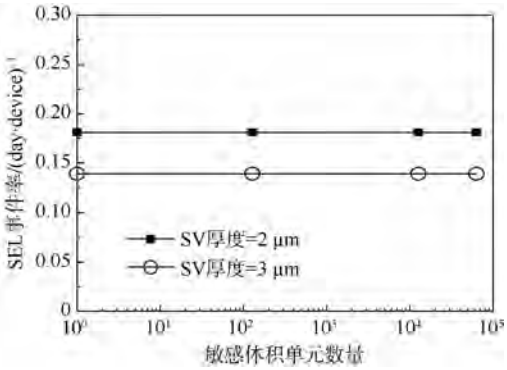


图 6　极轨道质子 SEL 事件率(OMERE SIMPA)

Fig. 6　Proton SEL predicted rates of polar LEO(OMERE SIMPA)

离作用对 SEE 的贡献通常是忽略不计的,但对于 SEE 阈值较低的器件则需要考虑[13-14]. 这款器件 SEL 阈值低、截面大,俘获带质子的能量范围为 0.1 ~400 MeV,高能质子倾斜入射时有效 LET 值可以远高于测试器件的 LET 阈值,质子的直接电离作用也可能会对 SEL 产生贡献.

图 7 是利用 OMERE 得到的极轨道俘获带质子 LET 谱. 利用质子的 LET 谱和重离子截面测试数据进行了 SEL 事件率计算,最终结果如图 8 所示.

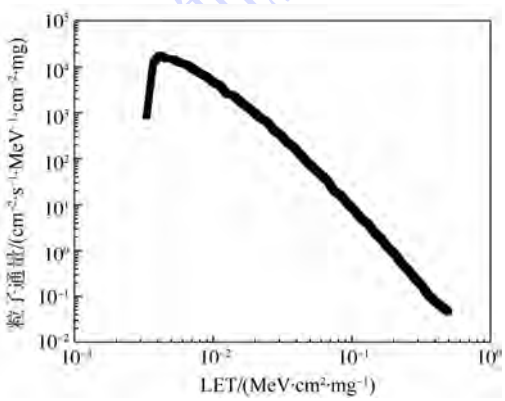


图 7　极轨道俘获带质子 LET 谱

Fig. 7　Trapped proton LET spectrum of polar LEO

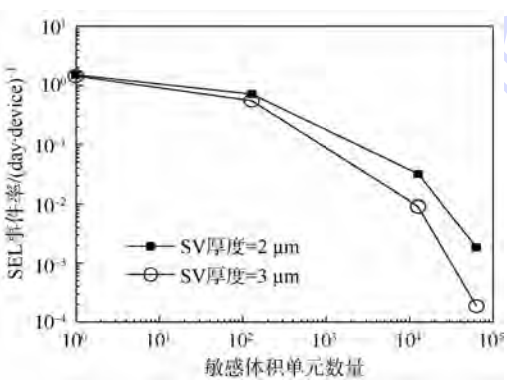


图 8　极轨道质子直接电离 SEL 事件率

Fig. 8　Proton SEL predicted rates of polar LEO by direct ionization

从图 8 中可以看出,计算结果与重离子 SEL 事件率计算结果相似,但敏感体积单元数对 SEL 事件率有很大影响. 在敏感体积厚度为 2 μm 时,敏感体积数由 1 增加到 63 360,SEL 事件率减小了 3 个数量级. 在敏感体积厚度为 3 μm 时,敏感体积数由 1 增加到 63 360,SEL 事件率减小了 4 个数量级.

3　敏感体积单元数量对 SEL 事件率的影响

3.1　重离子 SEL 事件率

2.1 节针对地球同步轨道和极轨道利用 OMERE 计算分析了敏感体积单元数对重离子 SEL 事件率的影响,2 种轨道的计算结果所表现出的趋势是一致的.

对应同样的敏感体积厚度,随着敏感体积单元数增加,SEL 事件率逐渐减小. 当敏感体积单元数从 1 增加到 128 时,SEL 事件率变化不大. 当敏感体积单元数增大到 12 672 和 63 360 时,SEL 事件率明显减小. 敏感体积厚度为 3 μm 时,对应地球同步轨道和极轨道 SEL 事件率分别减小了 48.7% 和 55.1%. 对于其他情形,敏感体积单元数修正引起的改变更小. 器件在轨 SEL 预估涉及众多因素,相对于地面模拟试验数据、飞行试验数据、空间辐射环境模型和粒子与器件相互作用模型等因素带来的误差[15],敏感体积单元数的影响不大.

另外,当敏感体积单元数比较小时 SEL 事件率与敏感体积厚度基本无关. 特别是按照传统的 SEL 预估方法,敏感体积单元数量为 1,对应 2 种敏感体积厚度的情况,SEL 事件率计算结果基本相等. 当敏感体积单元数增大到 12 672 和 63 360 时,敏感体积厚度增大,SEL 事件率明显减小. 敏感体积厚度的增大相当于增大了临界电荷量,因而随着敏感体积厚度的增大,SEL 事件率减小. 当敏感体积单元数为 1,根据试验的截面数据建立的敏感体积单元的横向尺寸远大于纵向尺寸. 空间辐射环境中的多数重离子具有低 LET 值而长射程的特点. 这些倾斜入射的重离子在敏感体积内沉积的电荷量要高于临界电荷量,临界电荷量的小幅度增加对 SEL 事件率基本没有影响,因而敏感体积厚度改变对传统 SEL 事件率预估几乎没有影响.

3.2　质子 SEL 事件率

2.2 节针对极轨道分别利用 OMERE SIMPA 模型和俘获带质子的 LET 谱计算分析了俘获带质子引起的 SEL 事件率. 利用 OMERE SIMPA 计

算,在相同的敏感体积厚度下,敏感体积单元数对 SEL 事件率计算没有影响. 根据 SIMPA 模型的计算方法,质子引起单粒子效应是通过核反应产生二次离子导致,不需要考虑入射质子空间方向上的差异,而敏感体积单元数量的修正影响的就是倾斜入射的粒子,因而质子 SEL 事件率与敏感单元数无关. 另外即使能够获得完整的质子单粒子效应截面地面辐照试验数据,根据 Bendel 双参数公式拟合截面进而预估质子单粒子效应,敏感体积单元数对结果也没有影响.

考虑到这款器件 SEL 阈值低、截面大,利用俘获带质子 LET 谱和重离子测试数据进行了 SEL 事件率计算. 敏感体积单元数对 SEL 事件率有很大影响. 在敏感体积厚度为 2 μm 时,敏感体积数由 1 增加到 63 360,SEL 事件率减小了 3 个量级. 这与俘获带质子有密切关系,根据计算的质子 LET 谱和对应能谱,能够通过直接电离作用引起 SEL 的是具有低 LET 值、高能量、长射程的质子. 对这款器件的计算表明,能量至少大于 10 MeV 的质子倾斜入射才可能通过直接电离诱发被测器件发生 SEL. 经过敏感体积单元数修正后,敏感体积单元的横向尺寸大幅减小,这些质子沉积的电荷不足以引起 SEL.

对比图 5、图 6 与图 8,在敏感体积厚度为 2 μm的情况下,敏感体积数为 1 时质子直接电离引起的 SEL 事件率比质子核反应引起的 SEL 事件率及宇宙线重离子引起的 SEL 事件率高一个量级,在敏感体积数为 63 360 时,质子直接电离引起的 SEL 事件率可以忽略. 这说明敏感体积单元数的修正对 SEL 事件率计算的影响很大,否则将会严重高估质子直接电离作用对 SEL 事件率的贡献[13]. 另外对于质子直接电离引起的 SEL 事件率,敏感体积单元数修正后,SEL 事件率对敏感体积厚度非常敏感,敏感体积厚度从 2 μm 增加到 3 μm,SEL 事件率减小了 1 个量级. 总之无论是重离子 SEL 事件率预估还是质子的 SEL 事件率估,敏感体积单元数修正后 SEL 事件率对敏感体积厚度变得相对敏感.

最后需要说明的是,以上计算分析了在不同的轨道、粒子、敏感体积厚度等条件下,敏感体积单元数对 SRAM K6R4016V1D SEL 事件率预估的影响. SEL 事件率预估结果还与器件本身的抗辐射参数有关. 质子单粒子效应预估有许多种不同的计算方法[16],本文采用的是基于重离子测试数据计算质子 SEL 事件率的 SIMPA 模型,对于其他方法计算结果的影响有待于进一步研究. 另外激光定位 SEL 敏感区的试验结果表明,单一 SEL 敏感区的长宽比很高,而本文主要关注的问题是敏感体积单元数量,在计算中设定敏感体积的长宽相等,这也是后续需要改进的部分.

4 结 论

利用脉冲激光定位 SRAM K6R4016V1D 单粒子闩锁效应敏感区的试验结果对器件在轨 SEL 事件率进行了改进计算. 结论如下:

1) 脉冲激光具有快速准确定位 SEL 敏感区的优势,可以在一定精度下确定器件的 SEL 敏感单元数量. 测试器件的 SEL 敏感区具有周期重复性,是由 63 360 个单一 SEL 敏感体积单元组成.

2) 对于重离子引起的 SEL 事件率,当敏感体积单元数较小时,SEL 事件率基本不随敏感体积单元数而变化;当敏感体积单元数增大到 63 360 时,SEL 事件率的减小较明显,SEL 事件率减小的最大幅度约为 50%.

3) 敏感体积单元数对质子核反应引起的 SEL 事件率没有影响,这与采用的计算模型有关;但敏感体积单元数量的修正对计算质子直接电离引起的 SEL 事件率非常必要,否则将会过高评估质子直接电离作用对总的 SEL 事件率的贡献.

参考文献 (References)

[1] Page Jr T E, Benedetto J M. Extreme latchup susceptibility in modern commercial-off-the-shelf(COTS) monolithic 1 M and 4 M CMOS static random-access memory(SRAM) devices[C]//Radiation Effects Data Workshop. Seattle: IEEE, 2005: 1-7.

[2] 韩建伟,张振龙,封国强,等. 单粒子锁定极端敏感器件的试验及对我国航天安全的警示[J]. 航天器环境工程,2008,25(3):265-267.
Han J W, Zhang Z L, Feng G Q, et al. The radiation test of SRAM devices for extreme single event latch-up susceptibility and a warning to our aerospace safety[J]. Spacecraft Environment Engineering, 2008, 25(3): 265-267(in Chinese).

[3] 余永涛,封国强,陈睿,等. SRAM K6R4016V1D 单粒子闩锁及防护试验研究[J]. 原子能科学技术,2012,46(增刊):587-591.
Yu Y T, Feng G Q, Chen R, et al. Experimental study on single event latchup of SRAM K6R4016V1D and its protection technology[J]. Atomic Energy Science and Technology, 2012, 46(Suppl.): 587-591(in Chinese).

[4] Burnell A J, Chugg A M, Harboe-Sørensen R. Laser SEL sensitivity mapping of SRAM cells[J]. IEEE Transactions on Nuclear Science, 2010, 57(4): 1973-1977.

[5] Faraud E, Pouget V, Shao K, et al. Investigation on the SEL sensitive depth of an SRAM using linear and two-photon absorption laser testing[J]. IEEE Transactions on Nuclear Science, 2011,

北航学报 赠阅

58(6):2637-2643.

[6] Dodds N A, Hooten N C, Reed R A, et al. SEL-sensitive area-mapping and the effects of reflection and diffraction from metal lines on laser SEE testing[J]. IEEE Transactions on Nuclear Science, 2013, 60(4):2550-2558.

[7] 余永涛,封国强,陈睿,等. CMOS SRAM 器件单粒子锁定敏感区的脉冲激光定位试验研究[J]. 航天器环境工程,2014,31(2):150-153.
Yu Y T, Feng G Q, Chen R, et al. Research of SEL sensitive region of CMOS SRAM by pulsed laser mapping facility[J]. Spacecraft Environment Engineering, 2014, 31(2):150-153(in Chinese).

[8] Ferlet-Cavrois V, Muschitiello M, Alessio D M. Single event latch-up(SEL) analysis of the 256 k × 16 SRAM samsung K6R4016V1D-TC10, RA0660[R]. Noordwijk: ESA/ESTECT-EC-QEC, 2013.

[9] Yu Y T, Feng G Q, Chen R, et al. Laser SEU sensitivity mapping of deep submicron CMOS SRAM[J]. Journal of Semiconductors, 2014, 35(6):064011-1-064011-4.

[10] OMERE softwareversion 3.6[CP/OL]. http://www.trad.fr/OMERE-Software.html.

[11] Petersen E L, Pickel J C, Adams J H, et al. Rate prediction for single event effects-a critique[J]. IEEE Transactions on Nuclear Science, 1992, 39(6):1577-1599.

[12] Douct B, Patin Y, Buisson J, et al. Characterization of proton interactions in electronic components[J]. IEEE Transactions on Nuclear Science, 1994, 41(3):593-600.

[13] Bezerra F, Ecoffet R, Lorfèvre E, et al. CARMEN2/MEX: an in-flight laboratory for the observation of radiation effects on electronic devices[C]//12th European Conference on Radiation and its Effects on Components and Systems(RADECS). Piscataway, NJ: IEEE, 2011:607-614.

[14] Artola L, Velazco R, Hubert G, et al. In flight SEU/MCU sensitivity of commercial nanometric SRAMs: operational estimations[J]. IEEE Transactions on Nuclear Science, 2011, 58(6):2644-2651.

[15] Bezerra F, Lorfevre E, Ecoffet R, et al. In flight observation of proton induced destructive single event phenomena[C]//European Conference on Radiation and its Effects on Components and Systems(RADECS). Piscataway, NJ: IEEE, 2009: 126-132.

[16] Petersen E L. Approaches to proton single-event rate calculations[J]. IEEE Transactions on Nuclear Science, 1996, 43(2):496-504.

SEL rate prediction for SRAM using pulsed laser sensitivity mapping

YU Yongtao[1,2], HAN Jianwei*[1], FENG Guoqiang[1], CAI Minghui[1]

(1. National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The classical approaches for single event latchup (SEL) rate prediction are based on the rectangular parallelepiped (RPP) model of only one sensitive volume. However, the experiment results of SEL sensitivity mapping of static random access memory (SRAM) show that the device has not only one sensitive volume (SV). The in-flight SEL rate of the device was corrected using the experiment results of pulsed laser SEL sensitivity mapping of SRAM K6R4016V1D. The SEL sensitivity maps of the SRAM by pulsed laser were first obtained and then the SV number of the device was calculated. The SEL rates of the device were predicted and discussed for different space orbits, radiation particles, SV thicknesses and SV number in particular. The results show that SEL rate caused by heavy ion decreases with SV number. The correction of the SV number is essential for SEL rate due to proton direct ionization; otherwise, the contribution of direct ionization of protons to SEL rate would be greatly overestimated.

Key words: SEL rate; sensitivity mapping; sensitive volume; direct ionization of proton; SRAM; pulsed laser

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0535

基于大涡模拟的平流层浮空器气动特性分析

王裕夫，陶国权*，刘东旭，武哲

（北京航空航天大学 航空科学与工程学院，北京 100191）

摘 要：为研究不同球体间距对“球形囊体型”平流层浮空器（SLV）气动特性的影响，采用有限体格式、结构网格和大涡模拟（LES）计算方法求解不可压缩的 Navier-Stokes（N-S）方程，对超临界雷诺数“球形囊体型”平流层浮空器绕流进行数值模拟，并对不同球体间距下的数值计算结果进行详细的分析比较. 通过对比试验数据，单球体数值模拟的阻力系数时均值与 Achenbach 的试验数据一致，验证了计算方法分析超临界雷诺数球体绕流问题的准确性. 研究不同间距的双球体阻力变化规律以及振动频谱特性，随上下游球体间距 G 的增加，合阻力先增大后减小，上游球体的阻力占优振动频率逐渐减小；$G=1.5D$（D 为球体直径）和 $G=2D$ 时，上下游球体有相同的占优振动频率. 随间距 G 的增加，两球体相互作用与上游球体对下游球体的尾涡结构影响逐渐减弱.

关 键 词：平流层浮空器；大涡模拟；球体绕流；阻力系数；频谱特征
中图分类号：O357.1
文献标识码：A **文章编号**：1001-5965(2015)04-0616-08

近年来，平流层浮空器（SLV，Stratosphere Lighter-than-air Vehicle）受到很大的重视，世界各主要国家均开展了相关的研究与试验. 平流层浮空器具有定点驻留、超长航时、视野范围广、生存能力强的特点，具有重要的民用和军用价值[1]. 平流层浮空器主要有“艇型”和“球形囊体型”两大类. “艇型”平流层浮空器具有气动阻力小的优点，但受限于目前的超压蒙皮材料技术，未来一段时间内“艇型”平流层浮空器在解决平流层的昼夜温差问题还有一定难度[2]. “球形囊体型”平流层浮空器是近期内在平流层长时驻空一种有效的途径.

“球形囊体型”相比“艇型”平流层浮空器存在阻力大、气动力波动幅度大等不足. 因此，研究“球形囊体型”平流层浮空器的气动特性并且通过合理设计减阻对其总体设计具有重要意义. 对于球体绕流问题，当雷诺数增加到某一值，球体表面边界层由层流变成湍流，推迟了边界层分离，压差阻力减少，阻力系数急剧下降，这种现象称为“失阻”，对应的雷诺数称为临界雷诺数. 由于“球形囊体型”平流层浮空器尺度大，对应的雷诺数就会很大，气动模型简化为超临界雷诺数范围的多球体绕流问题. 在超临界雷诺数范围时，球体尾流中的流动为湍流状态，流场具有复杂结构，呈明显的非定常性.

国外，Sakamoto 和 Haniu 系统地总结了前人的研究成果，并且通过流场显示试验和热线试验，测量了球体在很广的雷诺数范围内的斯特劳哈尔数 St[3]. Constantinescu 和 Squires 在总结前人研究成果的基础上，给出了球体的阻力系数随雷诺

收稿日期：2014-08-30；**录用日期**：2014-10-10；**网络出版时间**：2014-11-25 09:13
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0535.html
基金项目：北航蓝天新秀基金资助项目（YWF-13-T-RSC-071）
作者简介：王裕夫（1988—），男，四川成都人，博士生，justin-psp@163.com
*通讯作者：陶国权（1983—），男，浙江绍兴人. 讲师，ac97531@163.com，主要研究方向为飞行器总体设计，复合材料结构设计，无人机研制，螺旋桨优化设计.

引用格式：王裕夫，陶国权，刘东旭，等. 基于大涡模拟的平流层浮空器气动特性分析[J]. 北京航空航天大学学报，2015，41(4)：616-623. Wang Y F，Tao G Q，Liu D X，et al. Analysis of stratospheric lighter-than-air vehicle's aerodynamic characteristics based on large eddy simulation[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：616-623 (in Chinese).

数的变化情况,并对临界雷诺数时球体的绕流进行了分析,绘制了典型条件下球体的绕流图画[4]. Rodriguez等通过直接数值模拟方法研究了雷诺数为 3 700 和 10 000 时的球体绕流问题. 研究发现,尾流受控于 3 个不稳定因素:大规模涡流脱落频率,小规模 Kelvin-Helmholtz 不稳定剪切层频率以及再流通区域变化频率[5]. Moradian 等通过试验研究了雷诺数为 $2.2\times10^4\sim8\times10^4$ 时,雷诺数、湍流强度和整体尺度对球体阻力系数的影响[6]. Hassanzadeh 等通过大涡模拟(LES, Large Eddy Simulation)方法,研究了单球体和不同间距的并列双球体的绕流问题[7-8]. Pinar 等则通过试验研究了雷诺数为 5 000 时并列双球体的流场结构[9]. 国内,邹建峰等[10]采用自编的分块算法程序模拟了雷诺数在 20 ~1 000 之间的球体绕流场. 任安禄等[11]采用三维黏性不可压缩流场分块耦合计算方法研究球体绕流问题. 邹建峰和任安禄[12]对虚拟边界法加以改善并推广到三维多连通区域的数值模拟中去,研究了低雷诺数下串列双圆球的流场转捩现象.

国内外学者对球体绕流问题的研究主要集中在临界雷诺数以下,对于大尺度低速不可压超临界雷诺数范围的球体绕流问题研究几乎空白.

本文将"球形囊体型"平流层浮空器的气动模型简化为超临界雷诺数范围的串列双球体绕流问题,运用大涡模拟研究了两球体的间距对气动特性的影响.

1　数值模拟方法

1.1　控制方程

三维牛顿不可压流体控制方程为

$$\frac{\partial u_i}{\partial x_i}=0 \tag{1}$$

$$\rho\frac{\partial u_i}{\partial t}+\rho\frac{\partial}{\partial x_j}(u_iu_j)=-\frac{\partial p}{\partial x_i}+\mu\frac{\partial^2 u_i}{\partial x_j\partial x_j} \tag{2}$$

式中,ρ 为流体的密度;u_i 为流场的速度分量;p 为流场的压强;μ 为流体黏性系数.

在大涡模拟中,通过滤波函数将流动变量划分为大尺度量和小尺度量两部分:

$$f(x)=\bar{f}(x)+f'(x) \tag{3}$$

式中,$f(x)$ 为瞬时流动变量;$\bar{f}(x)$ 为大尺度量;$f'(x)$ 为小尺度量,也称为亚格子分量.

大尺度量用如下形式来表示:

$$\bar{f}(x)=\int_D f(x')G(x,x',\Delta)\,\mathrm{d}x' \tag{4}$$

式中,D 为计算流域;$G(x,x',\Delta)$ 为空间滤波函数,Δ 代表网格尺寸.

滤波函数 $G(x,x')$ 可以表示为

$$G(x,x')=\begin{cases}1/\Delta & |x-x'|\leqslant\Delta/2\\ 0 & |x-x'|>\Delta/2\end{cases} \tag{5}$$

对控制方程进行空间滤波,得到大涡模拟的基本方程为

$$\frac{\partial\bar{u}_i}{\partial x_i}=0 \tag{6}$$

$$\rho\frac{\partial\bar{u}_i}{\partial t}+\rho\frac{\partial}{\partial x_j}(\bar{u}_i\bar{u}_j)=-\frac{\partial\bar{p}}{\partial x_i}+\mu\frac{\partial^2\bar{u}_i}{\partial x_j\partial x_j}-\rho\frac{\partial\tau_{ij}}{\partial x_j} \tag{7}$$

式中,"$-$"代表经过空间滤波的变量;$\tau_{ij}=\overline{u_iu_j}-\bar{u}_i\bar{u}_j$ 为亚格子应力.

本文采用 Smagorinsky-Lilly 亚格子模型[13-14],亚格子应力定义为

$$\tau_{ij}=-2\nu_t\bar{S}_{ij}+\frac{\tau_{kk}\delta_{ij}}{3} \tag{8}$$

式中,τ_{kk} 为亚格子湍动能;$\nu_t=(C_s\Delta)^2|\bar{S}|$ 为涡黏系数;$\bar{S}_{ij}$ 为经过滤波后的速度变形张量:

$$\bar{S}_{ij}=\frac{1}{2}\left(\frac{\partial\bar{u}_i}{\partial x_j}+\frac{\partial\bar{u}_j}{\partial x_i}\right) \tag{9}$$

且 $|\bar{S}|=(2\bar{S}_{ij}\bar{S}_{ij})^{\frac{1}{2}}$;$\delta_{ij}$ 为 Kronecker 记号;C_s 为 Smagorinsky 常数.

1.2　计算模型

计算流域是横截面为正方形的长方体,囊体按球体处理,球体的直径为 D,坐标原点在上游球体球心,上下游球体球心的间距为 G(G 为 0 时则是单球体的情况),x 轴正方向指向压力出口,y 轴正方向指向上方,z 轴正方向由右手准则确定,速度入口平面距坐标原点为 $4D$,压力出口平面距坐标原点为 $10D$,计算域横截面为边长为 $8D$ 的正方形,来流速度为 V_0,如图 1 和图 2 所示.

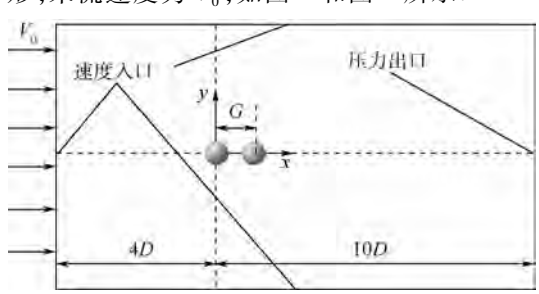


图 1　流场侧视图

Fig. 1　Side view of flow field

计算流域网格采用 O 形结构网格,附面层第一层网格的 y^+ 值为 0.5,空间最小网格步长位于

近壁面,约为 $0.6\times10^{-5}D$,球体表面网格与近壁面网格如图 3 和图 4 所示.

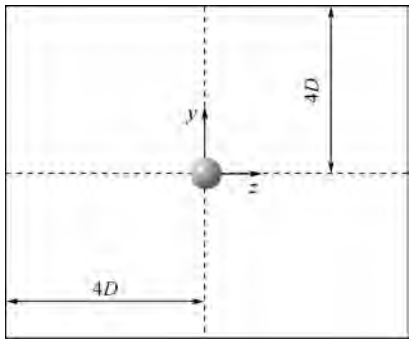


图 2 流场正视图

Fig. 2 Front view of flow field

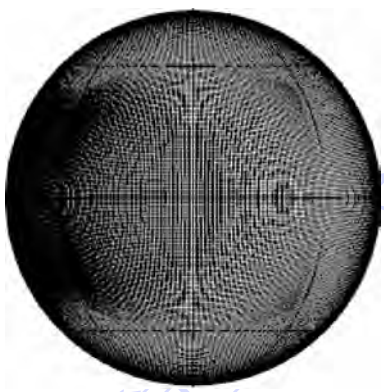


图 3 球体表面网格图

Fig. 3 Surface grid of sphere

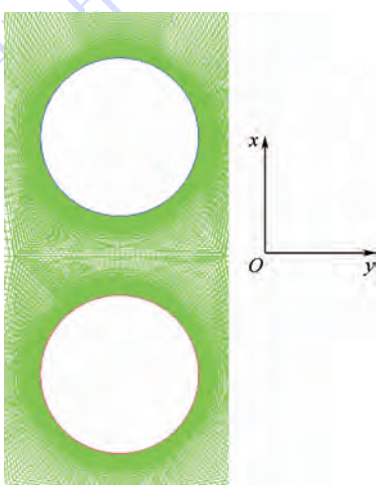


图 4 球体近壁面网格图

Fig. 4 Near wall grid of sphere

1.3 边界条件

边界条件设为速度入口、压力出口和壁面.由于已知来流速度,将入口边界定义为速度入口条件;球体表面定义为无滑移壁面边界条件;出口边界定义为压力出口边界条件.

1.4 计算方法

数值模拟采用 Fluent 软件平台,采用有限体积法对控制方程进行离散化,大涡模拟控制方程采用有限二阶隐式法推进,对流项的离散格式采用有限中心差分格式.

2 计算结果

2.1 方法验证:单球体($G=0$)

为了检验计算方法的准确性,先计算单球体在超临界雷诺数范围时($Re=1.41\times10^6$,1.88×10^6)的非稳态情况.

阻力系数定义为

$$C_D=\frac{F_x}{\frac{\pi}{8}\rho V_0^2D^2} \tag{10}$$

式中,F_x 为圆球受力沿流向的分量.

y 和 z 方向侧力系数定义为

$$C_y=\frac{F_y}{\frac{\pi}{8}\rho V_0^2D^2} \tag{11}$$

$$C_z=\frac{F_z}{\frac{\pi}{8}\rho V_0^2D^2} \tag{12}$$

式中,F_y 和 F_z 分别为圆球受力沿 y 和 z 轴方向的分量;C_y 和 C_z 为 y 和 z 方向侧力系数.

横向合力系数定义为

$$C_{\text{side}}=\sqrt{C_y^2+C_z^2} \tag{13}$$

阻力系数随时间变化如图 5 和图 6 所示.$Re=1.41\times10^6$ 时,阻力系数在 0.127 8 ~ 0.144 2 间波动,时均值为 0.136 7;$Re=1.88\times10^6$ 时,阻力系数在 0.131 5 ~ 0.160 2 间波动,时均值为 0.142 8. Achenbach 通过试验测量了球体在 $Re=5\times10^4\sim6\times10^6$ 的阻力系数[15],从 Achenbach 给出的试验数据插值出 $Re=1.41\times10^6$,1.88×10^6 时的阻力系数,试验值分别为 0.135 和 0.140. 计算结果对比如表 1 所示.

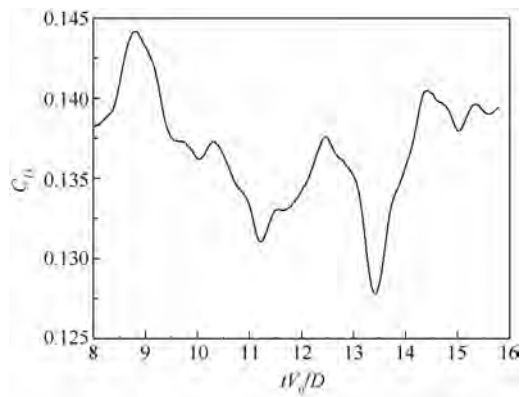


图 5 $Re=1.41\times10^6$ 时,阻力系数随时间变化图

Fig. 5 Time history of drag coefficient at $Re=1.41\times10^6$

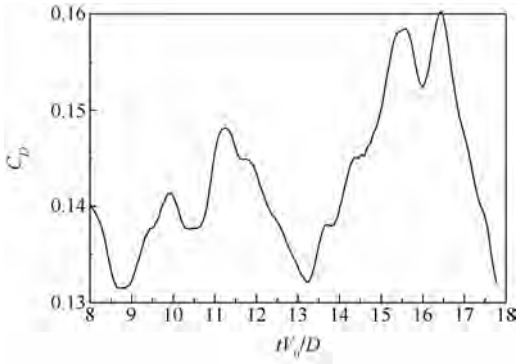


图 6　$Re = 1.88 \times 10^6$ 时,阻力系数随时间变化图

Fig. 6　Time history of drag coefficient at $Re = 1.88 \times 10^6$

表 1　计算值与试验值对比表

Table 1　Calculated value and experimental value contrasting table

Re	C_D 计算	C_D 试验	误差/%
1.41×10^6	0.136 7	0.135	1.26
1.88×10^6	0.142 8	0.140	2

由表 1 可以看出本文的计算结果与 Achenbach 的试验结果吻合良好,验证了本文方法在分析超临界雷诺数球体绕流问题时的准确性.

2.2　双 球 体

通过计算超临界雷诺数 $Re = 1.41 \times 10^6$ 时,不同间距的非稳态情况. 上下游球体球心的间距为 D 时,阻力系数随时间变化如图 7 所示. 图中:C_D-1 表示上游球体的阻力系数;C_D-2 表示下游球体的阻力系数;C_D-all 表示两球合阻力系数. 上游球体的阻力系数时均值为 0.009 3,下游球体的阻力系数时均值为 0.209 0,两球合阻力系数时均值为 0.218 3.

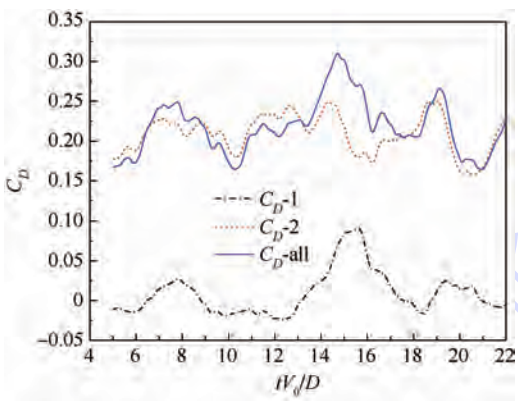


图 7　$G = D$ 时,阻力系数随时间变化图

Fig. 7　Time histories of drag coefficient when $G = D$

上下游球体球心的间距为 $1.5D$ 时,阻力系数随时间变化如图 8 所示. 上游球体的阻力系数时均值为 0.092 5,下游球体的阻力系数时均值为 0.314 2,两球合阻力系数时均值为 0.406 7.

上下游球体球心的间距为 $2D$ 时,阻力系数随时间变化如图 9 所示. 上游球体的阻力系数时均值为 0.110 1,下游球体的阻力系数时均值为 0.187 4,两球合阻力系数时均值为 0.297 5.

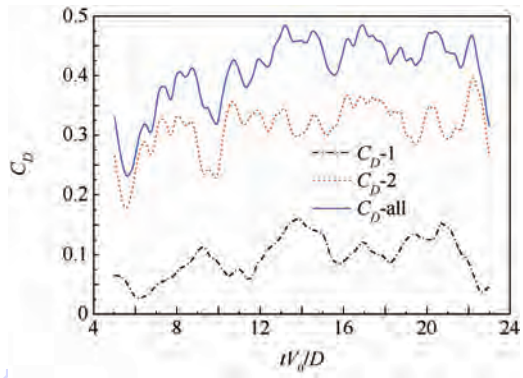


图 8　$G = 1.5D$ 时,阻力系数随时间变化图

Fig. 8　Time histories of drag coefficient when $G = 1.5D$

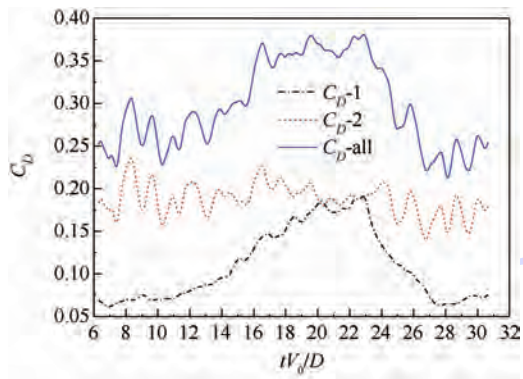


图 9　$G = 2D$ 时,阻力系数随时间变化图

Fig. 9　Time histories of drag coefficient when $G = 2D$

上下游球体球心的间距为 $3D$ 时,阻力系数随时间变化如图 10 所示. 上游球体的阻力系数时均值为 0.193 4,下游球体的阻力系数时均值为 0.108 7,两球合阻力系数时均值为 0.302 1.

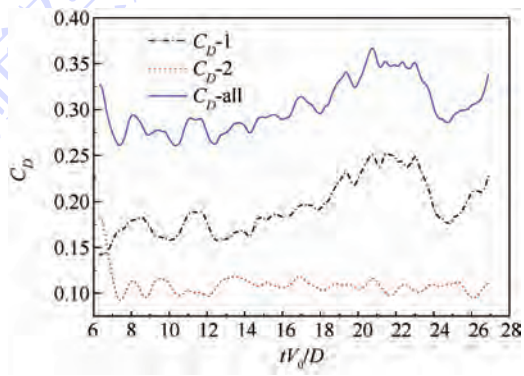


图 10　$G = 3D$ 时,阻力系数随时间变化图

Fig. 10　Time histories of drag coefficient when $G = 3D$

图 11 给出了阻力系数随间距的变化情况. 随着上下游球体距离 G 的增加,上游球体的阻力系数逐渐增大,下游球体的阻力系数先增大后减小,在 $G = 1.5D$ 时有最大值. 两球合阻力系数也先增

北航学报 赠阅

大后减小，$G=D$ 时为最小值，$G=1.5D$ 时为最大值，$G=2D$ 和 $G=3D$ 时基本相同.

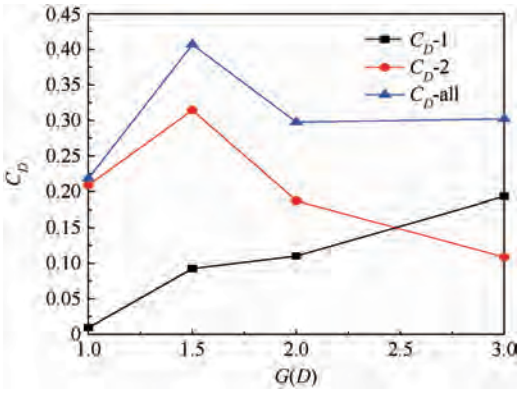


图 11 阻力系数随间距变化图

Fig. 11 Drag coefficient varies with distance

进一步分析阻力振动的频谱特性，运用快速傅里叶变换，将时域上的数据变化到频域上，得到了阻力系数和横向合力系数在频域上的幅度特征. 上下游球体阻力系数频谱特征如图 12 所示. 当$G=D$ 时，上游球体的占优频率为 0.118；下游球体的占优频率为 0.176，但下游球体还存在二、三、四次频，分别为 0.412，0.059，0.294，这 3 个次频的幅度值相差不大，与占优的主频形成竞争的态势. 当 $G=1.5D$ 时，上游球体的占优频率为 0.050，还存在二、三次频，分别为 0.151，0.252；下游球体的占优频率与上游球体相同，还存在与主频幅度值接近的二次频，频率为 0.504，三、四次频的频率分别为 0.202，0.303. 当 $G=2D$ 时，上下游球体存在相同的占优频率 0.041，下游球体还存在较多竞争的次频，频率范围为 0.123 ~ 0.739. 当 $G=3D$ 时，上游球体的占优频率为 0.049；下游球体的占优频率为 0.292，还存在较多竞争的次频，频率范围为 0.097 ~ 0.632.

图 13 显示了不同间距流向无量纲时均速度（u/V_0）的分布. 从图中可以看出，在球体背风面均有负速度分布. $G=D$ 时，由于两球接触而在球间形成凹槽，在上游球体的侧后方最小的流向速度达到了 −0.50，是 4 种间距中最小的，说明该位置有很强的流向回流，但是该间距对应的流向回流区域却是 4 种情况中最小的，球体所受的压差阻力最小，而球体的阻力主要与压差阻力相关，因

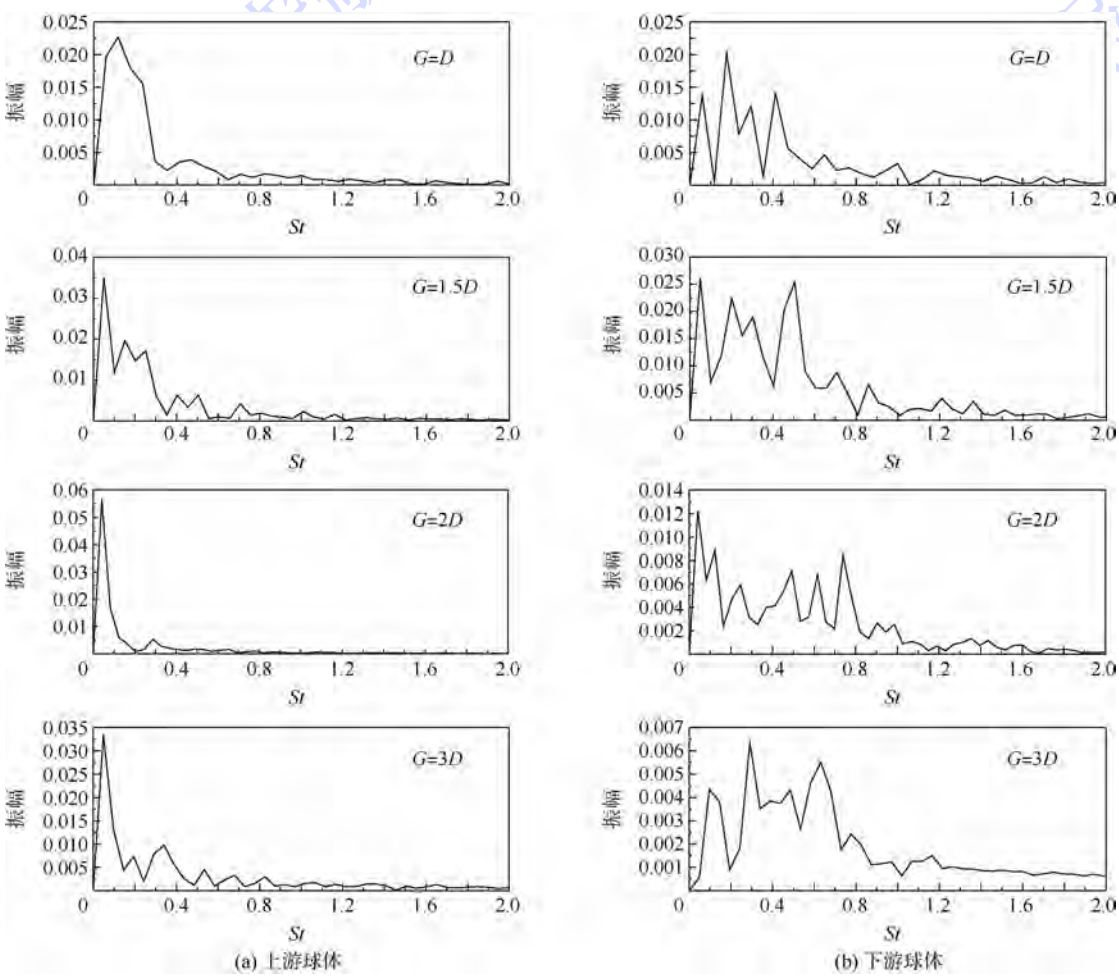


图 12 上下游球体阻力系数频谱特征图

Fig. 12 Frequency spectrum of upstream and downstream spheres' drag coefficients

北航学报 赠阅

此上游球体的阻力系数在 4 种情况中最小；下游球体的流向回流速度分布长度大约为 0.3D，是 4 种情况中最短的，球体所受的压差阻力最小，因此阻力系数在 4 种情况中也最小. 随着间距的增大，上游球体后的最小流向速度分布向中心线靠近，回流区域逐渐增加，上游球体所受的压差阻力也逐渐增大，阻力系数随之增大，如图 11 所示. 后 3 种情况的最小流向速度位置位于上游球体后，仅在 $G=2D$ 时，在下游球体后有一最小流向速度与上游球体相当. 间距从 $G=D$ 增加到 $G=3D$ 的过程中，流向速度最小值呈先增大后减小的趋势，$G=2D$ 时，流向速度最小值是 4 种情况中最大的，说明流向回流的强度最弱. $G=1.5D$ 时，中心线的流向速度皆为负值，说明在此间距条件下，流向回流充满整个区域，下游球体的流向来流速度很小，提前了边界层的分离，增大了压差阻力，导致阻力系数增大. 随间距增加，两球体间的相互作用逐渐减弱，下游球体流向来流速度逐渐增大到远前方值，边界层分离逐渐推迟，压差阻力逐渐减小，阻力系数也随之减小.

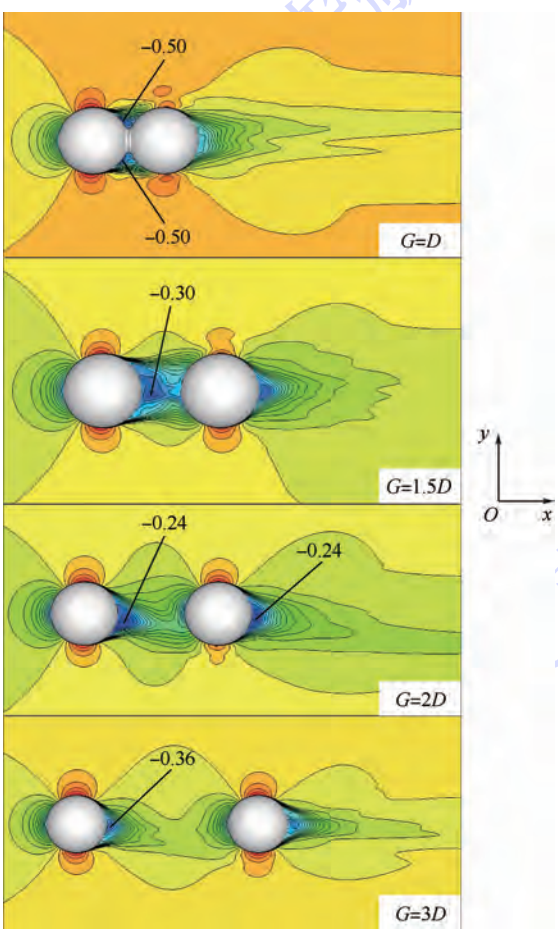


图 13　不同间距流向无量纲时均速度图

Fig. 13　Patterns of time-averaged streamwise velocity at different spacing

为了显示尾流中的涡结构，需要采用三维流场的涡识别方法. 本文采用 Jeong 和 Hussain 的 Q-定义方法[16]. 当 $G=D$ 时，尾涡结构变化如图 14 所示. 上游球体的上下面交替脱出“Ω”型涡，下游球体的尾涡结构受上游球体尾涡结构影响，较为复杂.

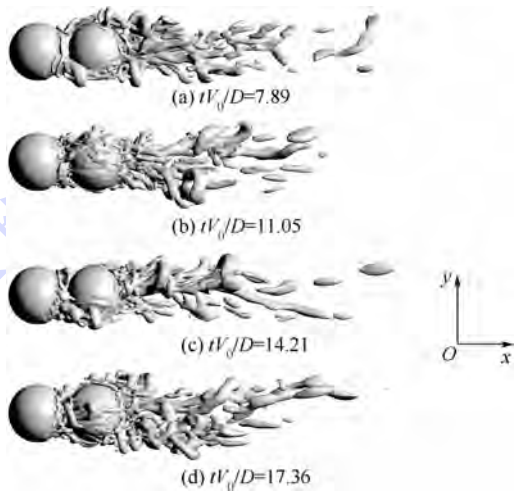


图 14　$G=D$ 时，不同时刻尾涡结构图

Fig. 14　When $G=D$, the trailing vortex structure at different time

当 $G=1.5D$ 时，尾涡结构变化如图 15 所示. 与 $G=D$ 时不同，上游球体的尾涡仅从下面脱出，绕过下游球体后有增大趋势. 与 $G=D$ 时相同，下游球体尾涡结构受上游尾涡影响，结构复杂.

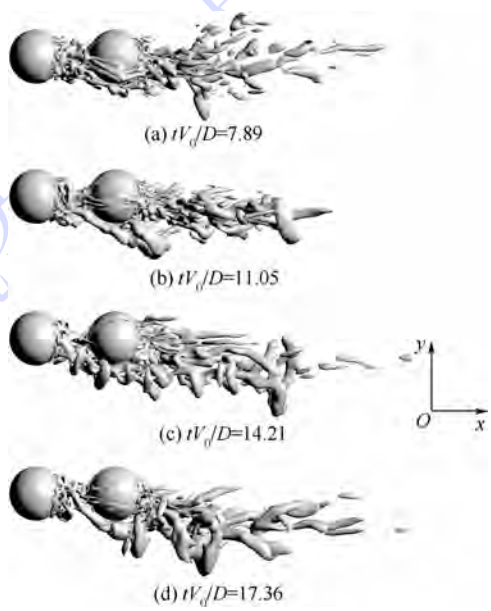


图 15　$G=1.5D$ 时，不同时刻尾涡结构图

Fig. 15　When $G=1.5D$, the trailing vortex structure at different time

当 $G=2D$ 时，尾涡结构变化如图 16 所示. 相比 $G=1.5D$ 时，上游球体脱出的尾涡在到达下游

球体之前能更充分的发展,两球体间大约有 2、3 个“Ω”型涡环.

当 $G=3D$ 时,尾涡结构变化如图 17 所示. 随着两球间距的加大,上游球体的尾涡结构对下游球体的影响逐渐减弱,下游球体脱出的尾涡结构能够观察到“Ω”型涡环.

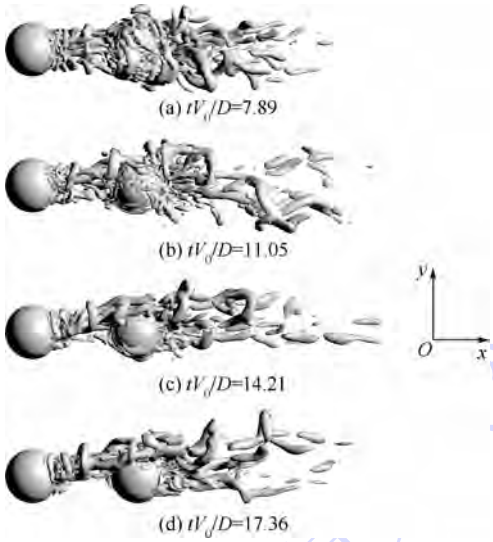


图 16 $G=2D$ 时,不同时刻尾涡结构图

Fig. 16 When $G=2D$, the trailing vortex structure at different time

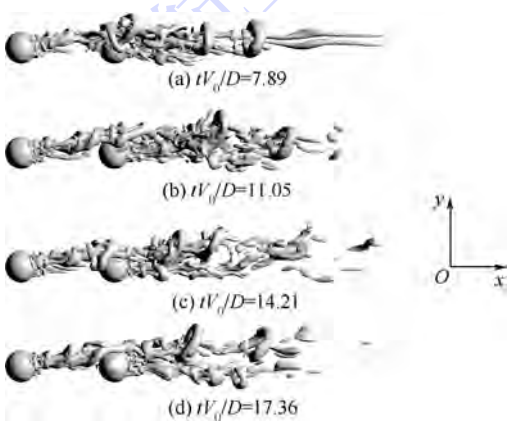


图 17 $G=3D$ 时,不同时刻尾涡结构图

Fig. 17 When $G=3D$, the trailing vortex structure at different time

3 结 论

采用大涡模拟法对“球形囊体型”平流层浮空器的气动模型(超临界雷诺数串列双球体)进行了数值研究,得出相关结论,对平流层浮空器的气动设计具有一定的参考意义.

1) 本文方法在分析超临界雷诺数球体绕流问题时具有准确性,可以为“球形囊体型”平流层浮空器气动设计提供参考依据.

2) 随上下游球体间距 G 增加,上游球体的阻力增大,下游球体的阻力先增大后减小,两球体合阻力先增大后减小,$G=D$ 时合阻力最小,$G=1.5D$时合阻力最大.

3) 随上下游球体间距 G 增加,上游球体的阻力振动频率逐渐减小. $G=D$ 时,下游球体的阻力振动频率略大于上游球体,$G=1.5D$ 和 $G=2D$ 时,上下游球体有相同的占优振动频率,$G=3D$ 时,下游球体的占优振动频率为上游球体的 6 倍.

4) $G=D$ 时,在上游球体的侧后方流向回流速度最强,随着间距的增大,上游球体后的最小流向速度分布向中心线靠近,$G=1.5D$ 时,流向回流充满中心线附近区域,随间距继续增加,两球体间的相互作用逐渐减弱.

5) $G=D$ 时,上游球体的上下面交替脱出“Ω”型涡,随间距增大,上游球体的尾涡仅从下面脱出,当间距达到 $G=3D$ 时,上游球体的尾涡结构对下游球体的影响较小.

参考文献(References)

[1] Colozza A. Initial feasibility assessment of a high altitude long endurance airship, NASA/CR-2003-212724[R]. Washington, D. C.: NASA, 2003.

[2] Volpe J. Lockheed Martin's HALE-D airship learns to fly makes a crash landing[EB/OL]. (2011-07-28). http://www.engadget.com/2011/07/28/lockheed-martins-hale-d-airship-learns-to-fly-makes-a-crash-la/.

[3] Sakamoto H, Haniu H. A study on vortex shedding from spheres in a uniform flow[J]. Journal of Fluids Engineering, 1990, 112(4): 386-392.

[4] Constantinescu G, Squires K. Numerical investigations of flow over a sphere in the subcritical and supercritical regimes[J]. Physics of Fluids, 2004, 16(5): 1449-1466.

[5] Rodriguez I, Lehmkuhl O, Borrell R, et al. Flow dynamics in the turbulent wake of a sphere at sub-critical Reynolds numbers[J]. Computers & Fluids, 2013, 80: 233-243.

[6] Moradian N, Ting D S K, Cheng S. The effects of freestream turbulence on the drag coefficient of a sphere[J]. Experimental Thermal and Fluid Science, 2009, 33(3): 460-471.

[7] Hassanzadeh R, Sahin B, Ozgoren M. Numerical investigation of flow around a sphere[J]. International Journal of Computational Fluid Dynamics, 2011, 25(10): 535-545.

[8] Hassanzadeh R, Sahin B, Ozgoren M. Large eddy simulation of flow around two side-by-side spheres[J]. Journal of Mechanical Science and Technology, 2013, 27(7): 1971-1979.

[9] Pinar E, Sahin B, Ozgoren M, et al. Experimental study of flow structures around side-by-side spheres[J]. Industrial and Engineering Chemistry Research, 2013, 52(40): 14492-14503.

[10] 邹建峰,任安禄,邓见. 圆球绕流场的尾涡分析和升阻力研究[J]. 空气动力学学报, 2004, 22(3): 303-308.
Zou J F, Ren A L, Deng J. Numerical investigations of wake and force for flow past a sphere[J]. Acta Aerodynamica Sinica, 2004, 22(3): 303-308(in Chinese).

北航学报 赠阅

[11] 任安禄,李广望,邹建峰. 中等雷诺数圆球绕流的数值研究[J]. 浙江大学学报,2004,38(5):644-648.
Ren A L, Li G W, Zou J F. Numerical study of uniform flow over sphere at intermediate Reynolds numbers[J]. Journal of Zhejiang University, 2004, 38(5): 644-648(in Chinese).

[12] 邹建锋,任安禄. 串列双圆球的流场转捩研究[J]. 浙江大学学报,2006,40(1):107-112.
Zou J F, Ren A L. Study on transitions for flow past two spheres in tandem arrangement[J]. Journal of Zhejiang University, 2006, 40(1): 107-112(in Chinese).

[13] Germano M, Piomelli U, Moin P, et al. A dynamic subgrid-scale eddy viscosity model[J]. Physics of Fluids A: Fluid Dynamics (1989-1993), 1991, 3(7): 1760-1765.

[14] Wilcox D C. Turbulence modeling for CFD[M]. 2nd ed. La Caflada, CA: DCW Industries, Inc., 1998.

[15] Achenbach E. Experiments on the flow past spheres at very high Reynolds numbers[J]. Journal of Fluid Mechanics, 1972, 54(3): 565-575.

[16] Jeong J, Hussain F. On the identification of a vortex[J]. Journal of Fluid Mechanics, 1995, 285: 69-94.

Analysis of stratospheric lighter-than-air vehicle's aerodynamic characteristics based on large eddy simulation

WANG Yufu, TAO Guoquan*, LIU Dongxu, WU Zhe

(School of Aeronautic Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: Flowing over stratospheric lighter-than-air vehicle (SLV) with spherical capsule shape at supercritical Reynolds number were simulated by using large eddy simulation (LES) method. Meanwhile, the results at different sphere spacing were analyzed and compared so that the influence of different sphere spacing on stratospheric lighter-than-air vehicle's aerodynamic characteristics was obtained. Compared with the results of the experiments, the calculations of time-average drag coefficients agreed well with Achenbach's experimental results in the single sphere. The method was verified accurate on the analysis of flow over spheres at supercritical Reynolds number. The drag coefficients of the two spheres at different sphere spacing and frequency spectrum were studied. With the increase of the upstream and downstream spheres' spacing, the whole drags increased first, then decreased, and the the dominant frequencies of upstream sphere decreased. The upstream and downstream spheres had the same dominant frequency of vibration when the spheres' spacing were $G=1.5D$ and $G=2D$. With the increasing of G, the interaction between the two spheres and the influence of upstream sphere's trailing vortex to downstream sphere's trailing vortex waned.

Key words: stratospheric lighter-than-air vehicle; large eddy simulation; flow over spheres; drag coefficient; frequency spectrum

北京航空航天大学学报

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0666

基于压缩感知的 DSSS 信号双阶段捕获方法

程艳合，杨文革*
（装备学院 光电装备系，北京 101416）

摘 要：针对直接序列扩频（DSSS，Direct Sequence Spread Spectrum）信号的捕获性能与硬件消耗或计算复杂度的相互制约问题，基于压缩感知理论，提出了一种双阶段压缩捕获方法，第 1 阶段进行快速粗捕获，第 2 阶段在第 1 阶段基础上实现精确捕获. 首先研究 DSSS 信号的相关域稀疏性，构造了稀疏变换矩阵；然后利用确定性沃尔什-阿达马矩阵，分别构造了 2 个阶段压缩测量矩阵及其检测算法；最后从检测概率和平均捕获时间两方面对提出算法的捕获性能进行了理论分析，并用蒙特卡罗法进行了验证. 理论分析和仿真实验表明，该方法能够在显著降低相关次数的前提下，达到传统基于并行相关方法的捕获性能水平.

关 键 词：直接序列扩频信号；压缩感知；测量矩阵；检测概率；平均捕获时间
中图分类号：TN914.4
文献标识码：A **文章编号**：1001-5965(2015)04-0624-08

直接序列扩频（DSSS，Direct Sequence Spread Spectrum）信号具有抗干扰、抗噪声、抗多径干扰、低截获、测距性能好和可码分多址等优点，被广泛应用于军事通信、导航和航天测控等领域. 扩频通信优越性以精确同步为前提，同步过程一般包括捕获阶段、跟踪阶段 2 个阶段，其中捕获是首要任务. 信号捕获的主要限制因素是载波频率和伪码相位的不确定性. DSSS 信号的捕获过程就是在整个多普勒频率和码相位上以固定间隔进行二维搜索，这对接收机的硬件或计算能力提出了很高的要求，尤其是在高动态情况下，长码 DSSS 信号的快速捕获实现困难，捕获范围和平均捕获时间（MAT，Mean Acquisition Time）是一对相互限制指标[1]. 为了实现信号的快速捕获，接收机可以采用大量的并行相关器来降低捕获时间，但是随着捕获范围的增大，接收机硬件复杂度也会急剧增大[2]. 在没有辅助情况下，同时降低相关器个数和捕获时间往往难以实现[3]. 另外一个选择就是基于快速傅里叶变换（FFT，Fast Fourier Transform）的 DSSS 信号捕获方法，通过专用数字信号处理器快速地实现大量的假设检验处理，能够降低接收机的硬件资源消耗. 基于 FFT 的捕获方法首先分别对接收信号和本地生成伪码进行 FFT 变换，然后在频域进行复数相乘，最后对相乘结果进行逆 FFT 变换，但是也存在计算量大和高动态适应性等问题[4-6]. 总之，相对于上述传统信号捕获方法，一种理想的捕获方法应该能够同时降低相关器和计算量需求.

最近，出现了一种新型 DSSS 信号捕获方法，其采用折叠或双折叠技术[7-8]. 折叠技术对本地生成 PN 码进行折叠处理，降低了 MAT，但是会不可避免地引入信噪比（SNR，Signal to Noise Ratio）恶化. 为了降低折叠技术带来的 SNR 恶化，双折叠技术同时对接收信号进行折叠处理，但是 MAT 会有所增大.

压缩感知（CS，Compressive Sensing）是另一种可能用来实现 DSSS 信号快捕的新兴技术. CS 技

收稿日期：2014-10-28；**录用日期**：2014-11-28；**网络出版时间**：2014-12-22 11:14
网络出版地址：www.cnki.net/kcms/detail/11.2625.V.20141222.1114.004.html
作者简介：程艳合（1987—），男，河北衡水人，博士生，cheng20130810@foxmail.com
*通讯作者：杨文革（1966—），男，江西金溪人，教授，yangttc1508@foxmail.com，主要研究方向为空间飞行器测控与通信技术.

引用格式：*程艳合，杨文革. 基于压缩感知的 DSSS 信号双阶段捕获方法[J]. 北京航空航天大学学报，2015，41(4)：624-631.*
Cheng Y H，Yang W G. Two-stage acquisition algorithm for DSSS signal based on compressive sensing[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：624-631 (in Chinese).

术可以利用信号的稀疏性，降低能够完美表达信号所需的采样点数目. 从理论上说，CS 技术找到了能够解决欠定线性方程组的方法[9-11]. 在 DSSS 系统中，假设扩频伪码序列已知，则接收信号能够被很小的信息表达，即其码相位和多普勒频率. 例如，当本地生成信号接近于接收信号时，虽然伪码长度可以为 1 023，但是如果信号在相关域表示，其稀疏度仅为 1，此时信号最稀疏. 在传统的 DSSS 信号捕获方法中，接收机检验 N 个码相位假设来获取 N 个测量值，需要 N 个相关器，而根据压缩感知技术，相关器数量可以远远小于 N. 因此，DSSS 信号捕获可以利用压缩感知技术，以期有效降低所需相关输出测量值个数. 应用压缩感知理论，首先必须找到能够最稀疏表达目标信号的稀疏变换矩阵，其次需要寻求合适的测量矩阵及其相对应的有效重构算法.

迄今为止，国内外学者已经提出了多种测量矩阵和重构算法，其中随机 Gaussian 矩阵和随机 Bernoulli 矩阵已经在压缩感知中被广泛使用，但是与其相对应的重构算法普遍都有较高的计算复杂度[12]. 因此，有学者提出了确定性压缩感知技术，该类压缩感知采用确定性测量矩阵，测量矩阵对应的重构算法计算复杂度较低[13]. 其中，确定性的 Walsh-Hadamard 矩阵由于具有类似于 FFT 的快速 Walsh-Hadamard 变换（FWHT，Fast Walsh-Hadamard Transform）算法而广受关注[14].

基于 DSSS 信号在其自相关域较强的稀疏性，本文提出了一种双阶段压缩捕获方法，不仅能够同时降低计算复杂度和硬件消耗，而且与折叠、双折叠技术相比，该方法的噪声鲁棒性明显提高. 与传统基于并行相关的捕获方法进行对比分析，仿真结果表明双阶段压缩捕获方法在有效降低相关运算量的前提下，可以达到基于并行相关方法的性能水平.

1　DSSS 信号模型分析

在扩频测控体制中一般采用调制效率较高的 PCM-CDMA-BPSK 调制体制，测控数据被伪随机码扩频后再经载波调制后发射. 测控基带设备接收到的中频扩频测控信号可表示为

$$x(t)=AP(t-\tau)D(t-\tau)\cdot \cos[2\pi(f_c+f_d)t+\theta]+n(t) \tag{1}$$

式中，A 为载波振幅；$P(t)$ 为伪码速率为 R_c 的扩频伪码；τ 为伪码相位；$D(t-\tau)$ 为比特率为 R_b（$R_b=1/T_b \ll R_c=1/T_c$）的数据码，T_b 为数据码时长，T_c 为伪码时长；f_c 为中频频率；f_d 为多普勒频率；θ 为未知的接收信号载波相位；$n(t)$ 为双边谱密度为 $N_0/2$ 的加性高斯白噪声，N_0 为接收信号噪声谱密度.

伪码捕获是通过搜索由所有可能的伪码相位和多普勒频率组合构成的二维假设平面，寻找接收信号对应的正确假设，进而得到信号的伪码相位 τ 和多普勒频率 f_d，如图 1 所示，其中 $[f_{min}, f_{max}]$ 是多普勒频率 f_d 变化范围. 假设接收机捕获码相位精度为半个码片之内，采样率（f_s）取 2 倍的伪码速率，即 $f_s=2R_c$.

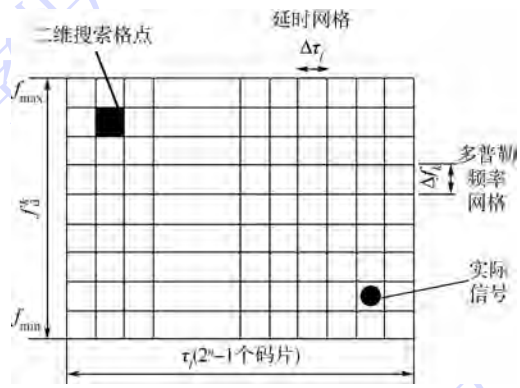


图 1　多普勒频率与伪码相位二维搜索

Fig. 1　Two-dimensional search for Doppler frequency and pseudo-code phase

在传统的捕获方法中，接收机首先要通过相关器对接收信号和本地生成参考信号进行相关处理，该本地信号可表示为

$$x_0(t)=P(t-\tau_l)e^{j2\pi(f_c+f_d^k)t} \tag{2}$$

式中，f_d^k 和 τ_l 分别为第 k 个多普勒频率假设对应的多普勒频率和第 l 个伪码相位假设对应的码相位；信号 $x(t)$ 与信号 $x_0(t)$ 之间的相关运算实际是在离散时间域内进行，但是为了简化代数分析模型，在连续时间域内进行理论分析. 假设相关积分时间为 T，且有 $T_c \ll T \ll T_b$，在相关时间内可以认为数据保持不变 $D(u-\tau)=D$. 则经过低通滤波的归一化相关器输出结果为

$$R(\Delta f_k,\Delta\tau_l)=\frac{1}{T}\int_t^{t+T}[x(u)x_0^*(u)]\cdot h_L(u)\,\mathrm{d}u\Big|_{t=t_0}=\frac{AD}{T}\int_{t_0}^{t_0+T}[P(u)P(u-\Delta\tau_l)e^{j(2\pi\Delta f_k u+\theta)}+n(u)P(u-\tau_l)e^{j2\pi(f_c+f_d^k)u}]\,\mathrm{d}u \tag{3}$$

式中，t_0 为积分开始时刻；x_0^* 为本地信号 x_0 的复共轭；$h_L(t)$ 为低通滤波器可以滤除相关器输出信号的高频分量. 输出结果由多普勒频率差异 $\Delta f_k=$

$f_d - f_d^k$ 和码相位差异 $\Delta\tau_l = \tau - \tau_l$ 共同决定. 在无噪声时,$R(\Delta f_k, \Delta\tau_l)$ 可以转变为

$$R(\Delta f_k, \Delta\tau_l) = \frac{AD}{T}\int_{t_0}^{t_0+T} P(u)P(u-\Delta\tau_l)\mathrm{e}^{\mathrm{j}(2\pi\Delta f_k u+\theta)}\mathrm{d}u = \sum_{k=0}^{T/T_c-1}\left[\int_{t_0+kT_c}^{t_0+(k+1)T_c-|\Delta\tau_l|}\mathrm{e}^{\mathrm{j}(2\pi\Delta f_k u+\theta)}\mathrm{d}u + \int_{t_0+(k+1)T_c-|\Delta\tau_l|}^{t_0+(k+1)T_c} P(u)\cdot P(u-\Delta\tau_l)\mathrm{e}^{\mathrm{j}(2\pi\Delta f_k u+\theta)}\mathrm{d}u\right] \tag{4}$$

根据式(4)中第 2 个积分项的积分范围,可知 $P(u)$ 和 $P(u-\Delta\tau_l)$ 相互独立,则该积分项可忽略,并且假设 $A=1, D=1$,式(4)可进一步近似为

$$R(\Delta f_k, \Delta\tau_l) \approx \mathrm{sinc}(\pi\Delta f_k T)\cdot\frac{\sin[\pi\Delta f_k(T_c-|\Delta\tau_l|)]}{\sin(\pi\Delta f_k T_c)}\cdot \mathrm{e}^{\mathrm{j}(2\pi\Delta f_k t_0+\theta)}\cdot\mathrm{e}^{\mathrm{j}(\pi\Delta f_k T)}\cdot\mathrm{e}^{-\mathrm{j}(\pi\Delta f_k|\Delta\tau_l|)} \tag{5}$$

式中,$\mathrm{sinc}(x)=\sin(x)/x$. 由式(5)可知,当 $\Delta f_k = 0$ 时,$R(0,\Delta\tau_l)$ 表示伪随机码的自相关函数(ACF, Auto Correlation Function),即

$$R(0,\Delta\tau_l) \approx \begin{cases} T(1-|\Delta\tau_l|/T_c) & |\Delta\tau_l| \leqslant T_c \\ 0 & \text{其他} \end{cases} \tag{6}$$

假设接收机采样率 $f_s = 2R_c$,则式(6)中只有 3 个连续伪码相位假设可以产生信号相关能量,自相关函数曲线如图 2 所示.

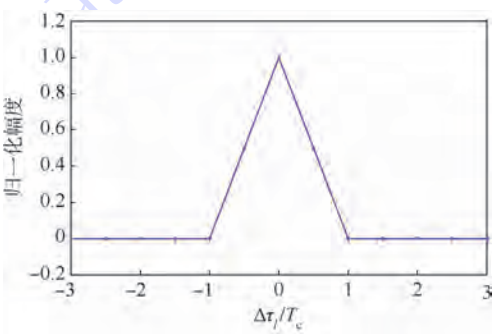


图 2 自相关函数曲线

Fig. 2 Auto correlation function curve

根据以上分析可知,DSSS 信号在其自相关域内具有较强的稀疏性. 假设采样率 $f_s = 2R_c$,接收信号 $\boldsymbol{x} = \{x[0], x[1], \cdots, x[2L-1]\}^{\mathrm{T}}$ 的稀疏变换矩阵为 $\boldsymbol{\Psi}$,则其构成元素可以表示为

$$\psi_{l,k} = P\left[\mathrm{mod}\left(\left\lceil\frac{(k-l)R_c}{f_s}\right\rceil, L\right)\right] \tag{7}$$

式中,L 为扩频伪码长度;l 和 k 表示该元素在 $\boldsymbol{\Psi}$ 中的位置,且有 $l,k \in \{0,1,\cdots,2L-1\}$;$\lceil\cdot\rceil$ 是向上取整运算;mod 为取模运算.

稀疏变换矩阵 $\boldsymbol{\Psi}$ 能够把接收信号 $\boldsymbol{x}$ 变换为低稀疏度表示形式,即

$$\boldsymbol{y}_0 = \boldsymbol{\Psi}\boldsymbol{x} \tag{8}$$

式(8)可以等价为 $2L$ 个并行相关器的相关运算过程,相位分辨率为半个码片.

2 双阶段压缩捕获方法研究

基于第 1 节 DSSS 信号模型分析,提出了基于确定性沃尔什-阿达马矩阵的双阶段压缩捕获算法,该算法主要包括 2 个处理阶段,算法流程图如图 3 所示. 其中,$\boldsymbol{\Phi}_1' = \boldsymbol{\Phi}_1\boldsymbol{\Psi}$ 为等效第 1 阶段测量矩阵,$\boldsymbol{\Phi}_1$ 为第 1 阶段测量矩阵;$\boldsymbol{\Phi}_2'$ 为等效第 2 阶段测量矩阵. 下面分别加以论述.

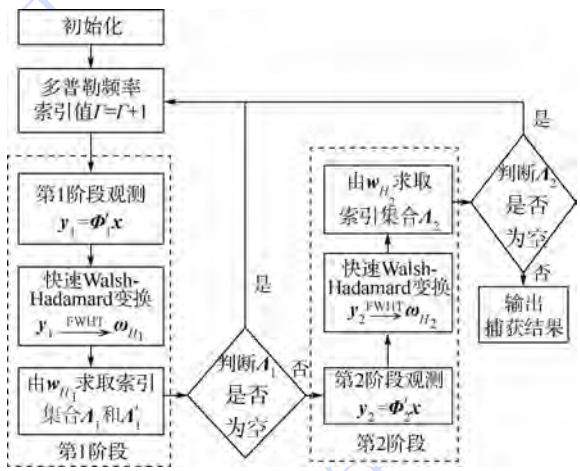


图 3 双阶段压缩捕获算法流程图

Fig. 3 Two-stage compressive acquisition algorithm flow chart

2.1 第 1 阶段压缩测量及检测

基于确定性沃尔什-阿达马矩阵,构造第 1 阶段测量矩阵 $\boldsymbol{\Phi}_1$,其构成元素为

$$\phi_{m,l+1}^1 = M_1^{-1/2} W_{M_{01}}\left(m, \left\lceil\frac{l+1}{\alpha}\right\rceil\right) \tag{9}$$

式中,$m(1\leqslant m\leqslant M_1)$ 和 $l(0\leqslant l\leqslant 2L-1)$ 表示该元素在 $\boldsymbol{\Phi}_1$ 中的位置;$M_1 = 2^{M_{01}}$,$\boldsymbol{W}_{M_{01}}$ 是大小为 $M_1\times M_1$ 的 Walsh-Hadamard 矩阵;$\alpha(\geqslant 1)$ 为 $\boldsymbol{W}_{M_{01}}$ 矩阵的列重复比. 由式(9)可知,$2L$ 列的测量矩阵 $\boldsymbol{\Phi}_1$ 是由 M_1 列的 Walsh-Hadamard 矩阵 $\boldsymbol{W}_{M_{01}}$ 扩展而得,矩阵 $\boldsymbol{W}_{M_{01}}$ 前 M_1-1 列分别重复 α 次构成矩阵 $\boldsymbol{\Phi}_1$ 的前 $\alpha(M_1-1)$ 列,矩阵 $\boldsymbol{W}_{M_{01}}$ 的最后一列重复构成矩阵 $\boldsymbol{\Phi}_1$ 的其余 $2L-\alpha(M_1-1)$ 列.

第 1 阶段测量过程如下:

$$\boldsymbol{y}_1 = \boldsymbol{\Phi}_1\boldsymbol{y}_0 = \boldsymbol{\Phi}_1\boldsymbol{\Psi}\boldsymbol{x} = \boldsymbol{\Phi}_1'\boldsymbol{x} \tag{10}$$

式中,$\boldsymbol{y}_1$ 为第一阶段测量值,大小为 $M_1\times 1$. 式(10)表述了第 1 阶段测量过程,测量矩阵 $\boldsymbol{\Phi}_1$ 通过 M_1 个 Walsh 码获得 M_1 个测量值,每个测量值

都是 $\boldsymbol{y}_0$ 的 α 个连续元素之和,根据式(6)可知,当 $\alpha \geqslant 3$ 时,该测量值能够包含信号的大部分能量,因为 Walsh 码具有正交性,所以接收机可以从 M_1 个 Walsh 码中找到具有最大测量值的 Walsh 码,并且该索引能够以 $\alpha/2$ 个码片的分辨率给出 ACF 输出峰值的位置. 注意,上述过程可以看作是由等效第一阶段测量矩阵 $\boldsymbol{\Phi}_1'$ 行向量构成的 M_1 个压缩相关器实现的相关运算处理,因此,提出的双阶段压缩捕获方法相对于传统基于并行相关器的方法,第 1 阶段测量值(或相关器)数量降低了 $\alpha \approx 2L/M_1$ 倍.

根据第 1 阶段的观测向量 $\boldsymbol{y}_1$,第 1 阶段检测算法通过 M_1 点 FWHT 生成向量 $\boldsymbol{w}_{\mathrm{H}_1}$,目的是为了寻找最大系数所对应的 Walsh 码[13]. I 路和 Q 路分别进行 FWHT,生成相应的向量 $\boldsymbol{w}_{\mathrm{H}_1}$ 组成部分,需要 $2M_1\mathrm{lb}M_1$ 次乘法,并且为了获得 $W_{\mathrm{H}_1} = |\boldsymbol{w}_{\mathrm{H}_1}|^2$,需要 $2M_1$ 次乘法和 $2M_1 - 1$ 次加法. 因此,第 1 阶段总共需要 $2M_1(\mathrm{lb}M_1 + 1)$ 次乘法. 把 $\boldsymbol{W}_{\mathrm{H}_1}$ 的每个元素与第 1 阶段检测门限 γ_1 进行比较,找出大于门限值的 $N_{\max}$ 个最大元素,假设这些元素索引由 $\boldsymbol{\Lambda}_1$ 表示,则有 $\boldsymbol{\Lambda}_1 \subset \{0,1,\cdots,M_1-1\}$ 和 $\|\boldsymbol{\Lambda}_1\|_0 \leqslant N_{\max}$. 假设第 1 阶段检测成功,$\boldsymbol{\Lambda}_1$ 的每个元素都可能包含 ACF 输出的正确峰值位置,由 $\boldsymbol{\Lambda}_1$ 扩展可得出峰值的所有可能位置的索引集合 $\boldsymbol{\Lambda}_1'$,如下:

$$\boldsymbol{\Lambda}_1' = \cup\left\{\alpha\boldsymbol{\Lambda}_1 + 1, \alpha\boldsymbol{\Lambda}_1 + 2, \cdots, \min\{\alpha(\boldsymbol{\Lambda}_1 + 1), 2L\}\right\} \tag{11}$$

2.2　第 2 阶段压缩测量及检测

基于第 1 阶段检测情况,第 2 阶段需要使用 $\|\boldsymbol{\Lambda}_1'\|_0$ 个并行压缩相关器检测集合 $\boldsymbol{\Lambda}_1'$ 中包含的 $\|\boldsymbol{\Lambda}_1'\|_0$ 个所有可能码相位,寻找 $\boldsymbol{y}_0$ 的 ACF 输出峰值位置(以半个码片精度),该阶段测量矩阵 $\boldsymbol{\Phi}_2$ 的构成元素为

$$\phi_{m,l}^2 = \begin{cases} M_2^{-1/2} W_{M_{02}}(m,k) & l = \boldsymbol{\Lambda}_1'(k) \\ 0 & l \neq \boldsymbol{\Lambda}_1'(k) \end{cases} \tag{12}$$

式中,k($k = 1,2,\cdots,\|\boldsymbol{\Lambda}_1'\|_0$)为 $\boldsymbol{\Lambda}_1'$ 的索引值;$M_{02} = \lceil \mathrm{lb}\|\boldsymbol{\Lambda}_1\|_0 \rceil$;$M_2 = 2^{M_{02}}$,$\boldsymbol{W}_{M_{02}}$ 是大小为 $M_2 \times M_2$ 的 Walsh-Hadamard 矩阵. 注意,测量矩阵 $\boldsymbol{\Phi}_2$ 有 $2L - \|\boldsymbol{\Lambda}_1'\|_0$ 个全零列向量,通过调整门限值 γ_1 和最大元素容限 $N_{\max}$,集合 $\boldsymbol{\Lambda}_1'$ 的大小,即 $\|\boldsymbol{\Lambda}_1'\|_0$,可以很小甚至为 0. 而且,一般情况下假设 $M_2 < M_1$.

第 2 阶段测量过程如下:

$$\boldsymbol{y}_2 = \boldsymbol{\Phi}_2\boldsymbol{y}_0 = \boldsymbol{\Phi}_2\boldsymbol{\Psi}\boldsymbol{x} = \boldsymbol{\Phi}_2'\boldsymbol{x} \tag{13}$$

式中,$\boldsymbol{y}_2$ 为第 2 阶段测量值,大小为 $M_2 \times 1$;$\boldsymbol{\Phi}_2' = \boldsymbol{\Phi}_2\boldsymbol{\Psi}$ 是等效第 2 阶段测量矩阵. 注意,上述过程可以看作是由等效第 2 阶段测量矩阵 $\boldsymbol{\Phi}_2'$ 行向量构成的 M_2 个压缩相关器实现的相关运算处理. 测量值 $\boldsymbol{y}_2$ 可以看做 Walsh 码的线性组合,如下:

$$\boldsymbol{y}_2 = \sum_{k=1}^{\|\boldsymbol{\Lambda}_1\|_0} M_2^{-1/2}\boldsymbol{y}_0(\boldsymbol{\Lambda}_1'(k))\boldsymbol{W}_{M_{02}}(\cdot,k) \tag{14}$$

式中,$\boldsymbol{W}_{M_{02}}(\cdot,\mathrm{k})$ 表示 $\boldsymbol{W}_{M_{02}}$ 的第 k 个 Walsh 码.

根据第 2 阶段观测向量 $\boldsymbol{y}_2$,第 2 阶段检测算法通过 M_2 点 FWHT 生成向量 $\boldsymbol{w}_{\mathrm{H}_2}$,目的也是为了寻找最大系数所对应的 Walsh 码. I 路和 Q 路分别进行 FWHT,生成相应的向量 $\boldsymbol{w}_{\mathrm{H}_2}$ 组成部分,需要 $2M_2\mathrm{lb}M_2$ 次乘法,并且为了获得 $W_{\mathrm{H}_2} = |\boldsymbol{w}_{\mathrm{H}_2}|^2$,需要 $2M_2$ 次乘法和 $2M_2 - 1$ 次加法. 因此,第 2 阶段总共需要 $2M_2(\mathrm{lb}M_2 + 1)$ 次乘法. 把 $\boldsymbol{W}_{\mathrm{H}_2}$ 的每个元素与第 2 阶段检测门限 γ_2 进行比较,找出大于门限值的最大元素,假设该元素索引为 $\boldsymbol{\Lambda}_2 \subset \{0,1,\cdots,M_2-1\}$,则 $\boldsymbol{\Lambda}_1'(\boldsymbol{\Lambda}_2)$ 就是最终捕获的码相位值,并且其捕获精度是半个码片,满足后续的跟踪要求.

3　性能分析

主要从检测概率、平均捕获时间 2 方面对提出的双阶段压缩捕获方法进行性能分析.

3.1　检测概率

首先给出接收信号信噪比 P_{SNR} 和稀疏化信号(即相关检测输出)信噪比 P_{SNR0} 定义,即

$$P_{\mathrm{SNR}} = 20\lg\left(\frac{A}{2\sigma_{\mathrm{v}}}\right) \tag{15}$$

$$P_{\mathrm{SNR0}} = 20\lg\left(\frac{G_0}{2\sigma_{\mathrm{v}}}\right) \tag{16}$$

式中,$G_0 = 2AL$ 是相关输出最大幅度值;σ_{v} 为输入信号噪声功率. 第 1 阶段中 I 路和 Q 路 FWHT 输出平均幅度分别是 $G_1\cos\theta$ 和 $G_1\sin\theta$,其中

$$G_1 = \begin{cases} 1.5G_0 & l_{\mathrm{T}}(\mathrm{mod}\,\alpha) = 0,1 \\ 2G_0 & 1 < l_{\mathrm{T}}(\mathrm{mod}\,\alpha) < \alpha - 1 \\ 0 & \text{其他} \end{cases} \tag{17}$$

式中,$l_{\mathrm{T}}(1 \leqslant l_{\mathrm{T}} \leqslant 2L)$ 为 $\boldsymbol{y}_0$ 的峰值位置,即接收信号的真实码相位. 式(17)仅仅针对直视信道,而实际信道是多径信道,并且大部分路径延时小于半个码片,即能在直视路径后的半个码片时间内进入接收机. 在这种情况下,ACF 输出与之前有所不同,然而直视仍是主要路径,所以 G_1 仍然可以按式(17)进行近似计算. 随着路径时延增加,多径信道能量呈指数衰减,即路径时延越小,信道越强,因此当没有直视路径只有延时路径时,ACF 输出仍保

北京航空航天大学学报

持三角形状,其峰值在最早到达的延时路径处出现.根据式(17),可得 G_1 的平均值 $\overline{G_1}=(2-1/\alpha)G_0$,则第 1 阶段输出判决量的噪声方差为

$$V_1\approx 2(2\alpha-1)LN_0 \tag{18}$$

第 2 阶段中 I 路和 Q 路 FWHT 输出平均幅度分别为 $G_0\cos\theta$ 和 $G_0\sin\theta$,因此 $W_{H_2}=|\boldsymbol{w}_{H_2}|^2$ 平均幅值 $\overline{G_2}=G_0$,则第 2 阶段输出判决量的噪声方差为

$$V_2\approx 2LN_0 \tag{19}$$

当集合 Λ_1 包含 l_T 时,$\boldsymbol{W}_{H_i}(i\in\{1,2\})$ 服从两自由度的非中心卡方分布,即

$$P_1^i(W_{H_i})=V_i^{-1}I_0\left(\frac{2G_i\sqrt{W_{H_i}}}{V_i}\right)\exp\left(\frac{-(W_{H_i}+G_i^2)}{V_i}\right) \tag{20}$$

式中,$I_0(\cdot)$ 是第一类零阶修正贝塞尔函数.

当集合 Λ_1 不包含 l_T 时,$\boldsymbol{W}_{H_i}(i\in\{1,2\})$ 服从两自由度的中心卡方分布,即

$$P_0^i(W_{H_i})=V_i^{-1}\exp\left(\frac{-W_{H_i}}{V_i}\right) \tag{21}$$

则第 1 阶段检测概率为

$$P_D^1=\int_{\gamma_1}^{\infty}P_1^1(y)\int_0^y P_0^1(x_{(M_1-N_{\max})})\,\mathrm{d}x\mathrm{d}y \tag{22}$$

式中,$x_{(u)}$ 为 u 阶统计值,其概率分布函数为

$$P_0^1(x_{(u)})=f_{x_{(u)}}(x)=\frac{M_1!\left[1-\exp\left(\frac{-x}{V_1}\right)\right]^{u-1}\exp\left[\frac{(u-M_1-1)x}{V_1}\right]}{V_1(u-1)!(M_1-u)!} \tag{23}$$

为了简化分析模型,在后续分析中假设 $N_{\max}=1$,即只考虑双阶段压缩捕获方法的最简单实现情况.根据文献[3]中的推论,可得第 1 阶段检测概率为

$$P_D^1=\int_{\gamma_1}^{\infty}P_1^1(y)\int_0^y P_0^1(x_{(M_1-N_{\max})})\,\mathrm{d}x\mathrm{d}y\geqslant \sum_{n=0}^{M_1-1}\frac{(-1)^n}{n+1}\binom{M_1-1}{n}\exp\left[-\frac{nG_1^2}{(n+1)V_1}\right]\cdot Q\left[\sqrt{\frac{2G_1^2}{(n+1)V_1}},\sqrt{\frac{2(n+1)\gamma_1}{V_1}}\right] \tag{24}$$

式中,$Q(a,b)$ 为马坎 Q 函数[15].

同上,第 2 阶段检测概率可以表示为

$$P_D^2=\sum_{n=0}^{M_2-1}\frac{(-1)^n}{n+1}\binom{M_2-1}{n}\exp\left(-\frac{nG_2^2}{(n+1)V_2}\right)\cdot Q\left[\sqrt{\frac{2G_2^2}{(n+1)V_2}},\sqrt{\frac{2(n+1)\gamma_2}{V_2}}\right] \tag{25}$$

式中 $M_2=\alpha$(由 $N_{\max}=1$ 易得).根据式(25)可知,由于 $M_2\ll 2L$,在同等输入信号 SNR 水平下,检测概率 P_D^2 要比传统基于并行相关方法的检测概率高.

综合式(20)、式(21)和文献[3]的推论,可得漏报概率和虚警概率分别为

$$P_M^i=\int_0^{\gamma_i}P_1^i(y)\,\mathrm{d}y\left[\int_0^{\gamma_i}P_0^i(x)\,\mathrm{d}x\right]^{M_i-1}=\left[1-\exp\left(-\frac{\gamma_i}{V_i}\right)\right]^{M_i-1}\left[1-Q\left(\sqrt{\frac{2G_i^2}{V_i}},\sqrt{\frac{2\gamma_i}{V_i}}\right)\right] \tag{26}$$

$$P_F^i=1-P_D^i-P_M^i \tag{27}$$

式中 $i(i\in\{1,2\})$ 表示第 i 阶段.上述 P_D^1,P_D^2,P_M^i 和 $P_F^i(i\in\{1,2\})$ 的前提是当前多普勒频率假设 f_d^* 与多普勒频率真实值 f_d 差异最小,即 $|f_d-f_I|<0.5\Delta f$.在较高 SNR 和较大 $\alpha(\gg 1)$ 情况下,有 $P_D^2>P_D^1$,$P_M^1\gg P_M^2$ 和 $P_F^1\gg P_F^2$,因此在后续分析中不再考虑 P_M^2 和 P_F^2 这 2 项指标.

注意,双阶段压缩捕获方法与传统基于并行相关方法的 P_D^1,P_M^1 和 P_F^1 拥有相同的表达式,只是具体参数有所不同,这和预期结果保持一致,因为 2 个阶段测量值 $\boldsymbol{y}_1$ 和 $\boldsymbol{y}_2$ 可以看作是对相邻相关器输出的有效合并.在输入信号 SNR 相同条件下,双阶段压缩捕获方法可以用较少的相关运算获得与传统基于并行相关方法相同的性能水平.

3.2 平均捕获时间

忽略很小的 FWHT 计算时间,假设第 1、第 2 阶段测量值 $\boldsymbol{y}_1,\boldsymbol{y}_2$ 的获取时间都是 $T_1=LT_c$,验证模式时间为 $T_V=N_2LT_1$,信号捕获的环形搜索状态转移流程图如图 4 所示.

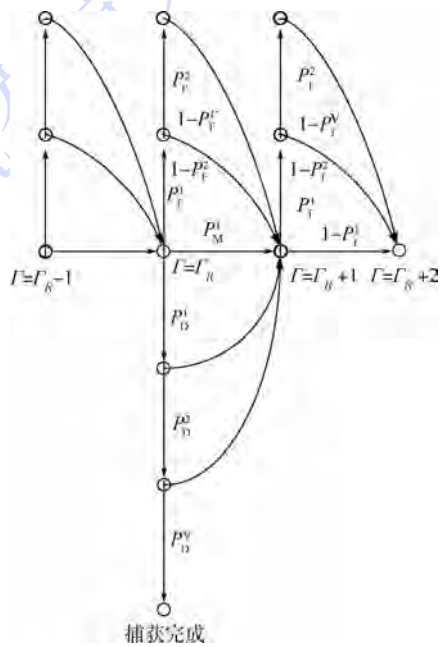


图 4 环形搜索状态转移流程图

Fig. 4 Circular search state shift diagram

如果当前多普勒频率假设 f_Γ 最接近多普勒频率真实值 f_d，第 1 阶段检测结果存在成功检测、漏报和虚警 3 种可能情况. 如果第 1 阶段检测漏报，多普勒频率假设 Γ 加 1，在下个节点重复第 1 阶段检测过程. 如果第 1 阶段成功检测或虚警，则启动第 2 阶段检测. 如果第 2 阶段成功检测，则启动验证模式. 直接正确假设检验（从节点 $\Gamma=\Gamma_R$ 到完成捕获）、正确假设漏报（从节点 $\Gamma=\Gamma_R$ 到节点 $\Gamma=\Gamma_R+1$）和非正确假设检验（从节点 Γ_i 到节点 Γ_{i+1}，且 $i\neq R$）3 种情况的传递函数分别为

$$H_D(R)=P_D^1P_D^2P_D^V R^{N_2+2} \tag{28}$$

$$H_M(R)=P_M^1R+P_F^1(1-P_F^2)R^2+P_F^1P_F^2(1-P_F^V)R^{N_2+2}+P_D^1(1-P_D^2)R^2+P_D^1P_D^2(1-P_D^V)R^{N_2+2} \tag{29}$$

$$H_0(R)=(1-P_f^1)R+P_f^1(1-P_f^2)R^2+P_f^1P_f^2(1-P_f^V)R^{N_2+1} \tag{30}$$

在实际中，验证模式函数多采用 N_1-N_2 检测器[16]，其检测概率和虚警概率分别为

$$P_D^V=Q\left(\sqrt{\frac{2G_2^2}{V_2}},\sqrt{\frac{2\gamma_2}{V_2}}\right) \tag{31}$$

$$P_F^V=\exp\left(-\frac{\gamma_2}{V_2}\right) \tag{32}$$

式中，$(\cdot)^V$ 表示验证检测模式. 注意，验证检测模式中使用的相关器与第 2 阶段的相同. 并且 $P_D^V\approx1$ 和 $P_F^V\approx P_f^V\approx0$ 可以做近似处理，式(28)～式(30)可以简化为

$$H_D(z)\approx P_D^1P_D^2z^{N_2+2} \tag{33}$$

$$H_M(R)\approx P_M^1R+(1-P_F^2+P_F^2R^{N_2})P_F^1R^2+(1-P_D^2)P_D^1R^2 \tag{34}$$

$$H_0(R)\approx(1-P_f^1)R+(1-P_f^2)P_f^1R^2+P_f^1P_f^2R^{N_2+1} \tag{35}$$

注意，在式(34)中忽略了第 1 阶段的成功检测，而第 2 阶段出现虚报、虚警及第 2 阶段成功检测，而验证模式出现虚报、虚警 2 种情况. 上述假设在一般情况下，尤其是较高 SNR 条件下，具有合理性. 根据文献[2]结论，系统传递函数可表示为

$$H(R)=\frac{H_D(R)H_0^{F_{\max}}(R)-H_D(R)}{[1-H_0(R)+\Gamma_{\max}-\Gamma_{\max}H_0(R)][H_M(R)H_0^{F_{\max}}(R)-1]} \tag{36}$$

根据系统传递函数与 MAT 相互关系 $\mu_R=T_1\left.\frac{\mathrm{d}H(R)}{\mathrm{d}R}\right|_{R=1}$，可得系统的 MAT 为

$$\mu_R=\Gamma_{\max}T_1(\Gamma_{\max}+1)^{-1}(3+2N_2+\Gamma_{\max}+\Gamma_{\max}P_f^1-P_f^1)/2+\Gamma_{\max}T_1[(\Gamma_{\max}+1)P_D^1P_D^2]^{-1}\cdot[1+P_D^1-(2+\Gamma_{\max})P_D^1P_D^2+P_F^1+N_2P_F^1P_F^2+\Gamma_{\max}(1+N_2P_f^1P_f^2)] \tag{37}$$

当 SNR 较低时，$P_D^1P_D^2$ 降低，式(37)的第 2 项会增加，从而导致平均捕获时间 μ_R 增加. 当信噪 SNR 较高时，式(37)可以简化为

$$\mu_R\approx[2+N_2+\Gamma_{\max}/2]T_1 \tag{38}$$

式(38)表明，MAT 与第 1、第 2 阶段测量值 $\boldsymbol{y}_1$，$\boldsymbol{y}_2$ 的获取时间、验证模式时间和多普勒频率搜索容限等因素成正比.

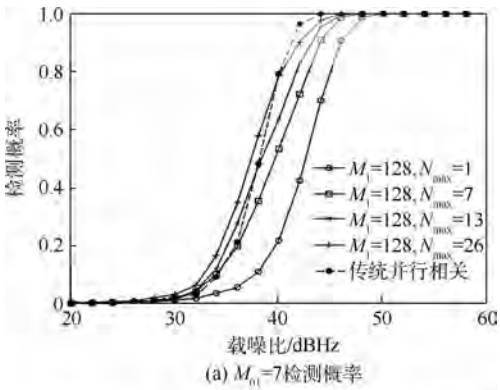
4　仿真校验

对双阶段压缩捕获方法进行蒙特卡罗实验验证，主要思路是从检测概率和平均捕获时间出发，对比分析双阶段压缩捕获方法和传统基于并行的相关方法的捕获性能. 仿真实验参数设置如下：考虑 BPSK 调制 DSSS 信号，数据率 $R_b=1$ kb/s，码速率 $R_c=10$ MHz，采样率 $f_s=2R_c$，多普勒频率 $f_d=100$ Hz，码相位 $\tau=1\,000$，相关处理时间 1 ms，即 $2L=2\,046$ 个采样值，第 1 和第 2 阶段捕获门限分别为 $\gamma_1=1.2$ 和 $\gamma_2=0.75$[3].

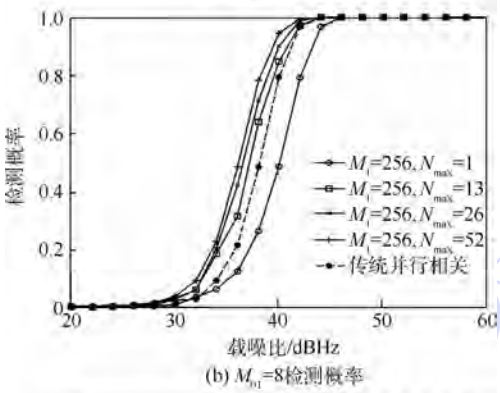
4.1　检测概率

在不同测量值个数情况下，通过蒙特卡洛仿真对比分析双阶段压缩捕获方法与传统基于并行相关方法的检测概率，其中载噪比 C/N_0 变化范围取[20,60] dBHz，测量矩阵 $\boldsymbol{\Phi}_1$ 与 $\boldsymbol{\Phi}_2$ 构造参数分别为 $M_{01}\in\{7,8,9,10\}$ 和 $M_{02}=\mathrm{lb}(2L/M_1)$，则有第 1 阶段测量值个数 $M_1\in\{128,256,512,1\,024\}$，第 1 阶段超过门限的最大值个数 $N_{\max}\in\{1,\lceil0.05M_1\rceil,\lceil0.1M_1\rceil,\lceil0.2M_1\rceil\}$，检测概率对比如图 5 所示.

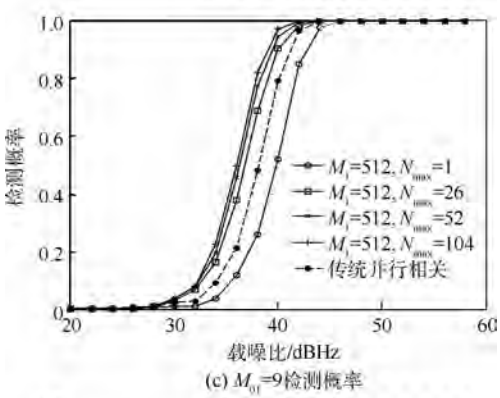
由图 5 可知，仿真结果与理论分析结果基本吻合，检测概率随着第 1 阶段测量值个数 M_1 和超过门限的最大值个数 $N_{\max}$ 增加而增大，并且最终趋向于一个稳定值. 当 $M_1=256$ 且 $N_{\max}=13$ 时，双阶段压缩捕获方法捕获性能与传统基于并行相关方法的基本持平，且有 $\alpha\approx2L/M_1=8$，第 1 和第 2 阶段的相关次数分别为 256 和 $\alpha N_{\max}=104$，即双阶段压缩捕获方法相关次数总共为 360，而传统基于并行相关的方法相关次数为 $2L=2\,046$，可知此时双阶段压缩捕获方法相关次数相对于传统基于并行相关的方法几乎降低了 6 倍，说明双阶段压缩捕获算法在显著降低相关运算量



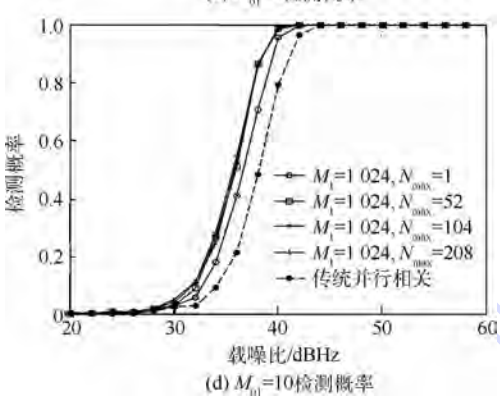
(a) M_{01}=7检测概率



(b) M_{01}=8检测概率



(c) M_{01}=9检测概率



(d) M_{01}=10检测概率

图 5 检测概率对比

Fig. 5 Comparison of detection probability

的前提下,能够达到传统基于并行相关捕获方法的检测概率水平.

4.2 平均捕获时间

实验条件设置与检测概率实验条件设置基本相同,测量矩阵 $\boldsymbol{\Phi}_1$ 构造参数取 $M_{01} \in \{7,8,9\}$,即第 1 阶段测量值个数 $M_1 \in \{128,256,512\}$,第 1 阶段超过门限的最大值个数 $N_{max}=4\alpha$,测量值获取时间 $T_1=1$ ms, $N_2=10$, $\Gamma_{max}=20$,平均捕获时间对比关系如图 6 所示.

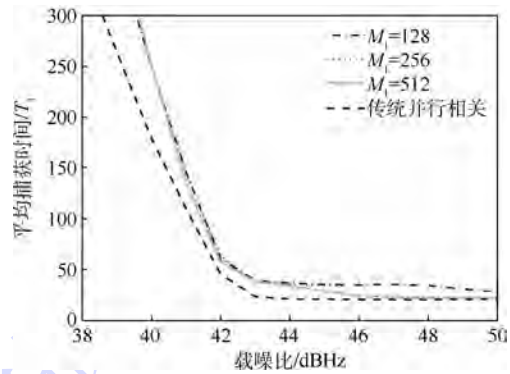


图 6 平均捕获时间对比关系

Fig. 6 Comparison of MAT

由图 6 可知,在载噪比较高(>42 dBHz)时,双阶段压缩捕获方法平均捕获时间趋向于一个稳定的最小值 $22T_1$,该最小值与式(38)得到的理论值 $2T_1+10T_1+(20/2)T_1$ 相一致;在载噪比较低时,双阶段压缩捕获方法平均捕获时间随着载噪比的降低而快速增加,这也与 3.2 节的理论分析保持一致,对应于式(37)的第 2 项,当载噪比较低时,分母 $P_D^1P_D^2$ 降低,分子 $1+P_D^1-(2+20)\cdot P_D^1P_D^2+P_F^1+10P_F^1P_F^2+20(1+10P_f^1P_f^2)$ 增高,导致平均捕获时间增加. 值得注意的是,双阶段压缩捕获方法平均捕获时间的变化趋势与基于传统并行相关方法的保持一致,而且第 1 阶段测量值个数 M_1 越大,双阶段压缩捕获方法平均捕获时间越能够较快地接近基于传统并行相关方法的平均捕获时间.

5 结 论

基于压缩感知理论,提出了一种双阶段压缩捕获方法,并进行了理论分析与实验对比验证,主要结论有:

1) 在显著降低相关运算量的前提下,该方法能够达到传统基于并行相关捕获方法的检测概率水平.

2) 该方法平均捕获时间的变化趋势与基于传统并行相关方法的保持一致,而且第 1 阶段测量值个数越大,双阶段压缩捕获方法平均捕获时间越能够较快地接近基于传统并行相关方法的平均捕获时间.

3) 双阶段压缩捕获方法能够有效利用 DSSS 接收信号在自相关域的稀疏性,可以有效解决 DSSS 信号在捕获过程中难以同时实现的高检测性能、低硬件消耗或计算量问题.

参考文献(References)

[1] 田日才. 扩频通信[M]. 北京:清华大学出版社,2006:1-26.
Tian R C. Spread spectrum communication[M]. Beijing:Tsinghua University Press,2006:1-26(in Chinese).

[2] Huang Z J, Zhou J, Chen B, et al. A novel two-dimensional space-time processing scheme for DS/CDMA code acquisition [C]//International Conference on Signal Processing, Communication and Computing. Kunming:IEEE,2013:1-6.

[3] Myeongsu H, Takki Y, Kang C, et al. A new adaptive code-acquisition algorithm using parallel sub filter structure[J]. IEEE Transactions on Vehicular Technology, 2006, 55(6): 1790-1796.

[4] Wu H B, Hu Y H. Fast acquisition of GPS signal in low SNR environments[C]//The 2nd International Conference on Information Science and Engineering. Hangzhou: IEEE, 2010: 4534-4537.

[5] Liu S X, Feng W Q, Xiong H G, et al. A FFT based long-code acquisition algorithm suitable for FPGA implementation[C]// 2011 International Conference on Electric Information and Control Engineering. Wuhan:IEEE,2011:5621-5624.

[6] Hong L. Research on global positioning system M-code acquisition method and the acquisition performance[J]. IET Communications,2014,8(5):587-596.

[7] Tian M H, Feng Y X, Liu F. On the scheme of rapid acquisition of long PN code in DSSS Signals[C]//The 5th International Conference on. Kunming:IEEE,2009:457-460.

[8] Li H, Cui X, Lu M, et al. Dual-folding-based rapid search method for long PN-code acquisition[J]. IEEE Transactions on Wireless Communication,2008,7(12):5286-5296.

[9] Feng W Q. Dual-channel method for fast long PN-code acquisition[J]. Communications, China,2014,11(5):60-70.

[10] Candes E J, Tao T. Decoding by linear programming[J]. IEEE Transactions on Information Theory,2005,51(12):4203-4215.

[11] Donoho D L. Compressive sensing[J]. IEEE Transactions on Information Theory,2006,52(4):1289-1306.

[12] Candes E J, Wakin M B. An introduction to compressive sampling[J]. IEEE Signal Processing Magazine, 2008, 25(2): 21-30.

[13] Calderbank R, Howard S, Jafarpour S. Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property[J]. Selected Topics in Signal Processing, 2010,4(2):358-374.

[14] Ouyang W L. Fast algorithm for walsh hadamard transform on sliding windows[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2010,32(1):165-171.

[15] Sofotasios P C, Freear S. Novel expressions for the Marcum and one dimensional Q-functions[C]//7th International Symposium on Wireless Communication Systems. York: IEEE, 2010: 736-740.

[16] Ward P W. GPS receiver search techniques[C]//IEEE Position Location and Navigation Symposium. Atlanta:IEEE,1996: 604-611.

Two-stage acquisition algorithm for DSSS signal based on compressive sensing

CHENG Yanhe, YANG Wenge*

(Department of Optical and Electrical Equipment, Equipment Academy, Beijing 101416, China)

Abstract: According to the restriction between the acquisition performance and the hardware consumption, a two-stage compressive acquisition method was proposed based on compressive sensing, specifically some coarse acquisition values can be got by the first stage, following by accurately detecting the signal in the values by the second stage. Firstly, the sparsity of DSSS signal in the correlation domain was explored, so as to build the sparsity transformation matrix. Secondly, the measurement matrixes and detection algorithms in both stages were produced according to the deterministic Walsh-Hadamard matrix. Finally, the acquisition performance of the provided method was analyzed theoretically in terms of the detection probability and the mean acquisition time, which was also verified by Monte Carle method. The theoretical analysis and the simulation results show that the novel method can use much less correlations to achieve the same performance as conventional parallel correlation-based methods.

Key words: direct sequence spread spectrum signal; compressive sensing; measurment matrix; detection probability; mean acquisition time

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0255

过渡状态下材料断裂韧性的计算方法

高潮，何宇廷*，崔荣洪，伍黎明

（空军工程大学 航空航天工程学院，西安 710038）

摘 要：材料的断裂韧性对于损伤容限设计和评估技术非常重要，但采用传统的试验方法获得过渡状态下材料的断裂韧性需要大量的时间和费用. 针对这些不足，提出了一种过渡状态下材料断裂韧性的计算方法. 分析了三维裂纹尖端的三轴应力约束参数，利用材料断裂能量准则研究了断裂韧性与拉伸性能之间的关系，得到了在三向应力状态下Ⅰ型裂纹体断裂破坏时的断裂韧性与材料单向拉伸性能参数之间的关系表达式和过渡状态下材料断裂韧性的计算方法. 计算了2219-T87铝合金在过渡状态下的断裂韧性理论值，并将其与试验结果和经验公式计算值进行比较. 结果显示，理论计算值处于试验点范围内，小于经验公式计算值，且偏于安全.

关 键 词：断裂韧性；过渡状态；平面应变；三轴应力约束；裂纹尖端
中图分类号：O346.1
文献标识码：A **文章编号**：1001-5965(2015)04-0632-09

通常情况下材料断裂韧性被看作常数，为平面应变状态下的断裂韧性值. 实际上，断裂韧性的值是随着试样厚度的变化而变化的，即断裂韧性是一个与应力状态有关的量，并不是仅与材料性质有关的常数. 在一些航空技术先进国家，已经通过大量的试验给出了许多常用材料的K_C-B曲线，即断裂韧性-厚度曲线，而中国基本没有建立航空结构材料的K_C-B曲线. 要解决这一问题，一是通过像先进国家那样通过大量的试验来确定K_C-B曲线，但将耗费巨额资金；二是由三维断裂理论入手，分析断裂的本质，进而提出不依赖于厚度的断裂韧性计算方法，得出材料的K_C-B曲线.

国内外有不少学者已对材料断裂韧性进行了研究，并取得了显著的成果. 文献[1-13]采用数值仿真或理论计算方法对平面应力或平面应变状态下材料的断裂韧性进行了分析，但没有涉及到过渡状态下材料断裂韧性的计算方法. 文献[14]引入了厚度修正经验函数，但所得过渡状态下的断裂韧性是经验修正的结果，模型的通用性受到限制. 杨继运等[15-18]建立了材料断裂韧性与试样厚度之间的计算模型，但在计算材料断裂韧性时，首先需要试验数据来确定模型中的参数值. 可见，国内外对过渡状态下材料断裂韧性的计算方法研究较少，而过渡状态下材料的断裂韧性值不可能通过试验来一一确定，所以迫切需要一种能计算过渡状态下材料断裂韧性的方法.

本文在已有研究成果的基础上提出一种过渡状态下材料断裂韧性的计算方法. 对三维裂纹尖端的三轴应力约束参数进行分析，通过断裂能量准则对断裂韧性与材料拉伸性能之间的关系进行研究，得到过渡状态下材料断裂韧性的计算方法，同时结合试验值对这一方法进行了验证.

收稿日期：2014-05-08；**录用日期**：2014-06-25；**网络出版时间**：2014-09-05 15:56
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0255.html
基金项目：国家自然科学基金资助项目(51201182)
作者简介：高潮(1986—)，男，陕西杨陵人，博士生，gaochao19861105@sina.com
*通讯作者：何宇廷(1966—)，男，四川阆中人，教授，hyt666@tom.com，主要研究方向为飞机结构可靠性与安全.

引用格式：高潮，何宇廷，崔荣洪，等. 过渡状态下材料断裂韧性的计算方法[J]. 北京航空航天大学学报，2015，41(4)：632-640.
Gao C，He Y T，Cui R H，et al. Calculation method of fracture toughness of materials at transition state[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：632-640(in Chinese).

1　三维裂纹尖端的三轴应力约束参数

传统的疲劳断裂理论往往是基于单轴试样的试验数据或应力强度参数进行分析的，此时应力状态的影响就以约束的概念来体现，而三轴应力约束是造成三维疲劳裂纹尖端场与二维裂纹尖端场不同的根本原因.

穿透厚度裂纹坐标系如图 1 所示，令 O 点为三维裂纹前缘曲线上的任意一点，x，y 和 z 分别为 O 点处曲线的主法向、副法向和切向分量，xOy 为 O 点处裂纹前缘曲线的法平面，r 和 θ 为该平面内的极坐标.

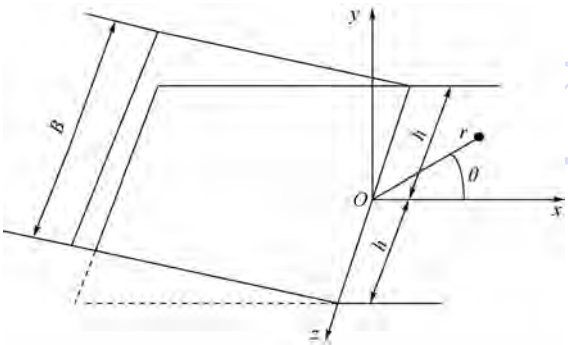


图 1　穿透厚度裂纹坐标系

Fig. 1　Coordinate system of through-wall crack

对于典型的穿透直裂纹，可将 O 点固定于试件中心面上，定义三轴应力约束参数 T_z 为

$$T_z=\frac{\sigma_z}{\sigma_r+\sigma_\theta} \tag{1}$$

式中，σ_z、σ_r 和 σ_θ 分别为极坐标下沿 z 方向、r 方向和 θ 方向的应力.

在弹性范围内，当 $T_z=0$ 时，结构处于平面应力状态；当 $T_z=\nu$（ν 为泊松比）时，结构处于平面应变状态；当 $0<T_z<\nu$ 时，结构处于过渡状态.

对于三轴应力约束参数 T_z，目前还没有理论上的解析表达显式. 文献[19]根据有限元数值仿真结果提出了穿透型裂纹的 T_z 拟合表达式：

$$T_z=\frac{1}{2}\left[1-(1-2\nu)\left(\frac{r}{\bar{r}_p}\right)^{\frac{n-1}{2.3n+1}}\right]\times$$
$$\left[1-1.218\left(\frac{r}{B}\right)^{1/2}-0.359\frac{r}{B}+0.361\left(\frac{r}{B}\right)^{3/2}\right]\times$$
$$\left[1-\left|\frac{z}{h}\right|^{0.94\left(\frac{r}{B}\right)^{-0.58}}\right]^2 \tag{2}$$

式中，$\bar{r}_p$ 为 $\theta=0°$ 时裂纹尖端穿透厚度的平均塑性区尺寸；h 为厚度 B 的一半；n 为材料的 Ramberg-Osgood 硬化指数.

2　材料断裂韧性的本质描述

三维裂纹体断裂的能量准则为

$$\begin{cases}G=R\\ \dfrac{\partial G}{\partial a}=\dfrac{\partial R}{\partial a}\end{cases} \tag{3}$$

式中，G 为能量释放率；R 为裂纹扩展阻力；a 为裂纹长度.

上述能量准则也是发生断裂的充要条件. 根据光滑试样的拉伸性能参数可以给出材料的断裂结果（裂纹失稳扩展的结果），因此，可用材料的拉伸性能参数（包括最终断裂时的参数）来分析裂纹的失稳扩展.

2.1　基本关系式

带裂纹结构开裂的条件[20]为

$$G=2\gamma+\frac{\partial U_F}{\partial a}+\frac{\partial U_U}{\partial a} \tag{4}$$

式中，$\dfrac{\partial U_F}{\partial a}$ 为裂纹尖端过程区消耗的塑性功率；$\dfrac{\partial U_U}{\partial a}$ 为裂纹尖端过程区以外的塑性区消耗的塑性功率；2γ 为表面能.

对于Ⅰ型裂纹体，能量释放率与应力强度因子的关系[21]为

$$G=\frac{1-\nu T_z}{E}K_I^2 \tag{5}$$

式中，K_I 为应力强度因子；E 为弹性模量.

则有

$$\frac{1-\nu T_z}{E}K_I^2=2\gamma+\frac{\partial U_F}{\partial a}+\frac{\partial U_U}{\partial a} \tag{6}$$

在裂纹尖端附近，过程区中为非连续介质，有着长大并联合的孔洞，这与光滑试样拉伸时颈缩开始后的变形情况在本质上是相同的. 过程区以外塑性区的变形情况与光滑试样拉伸时从屈服延伸率 δ_s 到均匀延伸率 δ_b 段的均匀塑性变形情况在本质上是相同的. 裂纹尖端前方弹性区的变形情况与光滑试样拉伸时弹性段的变形情况在本质上也是相同的. 因此，可参考文献[14]中的方法，将裂纹尖端场等效为单向应力和应变场：

$$\frac{\partial U_F}{\partial a}=W_F h_F \tag{7}$$

$$\frac{\partial U_U}{\partial a}=W_U h_U \tag{8}$$

式中：W_F 为单向拉伸应力-应变曲线下应力值 σ 大于抗拉强度 σ_b 的面积，代表颈缩后材料不可恢复的真实塑性应变能密度；h_F 为裂纹尖端过程区

北航学报 赠阅

的当量尺寸；W_U 为单向拉伸应力-应变曲线下屈服应力 σ_s 至 σ_b 之间的面积，代表连续、均匀塑性变形部分的应变能密度；h_U 为连续塑性区的当量尺寸.

研究对象为处于某一应力状态下带裂纹的单位厚度板中的情况. 典型的应力-应变曲线如图 2 所示，图 3 为裂纹尖端区域示意图.

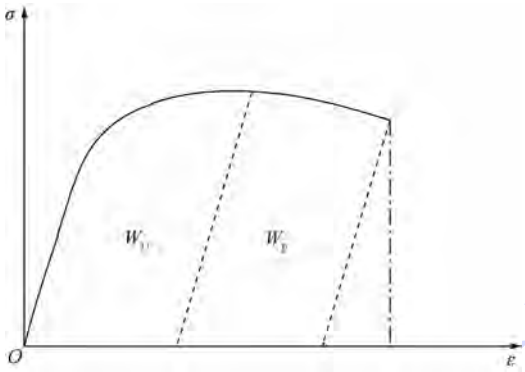


图 2 典型的应力-应变曲线

Fig. 2 Typical stress-strain curve

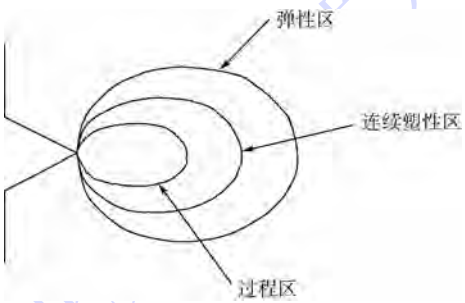


图 3 裂纹尖端区域示意图

Fig. 3 Illustration of crack tip regions

根据已有研究[14]，裂纹扩展单位面积所需表面能可近似表示为

$$2\gamma = Ea_0/20 \tag{9}$$

式中，a_0 为原子直径.

所以，式(6)可表示为

$$\frac{1-\nu T_z}{E}K_I^2 = \frac{Ea_0}{20} + W_F h_F + W_U h_U \tag{10}$$

式(10)即为含 I 型裂纹结构处于某三向应力状态下的开裂条件. 当 h_F 和 h_U 取为结构断裂时的值时，式(10)也是结构的断裂条件，K_I 就变为平面应力状态下的材料断裂韧性 K_{C1}.

2.2 $W_F h_F$ 的表征

2.2.1 平面应力状态下 $W_F h_F$ 的表征

h_F 为裂纹尖端过程区的有效当量尺寸，其最小值可认为等于与裂纹面成 45°方向的滑移带有效尺寸. 由于已有研究者测定在室温及高应力作用下滑移带沿裂纹面方向的宽度 ω 约为 10 μm[14]，所以有

$$h = \omega/\sin 45° = 14.142\ \mu m \tag{11}$$

对于有较大颈缩应变的金属材料，于平面应力状态下断裂时，h_F 可表示为[14]

$$h_F = \frac{8W_F E^2 a_0}{\pi\sigma_b^2 \sin 45°} \tag{12}$$

根据已有文献，W_F 可表示为[20]

$$W_F = \frac{1}{2}(\sigma_b + \sigma_K)(\varepsilon_f - \varepsilon_b) \tag{13}$$

式中，σ_K 为断裂应力；ε_b 为屈服应变；ε_f 为真实断裂应变.

2.2.2 三向应力状态下 $W_F h_F$ 的表征

对于一般的三向应力状态，可认为结构断裂时裂纹尖端过程区的当量尺寸 h'_F 为平面应力状态时的 k 倍，则有

$$h'_F = kh_F \tag{14}$$

对于 k 可作如下分析. 当裂纹尖端应力应变场以弹性解表征时，其对应的等效单向应力达 σ_b，对应点距离裂纹尖端的距离 r_b 为模拟过程区尺寸. 设平面应力状态下结构断裂时 r_b 为 r_{b1}，非平面应力状态下结构断裂时 r_b 为 r_{b2}，根据相似性原理，可取 r_{b2}/r_{b1} 为 k. 在三向应力状态下，任何主应力分量超过材料的真实断裂应力 σ_f 时，材料将发生局部开裂，从而进入非连续的过程区.

设各向同性材料有一立方单元体受主应力系数 σ_1，σ_2 和 σ_3 作用. 可定义应力状态参数：

$$\begin{cases}\alpha_1 = \sigma_1/\sigma_1 = 1\\ \alpha_2 = \sigma_2/\sigma_1\\ \alpha_3 = \sigma_3/\sigma_1\end{cases} \tag{15}$$

由广义虎克定律可得出弹性应变为

$$\varepsilon_{1e} = \frac{\sigma_1}{E}[\alpha_1 - \nu(\alpha_2 + \alpha_3)] \tag{16}$$

由弹塑性理论可得八面体剪应力为

$$\tau_{oct} = \sqrt{(\sigma_1-\sigma_2)^2 + (\sigma_2+\sigma_3)^2 + (\sigma_3-\sigma_1)^2}/3 = \sqrt{2}\sigma_1[\alpha_1^2 + \alpha_2^2 + \alpha_3^2 - (\alpha_2 + \alpha_2\alpha_3 + \alpha_3)]^{\frac{1}{2}}/3 \tag{17}$$

当 $\alpha_2 = \alpha_3 = 0$ 时，单向应力状态下八面体剪应力为

$$(\tau_{oct})_U = (\sqrt{2}/3)\sigma \tag{18}$$

从而可得

$$\sigma_1 = \sigma[\alpha_1^2 + \alpha_2^2 + \alpha_3^2 - (\alpha_2 + \alpha_2\alpha_3 + \alpha_3)]^{-\frac{1}{2}} \tag{19}$$

式(19)反映了三向应力状态下主应力 σ_1 与 σ 及应力状态参数 σ_1，σ_2 和 σ_3 之间的关系.

实际上，裂纹前缘由结构表面到中心部位的三轴应力约束参数 T_z 由小变大，对于Ⅰ型裂纹，在裂纹面上的裂纹尖端前缘某点（$\theta=0°$）处，有

$$\begin{cases}\sigma_1=\sigma_y\\ \sigma_2=\sigma_x=\sigma_1\\ \sigma_3=2T_z\sigma_1\end{cases}\tag{20}$$

式中，σ_x 和 σ_y 分别为沿 x 方向和 y 方向的应力.

根据式（15）和式（20）可得：

$$\begin{cases}\alpha_1=\alpha_2=1\\ \alpha_3=2T_z\end{cases}\tag{21}$$

当等效单向应力为 σ_b 时，根据式（19）可得：

$$\begin{cases}\sigma_{1b}=\sigma_b[\alpha_1^2+\alpha_2^2+\alpha_3^2-(\alpha_2+\alpha_2\alpha_3+\alpha_3)]^{-\frac{1}{2}}=\dfrac{\sigma_b}{1-2T_z} & \sigma_b\leqslant\sigma_f\\ \sigma_{1b}=\sigma_f & \sigma_b>\sigma_f\end{cases}\tag{22}$$

式中，σ_{1b} 为三向应力状态主应力.

当用 Ramberg-Osgood 模型[22]表征单向应力-应变关系时，σ_{1b} 对应的应变为

$$\varepsilon_{1b}=\frac{\sigma_{1b}(1-\nu-2\nu T_z)}{E}+\frac{1}{2}(1-2T_z)^n\left(\frac{\sigma_{1b}}{K}\right)^n\tag{23}$$

式中，K 为强化系数.

同时，裂纹尖端某点（$\theta=0°$）在 y 方向上的应变（弹性解）为

$$\varepsilon_y=\frac{1}{E}\cdot\frac{K_I}{\sqrt{2\pi r}}(1-\nu-2\nu T_z)\tag{24}$$

令 $\varepsilon_y=\varepsilon_{1b}$，可得：

$$r_b=\frac{1}{2\pi}\left[\frac{1-\nu-2\nu T_z}{E}\cdot\frac{K_I}{\frac{\sigma_{1b}(1-\nu-2\nu T_z)}{E}+\frac{1}{2}(1-2T_z)^n\left(\frac{\sigma_{1b}}{K}\right)^n}\right]^2\tag{25}$$

当结构于平面应力状态下发生断裂时，式（25）中 $T_z=0$，$K_I=K_C$，$r_b=r_{b1}$，所以有

$$r_{b1}=\frac{1}{2\pi}\left[\frac{1-\nu}{E}\cdot\frac{K_C}{\frac{\sigma_b(1-\nu)}{E}+\frac{1}{2}\left(\frac{\sigma_b}{K}\right)^n}\right]^2\tag{26}$$

当结构于三向应力状态下发生断裂时，设 $K_I=K_{C1}$，$r_b=r_{b2}$，则有

$$r_{b2}=\frac{1}{2\pi}\left[\frac{1-\nu-2\nu T_z}{E}\cdot\frac{K_{C1}}{\frac{\sigma_{1b}(1-\nu-2\nu T_z)}{E}+\frac{1}{2}(1-2T_z)^n\left(\frac{\sigma_{1b}}{K}\right)^n}\right]^2\tag{27}$$

由式（26）和式（27）可得：

$$k=\frac{r_{b2}}{r_{b1}}=\left[\frac{1-\nu-2\nu T_z}{1-\nu}\cdot\frac{\frac{\sigma_b(1-\nu)}{E}+\frac{1}{2}\left(\frac{\sigma_b}{K}\right)^n}{\frac{\sigma_{1b}(1-\nu-2\nu T_z)}{E}+\frac{1}{2}(1-2T_z)^n\left(\frac{\sigma_{1b}}{K}\right)^n}\cdot\frac{K_{C1}}{K_C}\right]^2\tag{28}$$

根据式（22）和式（28）即可确定 k.

如果定义 T_{z0} 为

$$T_{z0}=\frac{1}{2}\left(1-\frac{\sigma_b}{\sigma_f}\right)\tag{29}$$

则当 $T_z\geqslant T_{z0}$ 时，由式（20）和式（26）可得：

$$k=\left[\frac{1-\nu-2\nu T_z}{1-\nu}\cdot\frac{\frac{\sigma_b(1-\nu)}{E}+\frac{1}{2}\left(\frac{\sigma_b}{K}\right)^n}{\frac{\sigma_f(1-\nu-2\nu T_z)}{E}+\frac{1}{2}(1-2T_z)^n\left(\frac{\sigma_f}{K}\right)^n}\cdot\frac{K_{C1}}{K_C}\right]^2\tag{30}$$

当结构为平面应变状态时，式（28）和式（30）中的 $K_{C1}=K_{IC}$，K_{IC} 为平面应变状态下的材料断裂韧性，$T_z=\nu$；当结构为平面应力状态时，则 $T_z=0$，$K_{C1}=K_C$，$k=1$.

在求得 k 后，由式（14）即可确定三向应力状态下结构断裂时裂纹尖端过程区的当量尺寸 h_F'. 表示颈缩后材料的真实应变能密度 W_F 可按式（13）计算.

2.3　W_Uh_U 的表征

由式（8）可知，W_Uh_U 可直接从 $\partial U_U/\partial a$ 的推导求得. 采用裂纹尖端应力应变场的弹性解表达式进行推导，并在裂纹尖端相应的半径尺寸上用试验结果加以修正，从而得出 $\partial U_U/\partial a$ 的表达式. 同时，采用八面体剪应力与应力状态无关的特性来联系单行应力状态（单向拉伸应力）与三向应力状态（裂纹尖端场）.

将Ⅰ型裂纹尖端场表示成极坐标系下的

北航学报 赠阅

形式[20]：

$$\begin{cases}\sigma_r=\dfrac{K_{\mathrm{I}}}{\sqrt{2\pi r}}f_1(\theta)\\ \sigma_\theta=\dfrac{K_{\mathrm{I}}}{\sqrt{2\pi r}}f_2(\theta)\\ \sigma_z=T_z[f_1(\theta)+f_2(\theta)]\dfrac{K_{\mathrm{I}}}{\sqrt{2\pi r}}\\ \tau_{r\theta}=\dfrac{K_{\mathrm{I}}}{\sqrt{2\pi r}}f_3(\theta)\end{cases}\tag{31}$$

其中

$$\begin{cases}f_1(\theta)=\dfrac{5}{4}\cos\dfrac{\theta}{2}-\dfrac{1}{4}\cos\dfrac{3\theta}{2}\\ f_2(\theta)=\dfrac{3}{4}\cos\dfrac{\theta}{2}+\dfrac{1}{4}\cos\dfrac{3\theta}{2}\\ f_3(\theta)=\dfrac{1}{4}\sin\dfrac{\theta}{2}+\dfrac{1}{4}\sin\dfrac{3\theta}{2}\end{cases}$$

式中，$\tau_{r\theta}$为 r 与 θ 面的剪应力.

由弹塑性理论可知，八面体剪应力为

$$\tau_{\mathrm{oct}}=\frac{1}{3}\times\left[\sqrt{(\sigma_r-\sigma_\theta)^2+(\sigma_\theta-\sigma_z)^2+(\sigma_z-\sigma_r)^2}+6(\tau_{r\theta}^2+\tau_{rz}^2+\tau_{\theta z}^2)\right]^{\frac{1}{2}}\tag{32}$$

式中，τ_{rz}为 r 与 z 面的剪应力；$\tau_{\theta z}$为 z 与 θ 面的剪应力.

将式(31)代入式(32)，可得

$$\tau_{\mathrm{oct}}=\frac{K_{\mathrm{I}}}{3\sqrt{\pi r}}[f_{\mathrm{e}}(\theta)]^{\frac{1}{2}}\tag{33}$$

式中

$$f_{\mathrm{e}}(\theta)=(T_z^2-T_z+1)f_1^2(\theta)+f_1(\theta)f_2(\theta)(2T_z^2-2T_z-1)+(T_z^2-T_z+1)f_2^2(\theta)+3f_3^2(\theta)$$

令 $\psi(\theta)=[f_{\mathrm{e}}(\theta)]^{\frac{1}{2}}$，则式(33)可表示为

$$\tau_{\mathrm{oct}}=\frac{K_{\mathrm{I}}}{3\sqrt{\pi r}}\psi(\theta)\tag{34}$$

对于单向应力状态(单向拉伸载荷作用)，有

$$\tau_{\mathrm{oct}}=\frac{\sqrt{2}}{3}\sigma_{\mathrm{T}}\tag{35}$$

式中，σ_{T} 为单向拉伸真实应力.

令式(34)和式(35)相等，可得裂纹尖端应力场与单向拉伸应力之间的关系表达式：

$$\sigma_{\mathrm{T}}=\frac{K_{\mathrm{I}}}{2\sqrt{\pi r}}\psi(\theta)\tag{36}$$

在裂纹尖端取单位厚度板，图 4 为裂纹尖端塑性区示意图. σ_{Tm}为在塑性区中 σ_{T} 取极大值的线，假设 σ_{Tm}线上有任意一点 P 及其前方一点 P'，当裂纹扩展长度为 da 时，σ_{Tm}线将移到 σ'_{Tm}位置. 假设 P'在 σ'_{Tm}线上，这时 P 点将进入卸载区，其卸载按弹性规律进行. σ'_{Tm}线右方各点与裂纹未扩展时 σ_{Tm}线右方各点应力应变场的分布相同. 因此，在裂纹扩展长度延长至 da 的过程中，仅用 σ_{Tm}线与 σ'_{Tm}线之间的材料参数计算塑性功率即可，弹性卸载可略去不计.

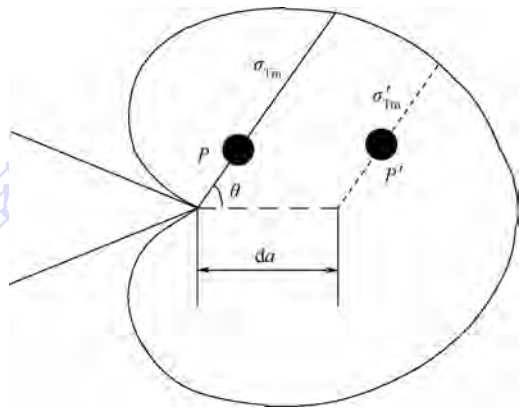


图 4 裂纹尖端塑性区示意图

Fig. 4 Plastic zone around crack tip

对于金属材料，其应力-塑性应变曲线可由 Ramberg-Osgood 模型表示为

$$\varepsilon_{\mathrm{Tp}}=\varepsilon_{\mathrm{pb}}\left(\frac{\sigma_{\mathrm{T}}}{\sigma_{\mathrm{b}}}\right)^n\tag{37}$$

式中，$\varepsilon_{\mathrm{Tp}}$为 σ_{T} 对应的塑性应变；$\varepsilon_{\mathrm{pb}}$为 σ_{b} 对应的塑性应变. Ramberg-Osgood 硬化指数 n 为

$$n=\frac{\ln\dfrac{\varepsilon_{\mathrm{pb}}}{\varepsilon_{\mathrm{ps}}}}{\ln\dfrac{\sigma_{\mathrm{b}}}{\sigma_{\mathrm{s}}}}\tag{38}$$

式中，$\varepsilon_{\mathrm{ps}}$为屈服应力 σ_{s} 对应的塑性应变，通常可取为 0.2%.

当塑性应变较大时，式(37)可近似表示为

$$\varepsilon_{\mathrm{T}}=\varepsilon_{\mathrm{b}}\left(\frac{\sigma_{\mathrm{T}}}{\sigma_{\mathrm{b}}}\right)^n\tag{39}$$

式中，ε_{T} 为 σ_{T} 对应的应变. 塑性应变能密度可近似表示为

$$\Phi_{\mathrm{p}}=\int_0^{\sigma_{\mathrm{T}}}\sigma_{\mathrm{T}}\,\mathrm{d}\varepsilon_{\mathrm{T}}=\frac{n}{n+1}\sigma_{\mathrm{T}}\varepsilon_{\mathrm{T}}\tag{40}$$

所以，图 4 中裂纹扩展长度为 da 时，σ_{Tm}线与 σ'_{Tm}线之间材料的塑性应变能为

$$\mathrm{d}U_{\mathrm{p1}}=\iint_A\Phi_{\mathrm{p}}\mathrm{d}A\tag{41}$$

式中，A 为 σ_{Tm}线与 σ'_{Tm}线所围成的塑性区面积.

由于裂纹尖端扩展长度为 da 时，σ_{Tm}线与 σ'_{Tm}线是沿着 $\theta=0°$线上下对称的，所以当裂纹扩展

长度为 da 时，有

$$\mathrm{d}U_{\mathrm{p}}=2\iint_{A}\Phi_{\mathrm{p}}\mathrm{d}A \tag{42}$$

式中，dU_{p} 为塑性应变能. 式(42)可进一步表示为

$$\frac{\partial U_{\mathrm{p}}}{\partial a}=2\int_{r_{\mathrm{u}}}^{r}\frac{n}{n+1}\sigma_{\mathrm{T}}\varepsilon_{\mathrm{T}}\sin\theta\,\mathrm{d}r \tag{43}$$

式中，r_{u} 为对应高应力区的边界.

因为有[14]

$$\frac{\varepsilon_{\mathrm{T}}}{\varepsilon_{\mathrm{b}}}=\sqrt{\frac{r_{\mathrm{u}}}{r}} \tag{44}$$

利用式(44)对式(43)进行积分，可得：

$$\frac{\partial U_{\mathrm{p}}}{\partial a}=\frac{2n^{2}}{n^{2}-1}\sigma_{\mathrm{b}}\varepsilon_{\mathrm{b}}2r_{\mathrm{u}}\sin\theta\left[\left(\frac{\varepsilon_{\mathrm{b}}}{\varepsilon_{\mathrm{T}}}\right)^{\frac{n-1}{n}}-1\right] \tag{45}$$

这里 r_{u} 采用试验数据进行修正. 对于中心穿透裂纹的无限大板[14]，有

$$2r_{\mathrm{u}}=\frac{\varepsilon_{\mathrm{Tf}}\varepsilon_{\mathrm{s}}}{\varepsilon_{\mathrm{b}}\varepsilon_{\mathrm{T}}}h\left[1-\left(\frac{\sigma_{\mathrm{T}}}{\sigma_{\mathrm{b}}}\right)^{n+1}\right]\psi^{2}(\theta) \tag{46}$$

式中，$\varepsilon_{\mathrm{Tf}}$为单向拉伸应力-应变曲线上 σ_{K} 对应的应变.

对于中心穿透裂纹的有限宽板，$2r_{\mathrm{u}}$ 可表示为

$$2r_{\mathrm{u}}=\frac{h}{Y^{6}}\cdot\frac{\varepsilon_{\mathrm{Tf}}\varepsilon_{\mathrm{s}}}{\varepsilon_{\mathrm{b}}\varepsilon_{\mathrm{T}}}\left[1-\left(\frac{\sigma_{\mathrm{T}}}{\sigma_{\mathrm{b}}}\right)^{n+1}\right]\psi^{2}(\theta) \tag{47}$$

$$Y=\sqrt{\sec\frac{\pi a}{W}} \tag{48}$$

式中，ε_{s} 为屈服应变；W 为板的宽度.

所以，对于无限宽板，有

$$\frac{\partial U_{\mathrm{p}}}{\partial a}=\frac{2n^{2}}{n^{2}-1}\sigma_{\mathrm{b}}\varepsilon_{\mathrm{b}}h\frac{\varepsilon_{\mathrm{Tf}}\varepsilon_{\mathrm{s}}}{\varepsilon_{\mathrm{b}}\varepsilon_{\mathrm{T}}}\times\left[1-\left(\frac{\sigma_{\mathrm{T}}}{\sigma_{\mathrm{b}}}\right)^{n+1}\right]\left[\left(\frac{\varepsilon_{\mathrm{b}}}{\varepsilon_{\mathrm{T}}}\right)^{\frac{n-1}{n}}-1\right]\psi^{2}(\theta)\sin\theta \tag{49}$$

对于有限宽度板，有

$$\frac{\partial U_{\mathrm{p}}}{\partial a}=\frac{2n^{2}}{n^{2}-1}\sigma_{\mathrm{b}}\varepsilon_{\mathrm{b}}\frac{h}{Y^{6}}\cdot\frac{\varepsilon_{\mathrm{Tf}}\varepsilon_{\mathrm{s}}}{\varepsilon_{\mathrm{b}}\varepsilon_{\mathrm{T}}}\times\left[1-\left(\frac{\sigma_{\mathrm{T}}}{\sigma_{\mathrm{b}}}\right)^{n+1}\right]\left[\left(\frac{\varepsilon_{\mathrm{b}}}{\varepsilon_{\mathrm{T}}}\right)^{\frac{n-1}{n}}-1\right]\psi^{2}(\theta)\sin\theta \tag{50}$$

根据式(34)和式(36)，可得：

$$\frac{\partial\sigma_{\mathrm{T}}}{\partial\theta}=\frac{K_{\mathrm{I}}}{\sqrt{2\pi r}}\cdot\frac{1}{2\psi(\theta)}\left\{2f_{1}(\theta)\frac{\partial f_{1}(\theta)}{\partial\theta}(T_{z}^{2}-T_{z}+1)+f_{1}^{2}(\theta)(2T_{z}-1)\frac{\partial T_{z}}{\partial\theta}+\left[\frac{\partial f_{1}(\theta)}{\partial\theta}f_{2}(\theta)+f_{1}(\theta)\frac{\partial f_{2}(\theta)}{\partial\theta}\right]\times(2T_{z}^{2}-2T_{z}-1)+f_{1}(\theta)f_{2}(\theta)(4T_{z}-2)\frac{\partial T_{z}}{\partial\theta}+2f_{2}(\theta)\frac{\partial f_{2}(\theta)}{\partial\theta}(T_{z}^{2}-T_{z}+1)+f_{2}^{2}(\theta)(2T_{z}-1)\frac{\partial T_{z}}{\partial\theta}+6f_{3}(\theta)\frac{\partial f_{3}(\theta)}{\partial\theta}\right\} \tag{51}$$

当 $r\to0$ 时，$T_{z}=T_{z}(z)$，与 θ 无关，可认为$\partial T_{z}/\partial\theta$ 为零，因此式(51)可写为

$$\frac{\partial\sigma_{\mathrm{T}}}{\partial\theta}=\frac{K_{\mathrm{I}}}{\sqrt{2\pi r}}\cdot\frac{1}{2\psi(\theta)}\times\left[\frac{3}{16}\sin3\theta(-T_{z}^{2}+T_{z}-1)+\frac{3}{4}\sin2\theta+\frac{1}{2}\sin\theta(-4T_{z}^{2}+4T_{z}-1)\right] \tag{52}$$

令$\dfrac{\partial\sigma_{\mathrm{T}}}{\partial\theta}=0$，可得：

$$\frac{3}{16}\sin3\theta(-T_{z}^{2}+T_{z}-1)+\frac{3}{4}\sin2\theta+\frac{1}{2}\sin\theta(-4T_{z}^{2}+4T_{z}-1)=0 \tag{53}$$

可以证明当式(53)成立时，必然有$\dfrac{\partial^{2}\sigma_{\mathrm{T}}}{\partial\theta}<0$，故 σ_{T} 取极大值. 求解式(53)即可得到 σ_{T} 取极大值时对应的角度.

令 $x_{1}=\dfrac{3}{16}(-T_{z}^{2}+T_{z}-1)$，$x_{2}=\dfrac{1}{2}(-4T_{z}^{2}+4T_{z}-1)$，则式(53)可变形为

$$16x_{1}^{2}\sin^{4}\theta+\left[\frac{9}{4}-8x_{1}(3x_{1}+x_{2})\right]\sin^{2}\theta+\left[(3x_{1}+x_{2})^{2}-\frac{9}{4}\right]=0 \tag{54}$$

由式(54)可直接求得 θ. 可见，θ 与 T_{z} 密切相关.

根据式(54)可求得两个根(θ_{1} 和 θ_{2})，且$\theta_{2}=\pi-\theta_{1}$. 由于在实际情况中 θ 随 T_{z} 的增大而增大，因此取 θ 随 T_{z} 的增大而增大的那一个作为 σ_{T} 取极大值的临界角. σ_{Tm} 的临界角 θ' 与 T_{z} 的值如表 1所示.

表 1　σ_{Tm} 的临界角 θ' 与 T_{z} 的值

Table 1　σ_{Tm} critical angle θ' and values of T_{z}

T_{z}	θ'/(°)
0	76.335
0.1	83.998
0.2	89.135
0.3	92.558
0.4	94.547
0.5	95.202

对于含穿透裂纹的无限宽度板，将上述各式代入式(10)可得：

北京航空航天大学学报

$$\frac{1-\nu T_z}{E}K_{C1}^2=\frac{Ea_0}{20}+kW_Fh_F+\frac{2n^2}{n^2-1}\sigma_b\varepsilon_b h\frac{\varepsilon_{Tf}\varepsilon_s}{\varepsilon_b\varepsilon_T}\left[1-\left(\frac{\sigma_T}{\sigma_b}\right)^{n+1}\right]\times\left[\left(\frac{\varepsilon_b}{\varepsilon_T}\right)^{\frac{n-1}{n}}-1\right]\psi^2(\theta)\sin\theta \tag{55}$$

对于含穿透裂纹的有限宽度板，式(55)相应地可表示为

$$\frac{1-\nu T_z}{E}K_{C1}^2=\frac{Ea_0}{20}+kW_Fh_F+\frac{2n^2}{n^2-1}\sigma_b\varepsilon_b\frac{h}{Y^6}\cdot\frac{\varepsilon_{Tf}\varepsilon_s}{\varepsilon_b\varepsilon_T}\left[1-\left(\frac{\sigma_T}{\sigma_b}\right)^{n+1}\right]\times\left[\left(\frac{\varepsilon_b}{\varepsilon_T}\right)^{\frac{n-1}{n}}-1\right]\psi^2(\theta)\sin\theta \tag{56}$$

同时，可有

$$K_{C1}=\sigma\sqrt{\pi a_C}Y \tag{57}$$

$$\sigma_T=\frac{W}{W-2a_C}\sigma \tag{58}$$

式中，a_C 为平面应力状态下的裂纹长度.

可用式(56)来计算材料的断裂韧性. 由于 k 是 K_C 和 K_{C1}(K_{IC})的函数，θ 又是 T_z 的函数，而 K_{C1} 又依赖于 k 和 θ 才能求解. 可见，只有采用迭代算法才能求解式(56).

3 过渡状态下材料断裂韧性计算模型及试验验证

过渡状态是指介于平面应力和平面应变之间的状态，也就是 $0<T_z<\nu$ 的状态. 大量试验结果表明，随着试样厚度由大变小(对应着由平面应变过渡到平面应力状态)，断裂韧性由 K_{IC} 增大过渡到 K_C. 试样的厚度效应主要反映在平面应力及平面应变状态材料自身性能的比值上，也就是反映在板件裂纹尖端前缘材料的平均三轴应力约束参数上. 板件表层区域总是处于平面应力状态，其 $T_z=0$. 而中心区域处于由 T_z 控制的状态，当板足够厚时，$T_z=\nu$. 在厚板中，表层区域所占比例很小，当其影响可以忽略不计时，板件的材料特性就与厚度有关了，板件处于平面应变状态. 在薄板中则有着相反的情况.

3.1 经验模型

对于过渡状态下材料断裂韧性的变化规律，人们试图以 K_{C1} 作为材料常数，利用经验公式或某些理论推算过渡状态下的 K_{C1} 值. 文献[23]提及了一个经验公式：

$$\begin{cases}\dfrac{K_{C1}}{K_{IC}}=1+B_k e^{-A_k\frac{B}{B_0}}\\ B_0=2.5\left(\dfrac{K_{IC}}{\sigma_s}\right)^2\end{cases} \tag{59}$$

式中，A_k 和 B_k 为材料常数，由试验确定；B_0 为 K_{C1} 对应的厚度.

试验表明该经验公式能较好地拟合 K_{C1}-B 曲线. 文献[20]给出了用材料拉伸性能参数估算 A_k 和 B_k 的方法. 从而使得式(59)仅由 K_{IC} 即可求得 K_{C1}-B 曲线. A_k 和 B_k 的估算方法[20]为

$$A_k=\left(\frac{\sigma_f}{\sigma_b}\right)^2\cdot\frac{1\,500\,\sigma_bK_{IC}}{(1+7.78\times10^3\times n^{-4})E\sigma_s} \tag{60}$$

$$B_k=\frac{\sigma_f^2-\sigma_s^2}{\sigma_s^2}\cdot(1+8.1\times10^{-3}\times n^{-4}) \tag{61}$$

计算结果表明，利用式(59)~式(61)能由 K_{IC} 较好地预测 K_{C1}-B 曲线.

3.2 理论计算模型

第 2 节中推导了在三维状态下材料断裂破坏的临界条件关系式，只要能够找出材料断裂时裂纹尖端前缘的三轴应力约束参数，即可计算出该状态下材料的断裂韧性值. 对于过渡状态情况，可仿照平面应变状态下的方法进行计算，认为断裂的厚度效应是一种三维的平均效应，并且主要由裂纹尖端塑性区内的平均三维约束 T_{zp} 来决定. 对于穿透 I 型裂纹体：

$$T_{zp}=\frac{1}{B\bar{r}_p}\int_0^{\bar{r}_p}\int_0^B T_z\,\mathrm{d}z\,\mathrm{d}r \tag{62}$$

对于穿透裂纹有限板，文献[24]给出了 $\bar{r}_p$ 的计算公式：

$$\bar{r}_p=\frac{1}{\pi}\left(\frac{K_I}{\sigma_s}\right)^2(1-2T_z)^2 \tag{63}$$

式(63)只在小范围屈服条件下适用，可以计算有限厚度试样在不同位置的裂纹尖端塑性区尺寸.

联合式(2)和式(63)，对式(62)采用迭代算法计算 T_{zp}. 迭代过程如下：对于给定厚度的试样，在平面应力状态下，$T_z=0$；在平面应变状态下，$T_z=\nu$. 根据线性关系估算出给定厚度的 T_z 值，根据式(63)计算裂纹尖端塑性区尺寸 $\bar{r}_p$，再根据式(2)与式(62)计算裂纹尖端塑性区内的平均三维约束 T_{zp}. 根据 T_{zp} 的值与式(63)重新计算新的 $\bar{r}_p$，再根据式(2)与式(62)重新计算新的 T_{zp}，如此反复迭代计算直至 T_{zp} 的值满足一定的误差时停止迭代，即可得到裂纹尖端塑性区内的平均三

维约束 T_{zp}. 求出 T_{zp}后可用第 2 节中的方法计算材料在过渡状态下的断裂韧性值.

以 2219-T87 铝合金为例，采用经验模型与理论计算模型分别计算了该材料的 K_{C1}-B 曲线，2219-T87 铝合金的 K_{C1}-B 曲线如图 5 所示.

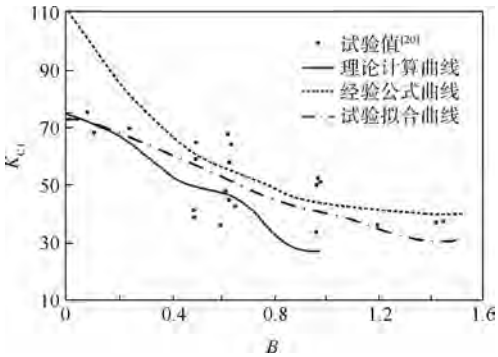


图 5　2219-T87 铝合金的 K_{C1}-B 曲线

Fig. 5　K_{C1}-B curves for 2219-T87 aluminum alloy

从图 5 中可以看出，理论计算值曲线的绝大部分均在试验点的范围内，小于经验公式计算值，且偏于安全，作为工程应用中的参考值是很有价值的. 计算曲线出现凸凹反复变化的情况可能是由以下两个原因造成的：①T_z 的表达式有待于完善；②由于 T_z 是 K_{C1}的函数，而 K_{C1}又是 k 及 T_z 的函数，因此迭代求解过程的算法还有待改进.

4　结　论

研究了过渡状态下材料断裂韧性的计算方法. 通过研究三维状态下材料断裂韧性与拉伸性能之间的关系，得到了一种过渡状态下材料断裂韧性的计算确定方法，并对 2219-T87 铝合金在过渡状态下的断裂韧性进行了理论计算. 理论计算结果表明：

1）在材料的拉伸性能已知时，过渡状态下的材料断裂韧性可通过理论计算得到，且理论计算值是偏于安全的.

2）通过理论计算得到的过渡状态下材料的断裂韧性值还存在一些问题，在以后的研究中需要不断进行完善.

参考文献（References）

[1] Kamaya M. Estimation of elastic-plastic fracture toughness by numerical simulation based on a stress-based criterion for ductile crack initiation[J]. International Journal of Pressure Vessels and Piping, 2014, 117: 2-8.

[2] Steffensen S, Madsen N D, Jensen H M. Numerical estimation of fracture toughness from indentation-induced circumferential cracking in thin films on ductile substrates[J]. International Journal of Solids and Structures, 2013, 50(20): 3406-3417.

[3] Li Y, Zhou M. Prediction of fracture toughness of ceramic composites as function of microstructure[J]. Journal of the Mechanics and Physics of Solids, 2013, 61(2): 472-488.

[4] Margolin B Z, Gulenko A G, Shvetsova V A. Improved probabilistic model for fracture toughness prediction for nuclear pressure vessel steels[J]. International Journal of Pressure Vessels and Piping, 1998, 75(12): 843-855.

[5] Wang Q Z, Zhang S, Xie H P. Rock dynamic fracture toughness tested with holed-cracked flattened Brazilian disc diametrically impacted by SHPB and its size effect[J]. Experimental Mechanics, 2010, 50(7): 877-885.

[6] Gong N P, Gao Y, Jiang X Y, et al. Research on dynamic fracture toughness of granite and finite element analysis[J]. Procedia Engineering, 2012, 37: 107-112.

[7] He Y T, Li F, Zhang G Q, et al. A theoretical approach for estimating fracture toughness of ductile metals[J]. Journal of Shanghai Jiaotong University, 2004, 38(Supple): 114-119.

[8] 徐志强，黎立云，刘一，等. I 型裂纹断裂韧性数值与理论计算比较分析[J]. 山东科技大学学报，2013，32(2)：61-65.
Xu Z Q, Li L Y, Liu Y, et al. Comparative analysis of numerical and theoretical calculation of fracture toughness model I fracture [J]. Journal of Shandong University of Science and Technology, 2013, 32(2): 61-65(in Chinese).

[9] 李江，李付国，薛凤梅，等. 基于裂纹尖端钝化能的断裂韧性估算方法[J]. 稀有金属材料与工程，2011，40(2)：577-581.
Li J, Li F G, Xue F M, et al. An estimation of fracture toughness based on the passivation energy in the crack tip[J]. Rare Metal Materials and Engineering, 2011, 40(2): 577-581(in Chinese).

[10] 马秋荣，张腾，李鹤，等. 不同厚度国产 X100 高钢级管线钢的断裂韧性[J]. 机械工程材料，2013，37(3)：69-72.
Ma Q R, Zhang T, Li H, et al. Fracture toughness of domestic X100 high strength linepipe steel with different thickness[J]. Materials for Mechanical Engineering, 2013, 37(3): 69-72(in Chinese).

[11] 方颖，李辉，慧虎，等. 基于 Master Curve 方法的 A508-III 钢断裂韧性研究[J]. 核动力工程，2011，32(1)：31-34.
Fang Y, Li H, Hui H, et al. Study on fracture toughness prediction for A508-III steel based on master curve approach[J]. Nuclear Power Engineering, 2011, 32(1): 31-34(in Chinese).

[12] 马娜，王理，陈勇. A508-III 钢韧脆转变温度范围的断裂韧性研究[J]. 核动力工程，2012，33(2)：57-60.
Ma N, Wang L, Chen Y. Research on fracture toughness of A508-III steel in transition range[J]. Nuclear Power Engineering, 2012, 33(2): 57-60(in Chinese).

[13] 何宇廷，傅祥炯. 铝合金 CCT 试样裂纹扩展三维尺寸效应金相分析[J]. 空军工程大学学报：自然科学版，2002，3(2)：5-8.
He Y T, Fu X J. Study of three-dimensional effect on crack of CCT aluminium alloy specimens by fractograph analysis[J]. Journal of Air Force Engineering University: Natural Science Edition, 2002, 3(2): 5-8(in Chinese).

[14] Farahmand B, Bockrath G E. A theoretical approach for evaluating the plane strain fracture toughness of ductile metals[J]. Engineering Fracture Mechanics, 1996, 53(6): 975-990.

[15] 杨继运,张行.材料断裂韧度与试样厚度关系研究[J].机械强度,2003,25(1):76-80.
Yang J Y, Zhang X. Research on the relation between fracture toughness and sample thickness[J]. Journal of Mechanical Strength, 2003, 25(1): 76-80(in Chinese).

[16] 杨继运,张行,张珉.裂纹扩展阻力曲线与试样厚度关系[J].北京航空航天大学学报,2003,29(7):575-578.
Yang J Y, Zhang X, Zhang M. Research on relation between crack extension resistance curve and sample thickness[J]. Journal of Beijing University of Aeronautics and Astronautics, 2003, 29(7): 575-578(in Chinese).

[17] 杨继运,张行.含裂纹板断裂韧度厚度效应的理论研究[J].机械强度,2006,27(5):672-680.
Yang J Y, Zhang X. Research on thickness effect on cracked plate fracture toughness[J]. Journal of Mechanical Strength, 2005, 27(5): 672-680(in Chinese).

[18] 杨继运,张行,张珉.含裂纹板断裂韧度厚度效应的理论与应用[J].机械工程学报,2005,41(11):32-42.
Yang J Y, Zhang X, Zhang M. Theory and application on thickness effect on cracked plate fracture toughness[J]. Chinese Journal of Mechanical Engineering, 2005, 41(11): 32-42(in Chinese).

[19] Guo W L. Elastoplastic three-dimensional crack border field-III fracture parameters[J]. Engineering Fracture Mechanics, 1995, 51(1): 51-57.

[20] 何宇廷.三维断裂准则及疲劳裂纹扩展模型研究[D].西安:西北工业大学,1997.
He Y T. A study of three dimensional fracture criteria and fatigue crack propagation model[D]. Xi'an: Northwestern Polytechnical University, 1997(in Chinese).

[21] 伍黎明.7050 铝合金耳片接头三维疲劳裂纹扩展特性研究[D].西安:空军工程大学,2013.
Wu L M. Study on the three-dimensional fatigue and fracture characteristics for 7050 aluminum alloy lug joints[D]. Xi'an: Air Force Engineering University, 2013(in Chinese).

[22] Yu H T, Wang J H, Yuan Y. An Improved hysteretic model for soils[J]. Journal of Shanghai Jiaotong University: Science Edition, 2013, 18(6): 655-659.

[23] 中国航空研究院.军用飞机疲劳损伤容限耐久性设计手册(第三册)[M].北京:中国航空研究院出版,1994:281-282.
Chinese Aviation Institute. Military aircraft fatigue, damage tolerant and durability handbook(3rd part)[M]. Beijing: Chinese Aviation Institute Press, 1994: 281-282(in Chinese).

[24] 崔振源.断裂韧性测试原理和方法[M].上海:上海科学技术出版社,1961:133-135.
Cui Z Y. The theory and method of fracture toughness test[M]. Shanghai: Shanghai Science and Technology Press, 1961: 133-135(in Chinese).

Calculation method of fracture toughness of materials at transition state

GAO Chao, HE Yuting*, CUI Ronghong, WU Liming

(Aeronautics and Astronautics Engineering College, Air Force Engineering University, Xi'an 710038, China)

Abstract: The fracture toughness of materials is very important for damage tolerance design and assessment technology, while the test method to obtain the fracture toughness in the transition state is costly and time-consuming. In order to overcome these disadvantages, a determination method of fracture toughness based on calculation in the transition state was presented. The triaxial stress constraint parameter for the three-dimensional crack tip was analyzed, the relation between fracture toughness and material tensile properties was researched with the energy criterion for material fracture, meanwhile the relation expression in the three-dimensional stress state for I type crack fracture and the determination method of fracture toughness of materials based on calculation in the transition state were obtained. The fracture toughness in the transition state for 2219-T87 aluminum alloy was calculated with the determination method and empirical formula most of the theoretical calculation values were safer compared with test results and smaller than the values calculated by the empirical formula.

Key words: fracture toughness; transition state; plane strain; triaxial stress constraint; crack tip

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0256

面向机身柔性装配的在线编程技术

全下文载

武锋锋，李东升*，王亮

（北京航空航天大学 机械工程及自动化学院，北京 100191）

摘 要：为满足机身数字化装配的“一架多型”和“一架多态”深层次柔性需求，使“桥架式”柔性工装具有更好的自主适应能力，必须建立相应的运动学模型，通过在线编程技术合理地规划运动控制策略.针对机身定位形式与柔性工装特点建立了柔性工装运动学模型，研究了基于仿真模型的在线编程原理、装配基准点构建方式和运动控制策略规划技术，提出了结构功能结合工艺需求驱动的仿真环境创建方法.为快速生成运动控制指令，开发了面向机身柔性装配的在线编程系统.以某飞机后机身装配为例，验证了系统的精准性、高效性和超强自主适应能力.

关 键 词：机身；柔性装配；在线编程；自主适应；运动学模型；仿真
中图分类号：V262.1
文献标识码：A **文章编号**：1001-5965(2015)04-0641-08

随着飞机设计中对整体结构的采用[1]，传统刚性工装很难适应结构件本身的制造差异，会产生不可忽视的有害装配应力，影响装配质量.柔性制造系统基于具体的工艺需求，通过一定的算法获取调控指令，引导控制系统将定位器平稳、精确地重构到指定位置，可适应这种差异性，实现对装配件的合理定位与夹紧[2].

从现有成果来看，国内对于壁板组件[3-4]、前襟翼小部件[5]装配、总装调姿[6-8]与对接[9]数字化柔性装配系统的研究较为成熟，柔性化主要体现在两方面：①快速响应产品变化的能力；②夹紧定位结构的应变能力[10].调形重构仅仅发生在工装准备期间或部件调姿阶段，装配工作一旦开展，柔性工装定位器就不再运动，与刚性工装无异，不具备对装配过程中不同装配状态的适应性.

机身装配过程中，对象结构复杂、待装配件成百上千、工艺涉及内容繁多且装配现场环境多变，因此机身数字化柔性工装对装配状态适应性的要求更为突出，国内外对机身类大部件数字化柔性装配方面的研究非常有限.文献[11-13]研究的部件柔性装配平台可根据实际需求，基于测量系统反馈的柔性模块“六足(Hexapods)”顶板位置数据，通过专用算法获得六足各腿的相对调形量，引导专用接头运动到位，可适应定位件的差异，但没有对装配状态的适应性进行进一步相关研究.国内郭洪杰等[14-15]针对飞机快速研制期间工装准备的短周期、低成本特征，研发了“桥架式”机身部件柔性工装，具备与型号的“一对多”及与装配状态的“一对多”特性，但却没有对这一特性进行详细研究.

本文从算法编程角度研究了“桥架式”柔性工装对产品、环境、装配状态等的自主适应性能，基于柔性工装运动学模型，分析了面向机身柔性装配的在线编程技术，并通过构建面向在线编程

收稿日期：2014-05-08；**录用日期**：2014-08-25；**网络出版时间**：2014-10-30 11:03
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0256.html

作者简介：武锋锋(1983—)，女，山西洪洞人，博士生，wffqll00@163.com
*通讯作者：李东升(1965—)，男，甘肃静宁人，教授，lidongs@buaa.edu.cn，主要研究方向为现代飞机制造工艺与先进装备、航空钣金数字化制造与钛合金构件精密成形技术和超大型曲面的超精密柔性成形制造理论与应用技术.

引用格式：武锋锋，李东升，王亮. 面向机身柔性装配的在线编程技术[J]. 北京航空航天大学学报，2015，41(4)：641-648.
Wu F F，Li D S，Wang L. Online programming technique for flexible assembly of fuselage [J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：641-648 (in Chinese).

北航学报 赠阅

仿真环境增强了编程效率与规范性. 以某机型后机身部件装配过程为例,验证了所开发系统的有效性及适应性.

1 柔性工装调形运动原理

1.1 机身数字化柔性工装特点

机身属于大部件范畴,对工装的适用性和开敞性要求更为严格. 首先,机身装配包含两个基本装配状态:骨架装配状态和壁板装配状态,两个状态对工装有不同的定位及夹紧要求. 其次,在工艺与工装设计合理性验证的仿真中,由于机身结构复杂且待装配件众多,几乎不可能将所有标准件及一些细小零件的装配过程进行人机验证,而实际装配过程中又确实存在一些将工装视为刚性工装,从而难以装配的状况,因此柔性工装需要具备随时进行避让性调形运动的功能. 最后,整体框与整体壁板的应用使得工装必须具备一定的调节能力,以适应其制造差异.

按设计依据是否独立于产品设计细节,柔性工装分为仅需要基本信息的通用结构和依赖于最终细节的专用结构[16]. 通用结构即工装设备,装配对象更换后,仅需要改变零组件的空间位置,不必重新设计其几何外形. 专用结构与机身产品直接接触,随装配对象不同而具有不同的空间位置和几何外形,需根据具体装配对象进行定制,与设备通过"十字转接"等机构实现快速更换,实现柔性工装的"一架多型"功能.

根据是否参与调形运动,"桥架式"柔性工装设备分为主体支撑与调形构件. 主体支撑包括 1 个静态框架和 4 根具备导向功能的静横梁,起支撑与导向作用,能保证调形运动的平稳运行. 调形构件包括动横梁和定位立柱,按上下两层分布,每层布置 5 根动横梁,每根动横梁上再布置 2 ~ 6 个定位立柱. 在伺服电机驱动下,动横梁沿静横梁运动完成工装定位单元(定位立柱 + 专用结构组件)的 x 向调形,定位立柱沿平行与垂直动横梁方向带动定位组件完成 y 向和 z 向调形,最终在工装内部形成可任意定位的工作空间. 图 1 为柔性工装的结构分解图.

为在降低控制系统故障率及研制成本的同时保证调形效率,柔性工装在每根横梁上分别加装了一种分时复用机构-滑板小车,实现了"串-并"混合模式调形[17]. 图 2 为 x 向和 y-z 向滑板小车结构图. 通过零位、识别等传感器和气动装置辅助完成滑板小车的回零、对象识别、锁定与解锁等操作,伺服电机带动滑板小车串行驱动组内(同一横梁上)全部控制对象完成调形工作,组与组之间并行开展.

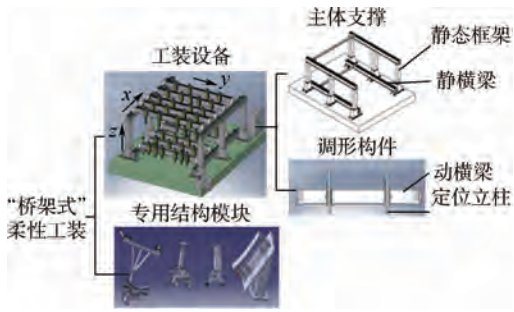


图 1 柔性工装的结构分解图

Fig. 1 Structural decomposition diagram of flexible tooling

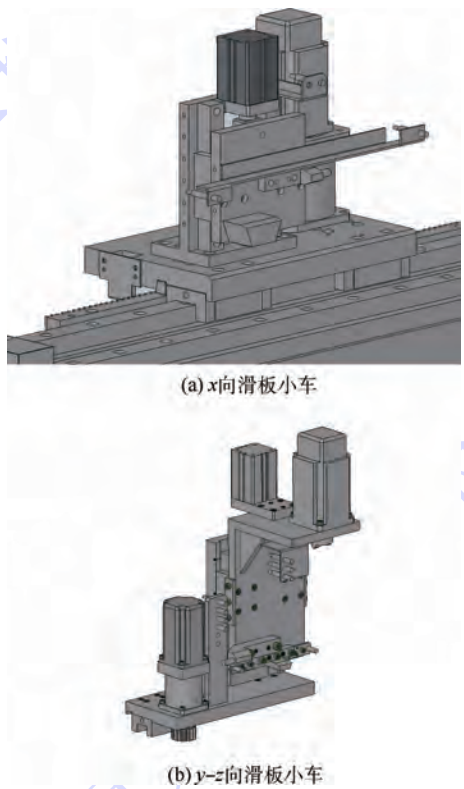


图 2 x 向和 y-z 向滑板小车结构图

Fig. 2 Structural diagrams of x and y-z sliding plate

x 向滑板小车上装配有 1 台伺服电机,y-z 向滑板小车配备有 2 台伺服电机. 柔性工装共有 4 根静横梁和 10 根动横梁,共需要 24 台伺服电机,伺服电机分布如图 3 所示.

1.2 基于柔性工装的机身定位形式

机身作为典型的框架式结构,采用"以骨架为基准"的装配工艺. 首先,利用工装上的专用接头定位框,然后以框为基准定位纵梁及隔板等零组件,连接形成机身骨架. 接着,以骨架为基准放置蒙皮壁板、压紧并连接,形成完整的机身部件. 最后下架,转移至大部件对接平台[18].

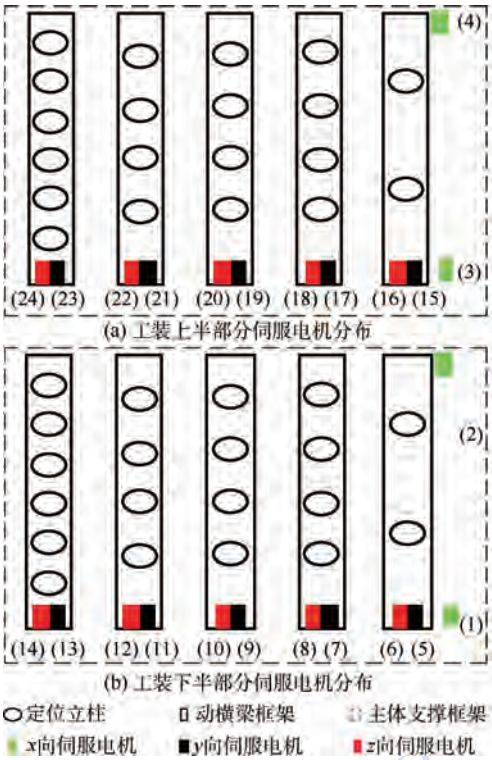


图 3　伺服电机分布

Fig. 3　Servo motor distribution

传统的机身骨架与蒙皮壁板装配工作分别在骨架装配型架与壁板装配型架上完成，中间需要经历下架及上架这个重新定位的过程，影响装配进度与质量. 通过合理设计专用结构，基于定位单元的可调性，“桥架式”柔性工装体系可将骨架装配与蒙皮壁板装配工作集中在一个工装上完成，减少不必要的重复定位操作和工装研制成本，缩短装配周期，提高机身部件的装配质量.

“桥架式”柔性工装体系不仅可以利用工装的可重构调形特征实现基于数字链传递的数字化定位，而且还可通过构建基于在线编程的仿真环境模拟现场环境，根据装配状态与避让需求规划柔性工装的运动控制策略，实现工装与产品装配状态“一对多”的“一架多态”功能.

1.3　基于三维几何的柔性工装运动学建模

柔性工装快速精确定位与调形时，目标位置可用定位单元飞机坐标系下的最终三维坐标值表示，工装运动可以看成是电机驱动滑板小车带动调形构件移动至目标位置的过程. 运动学模型主要根据运动学方面的知识直接求得电机运动与目标位置之间的关系. 逆运动学模型可用定位单元的目标坐标值作为输入来确定电机编号，识别对象、位移等输出.

机型 k 动横梁 l 上第 m 个定位立柱目标坐标值为 $(X^{g}_{klm}, Y^{g}_{klm}, Z^{g}_{klm})$，$m = 1 \sim 6$，$l = 1 \sim 10$. 工装原点在机型 k 飞机设计坐标系下的坐标值为 (X_k, Y_k, Z_k)，则各定位单元工装坐标系下的目标坐标值为

$$(x^{g}_{klm}, y^{g}_{klm}, z^{g}_{klm}) = (X^{g}_{klm}, Y^{g}_{klm}, Z^{g}_{klm}) - (X_k, Y_k, Z_k) \tag{1}$$

柔性工装零位校准后，各电机零位工装坐标系下坐标值为 (X_n, Y_n, Z_n)，$n = 1 \sim 24$ 为电机编号，由图 3 中电机的分布情况可直接得出 x 向调形电机编号为

$$n = \begin{cases} 1,2 & l \leqslant 5 \\ 3,4 & l > 5 \end{cases} \tag{2}$$

各组调形对象 ID 编号独立，因此锁定对象 ID 编号为

$$q = \begin{cases} l & l \leqslant 5 \\ l - 5 & l > 5 \end{cases} \tag{3}$$

电机位移量：

$$x_n = x^{g}_{kl1} - X_n \tag{4}$$

y 向调形电机编号：

$$n = 2l + 3 \tag{5}$$

锁定对象 ID 为立柱编号 m，电机位移量：

$$y_n = y^{g}_{klm} - Y_n \tag{6}$$

z 向调形电机编号：

$$n = 2l + 4 \tag{7}$$

y 向调形完毕，不松开锁定对象，z 向调形电机继续调整定位单元 z 向位移：

$$z_n = z^{g}_{klm} - Z_n \tag{8}$$

2　基于仿真的在线编程算法

2.1　基于仿真的在线编程原理

在机身数字化研制过程中，专用运动控制指令在线编程起着重要的桥梁作用. 机身装配过程中，需要基于 CAD 系统(CATIA)中的产品及工装等数模、结合现场情况、根据工艺要求在线规划运动控制策略，为工装运动控制提供必要的数据依据，以保障机身装配工作的顺利进行. 图 4 为机身部件柔性装配在线编程系统的定位.

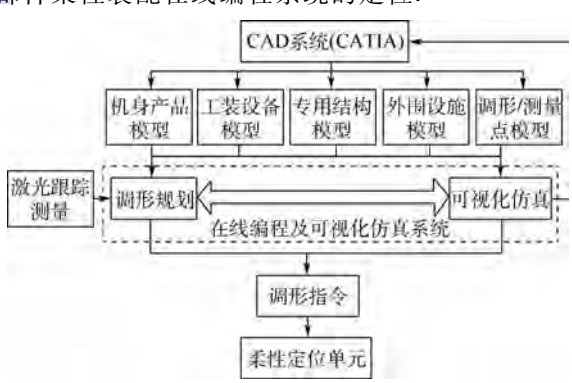


图 4　机身部件柔性装配在线编程系统的定位

Fig. 4　Position of online programming system for flexible assembly of fuselage structure

北航学报 赠阅

基于预先设计的数字化实体模型,检索定位单元三维坐标值,按柔性工装逆运动学模型算得电机运动控制指令,驱动滑板小车依次将调形构件移动到装配理论位,然后装配专用结构组件,检测完成机身装配前的工装准备工作.

产品装配期间,为便于零组件的定位、夹紧与装配,保证工装的开敞性,柔性工装需要根据工艺需求与装配状态随时进行调形运动. 按激光跟踪测量系统给出的定位单元现场精确位置信息、产品装配进度及工艺要求,在仿真环境中进一步规划新的调形运动,生成相应运动控制指令. 基于模型的在线编程包括 3 个模块:创建仿真环境、调形规划和指令后处理.

2.2 柔性装配基准点构建

为保证数字化柔性工装系统中数据的一致性与运动控制规划计算的方便,在工装上除原点外还设定了调形基准点及检测基准点. 图 5 为调形基准点与检测基准点分布图.

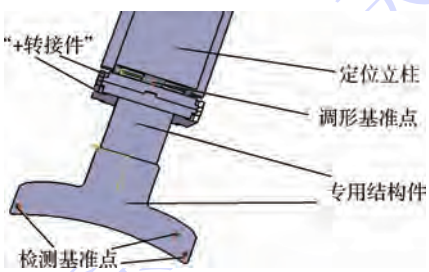


图 5 调形基准点与检测基准点分布图

Fig. 5 Scattergram of adjustment and measurement datum points

调形基准点位于各定位立柱的固定位置,标定各定位立柱的坐标信息,并可在工装准备期间用于检测调形构件的到位情况. 检测基准点选择与产品直接接触定位件上的关键特征点(按"3-2-1"原则设计的 3 个保证装配精度的 OTP 点),与装配产品具体型号相关,用于分析定位单元的现场位置信息,检测到位状况.

2.3 工装运动策略规划

工装准备期间,必须首先启动设备归零功能,使各调形构件位于自身零位,然后基于工装的逆运动学模型获取控制策略,启动自动调形模式,令各定位立柱运动至专用结构组件装配所需位置,即产品装配理论位,通过激光跟踪测量系统检测调形基准点的坐标值校准到位精度,最后安装专用结构组件,完成产品装配前的工装准备工作.

产品装配期间,从激光跟踪测量系统反馈飞机设计坐标系下定位单元的现场坐标值到运动控制策略的生成,需要基于坐标系转换原理通过数学计算获取数字化测量结果、定位单元调形基准点坐标值与控制指令之间的相互关系. 工装运动策略规划包含以下几步.

1) 创建仿真环境. 运动控制策略规划的前提条件是在数字化三维仿真软件平台上创建仿真环境. 创建仿真环境分为两步:

首先在三维软件平台上加载柔性工装、外围设施及基于装配进度的机身已装配零组件数模.

其次,基于激光跟踪测量系统给出的检测基准点位置信息,在仿真环境中将柔性工装各定位单元移动至现场位置,模拟现场状态.

机型 k 装配数模中动横梁 l 上第 m 个定位立柱检测基准点坐标值为 $(X_{klmi}, Y_{klmi}, Z_{klmi})$, $i = 1 \sim 3$, $m = 1 \sim 6$, $l = 1 \sim 10$. 本处及后续出现的不存在点坐标值以 $(-10\,000, -10\,000, -10\,000)$ 代替,不影响整个计算流程以及结果的获取.

激光跟踪测量系统给出飞机坐标系下检测基准点坐标值 $(x_{klmi}, y_{klmi}, z_{klmi})$. 将各定位单元移至现场位置,仿真环境中调形构件动横梁 l 及其上布置的 6 个定位单元的 x 向移动距离为

$$\Delta L_{kl} = \frac{1}{3}\sum_{i=1}^{3}(x_{kl1i} - X_{kl1i}) \tag{9}$$

l 上定位单元 m 的 y 向与 z 向移动距离为

$$\Delta Y_{klm} = \frac{1}{2}\sum_{i=1}^{2}(y_{klmi} - Y_{klmi}) \tag{10}$$

$$\Delta Z_{klm} = z_{klm1} - Z_{klm1} \tag{11}$$

机型 k 飞机设计坐标系下,调形基准点坐标值为 $(X_{klm}, Y_{klm}, Z_{klm})$. 机型不变,调形基准点与检测基准点关系固定:

$$(\Delta X_{klm}, \Delta Y_{klm}, \Delta Z_{klm}) = \left(\frac{1}{3}\sum_{i=1}^{3}(X_{klm} - X_{klmi}), \frac{1}{2}\sum_{i=1}^{2}(Y_{klm} - Y_{klmi}), Z_{klm} - Z_{klm1}\right) \tag{12}$$

则 x 向调形构件 l 工装坐标系下的坐标值为

$$x_{kl} = X_{kl1} - X_k + \Delta L_{kl} \tag{13}$$

y 向和 z 向调形构件动横梁 l 上的第 m 个定位立柱工装坐标系下的坐标值为

$$y_{klm} = Y_{klm} - Y_k + \Delta Y_{klm} \tag{14}$$

$$z_{klm} = Z_{klm} - Z_k + \Delta Z_{klm} \tag{15}$$

2) 规划调形运动. 根据工艺和现场需求逐步规划调形运动并进行碰撞检测验证,记录调形对象、方向、距离与顺序等信息.

首先,基于模拟现场状态的仿真环境,按照工艺与现场操作要求为指定调形构件设定目标位移值,指定调形对象、方向与距离. 例如,设定动横梁

l 移动量 Δx_{kl}，或者动横梁 l 上第 m 个定位单元的 y 向与 z 向移动量 Δy_{klm} 和 Δz_{klm}，移动量的正负号代表该方向的正负.

其次，在仿真环境中进行运动仿真，校验运动目标设定的合理性，检测调形构件运动过程中潜在的接触、干涉与超量程规划现象. 运动仿真过程中，若调形对象与产品、工装其他部位以及外围设备均无干涉，则记录该单步运动规划，否则取消该步运动，即将该步运动复位，重新设定调形目标.

最后，记录检验安全的调形过程. 单条记录涉及调形对象、方向与距离，记录之间的先后顺序即为调形顺序. 记录为一四维数组，若为 x 向调形，则第 1 维为动横梁 l 工装坐标系下的目标坐标值 $x^{g}_{kl} = x_{kl} + \Delta x_{kl}$；第 2 维无意义，以 0 标记；第 3 维记录动横梁编号 l；第 4 维以 -1 标记，为指令后处理过程中判别是否为 x 向调形的标识.

若为 y 向或 z 向调形，则第 1 维为动横梁 l 第 m 个定位单元工装坐标系下的 y 向目标坐标值 $y^{g}_{klm} = y_{klm} + \Delta y_{klm}$；第 2 维为 z 向目标坐标值 $z^{g}_{klm} = z_{klm} + \Delta z_{klm}$；第 3 维为动横梁编号 l；第 4 维为定位单元编号 m.

3）指令后处理. 规划记录涉及调形构件运动的先后顺序、对象、方向及运动量，不能直接作为调形驱动器伺服电机的运动控制指令，需将这些记录转换为电机编号、调形对象 ID 与电机绝对移动量的表示形式才可为柔性工装控制系统所接收，图 6 为指令后处理流程图.

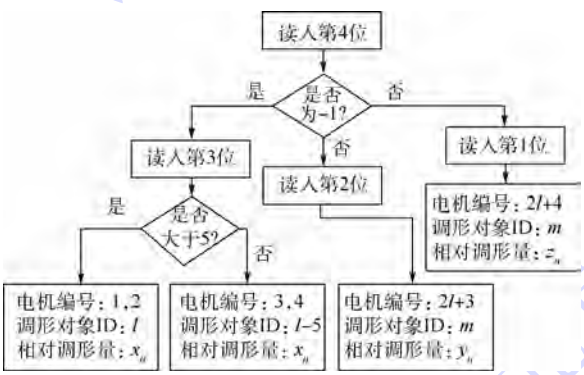


图 6　指令后处理流程图

Fig. 6　Flowchart of instruction's post-processing

3　面向在线编程的仿真环境构建

分析基于仿真的柔性工装在线编程过程可以发现，合理、规范地构建仿真环境对程序适应性与效率至关重要. 首先，编程系统对机身数字化装配“一架多型”和“一架多态”深度柔性需求的响应均体现在仿真环境构建过程中. 其次，调形运动规划过程中除碰撞检测模块，其余均为一般数学计算，而碰撞检测的效率与仿真环境构建的合理性和规范性直接相关.

机身柔性装配的原始三维数模是在线编程系统构建仿真模型的依据和数据获取的主要来源，通常存在大量的冗余数据，包含大量几何与拓扑、材料、工艺、产品管理、功能和技术要求等信息，用其直接创建仿真环境会造成因数据量过大而导致计算机硬件负担过重的现象. 因此构建仿真环境时，应抑制装配数模中一些不必要的信息，基于结构功能对模型进行轻量化处理，添加与编程计算相关的关键点信息，并通过装配 PBOM 信息控制产品装配状态.

3.1　基于结构功能性的模型轻量化

模型轻量化处理时，首先按结构在柔性工装调形运动过程中充当的功能角色对装配数模进行功能模块划分，然后抑制各个模块内部结构，对数模做仅保留外部轮廓和装配位置信息的处理，完成模型轻量化过程.

针对后机身装配数模，可分为柔性工装设备、专用结构组件、产品与外围设施 4 大类. 其中，按运动自由度，设备数模分为主体支撑、10 根 x 向动横梁、60 个 y 向滑板与 60 根 z 向定位立柱. 每根定位立柱配备 1 个专用结构组件，可从装配数模中分离出 60 个专用结构组件. 产品数模按照当前装配进度保留已装配部分，舍去未装配部分. 外围设施单独作为一部分. 在 CATIA 软件平台中，分别将这 193 个模块组件的内部零件进行合并，作为 1 个零件操作（对于不存在的滑板、定位立柱和专用结构组件，可创建一个空 PART 文件），然后依次加载到仿真环境中，完成基于结构功能性的轻量化模型构建.

3.2　基于轻量化模型的关键点创建

编程算法中离不开关键点信息的检索与提取，而数模经过轻量化处理后，点坐标信息都被抑制了，因此需要根据需求重新创建. 关键点包括工装原点、调形基准点与检测基准点 3 类：

1）工装原点位于工装静横梁上，位置固定，可直接创建于主体支撑组件轻量化数模中. 根据设计给出的工装原点在飞机坐标系下的坐标值计算转换矩阵，移动工装数模，使工装设备轻量化数模位于飞机设计坐标系中.

2）调形基准点位于每根定位立柱的固定位置，可通过程序自动创建于 z 向定位立柱零件体上. 此外，为响应柔性工装“一架多型”功能（即创建仿真环境时，当机型发生更换，已经过轻量化处理的柔性工装设备数模不需要重新加载，仅重新

北航学报 赠阅

加载与机型对应的专用结构组件轻量化数模），需要在每根定位立柱对应的专用结构组件数模与调形基准点在空间上重合的点位处同时创建一个辅助调形基准点. 当机型更换时，加载专用结构组件轻量化数模，检索辅助调形基准点坐标值作为定位立柱目标位置，基于工装的逆运动学模型获取控制策略，然后驱动柔性工装设备的各调形构件轻量化数模运动，使调形基准点所在端面与辅助调形基准点所在端面重合，完成仿真环境中柔性工装数模的创建.

3）创建检测基准点时，需在装配数模上获取其坐标值，然后创建于专用结构组件轻量化数模零件体中. 获取的数据为检测基准点飞机坐标系下坐标值，工装专用结构组件数模也是基于飞机坐标系而设计的，且在轻量化过程中保留了装配位置信息，即坐标系信息没有发生变化，可直接应用获取的坐标值数据创建检测基准点. 调形运动仿真过程中，点与定位单元作为一体变换位置.

3.3 装配工艺信息模型构建

在创建轻量化模型的过程中，需要根据装配进度将产品分为已装配部分和未装配部分，可利用装配 PBOM 信息根据现场装配进度所属工艺编号自动划分产品已装配部分和未装配部分，简化产品轻量化操作. 图 7 为装配数模处理模块，左侧为装配数模处理模块的轻量化数模加载结构树，右侧为已装配件选择模块.

图 7 装配数模处理模块

Fig. 7 Processing module of assembly model

结构功能结合工艺驱动的仿真环境创建方法最显著的特点是增加了编程系统的可修改性. 专用装配数模处理模块增强了仿真环境构建的规范性和效率. 当条件发生改变时，不需要重新加载全部信息，针对具体情况修改部分内容即可. 例如，当机型发生改变时，仅需要重新加载专用结构、外围设施、辅助调形基准点、检测基准点和 PBOM 信息；而现场装配进度不同，则根据所选工艺编号加载相应状态的飞机产品信息.

4 应用案例

4.1 系统运行

在线编程系统是机身数字化柔性装配平台不可或缺的一部分，通过规范数据传输格式实现了与柔性装配系统其他软硬件之间的无障碍通讯，在线编程系统流程图如图 8 所示.

将待装配的后机身装配数模读入系统后，按要求处理数模，构建仿真环境. 首先，参照轻量化数模加载结构树，将处理过的各种数模及点信息依次加入仿真环境中，程序依次检索并提取工装原点、辅助调形基准点与调形基准点在飞机坐标系下的坐标值，经过计算将工装设备各调形构件移动至与专用结构组件完全匹配. 在专用结构组件上创建检测基准点时，逐个选取并查询其坐标

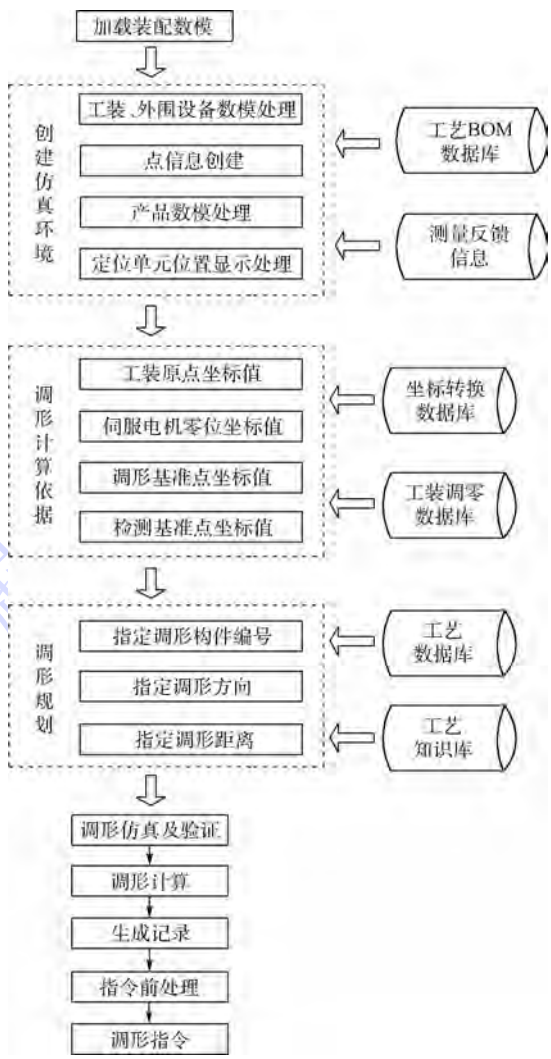


图 8 在线编程系统流程图

Fig. 8 Flowchart of online programming system

值后手动创建的方法比较耗时,且容易出错,可采用程序自动检索并创建的方法. 接着,选择装配工艺编号,将产品数模的已装配部分分离出来,通过程序自动进行轻量化处理,然后加载到仿真环境中. 最后,依据激光测量系统反馈的现场定位单元位置信息,在仿真环境中重新显示工装定位单元的位置,完成仿真环境的构建.

为了在调形规划过程中同步完成调形计算,需要加载工装坐标系下各伺服电机零位坐标等辅助信息,提取各定位单元调形基准点和检测基准点在飞机坐标系下的坐标值等.

基于模拟现场的仿真环境,逐步规划调形构件编号、调形方向及距离,通过仿真验证、调形计算、记录生成以及指令后处理过程,最终生成柔性工装控制系统可接受的运动控制指令,调形指令生成对话框如图 9 所示.

图 9　调形指令生成对话框

Fig. 9　Dialog box of adjustment instruction generation

4.2　型号验证

选择某型号飞机后机身装配过程对在线编程系统进行验证,工装调形实例验证如图 10 所示,实例证明系统可精准、高效地生成规范的柔性工装运动控制指令,与装配平台其他系统间实现数据的可靠交互.

图 10　工装调形实例验证

Fig. 10　Adjustment instance validation of tooling

相比 CATIA 自有编程模块和人工编程,在线编程系统可大幅度提高编程效率、提升系统对环境改变的自主适应能力、规范创建的仿真环境及运动指令的格式并增强指令生成的可靠性. 装配数模处理模块可简化仿真环境创建过程、规范创建结果、快速响应机型与装配现场状况的改变、辅助“桥架式”柔性工装自主适应“一架多型”和“一架多态”机身数字化装配的深度柔性要求并减少工装准备周期. 调形指令生成对话框可降低调形规划中的误操作并增加生成指令的规范性与可靠性. 表 1 为在线编程系统效率对比.

表 1　在线编程系统效率对比

Table 1　Comparison of effectiveness of online programming system

参数	在线编程系统	CATIA 自有模块	人工编程
可靠性	高	中等	低
对编程者的要求	基本知识	一定的编程能力	编程能力
编程时间/h	1 ~ 2	12	24 ~ 36
模型变化所需修改时间/h	1 ~ 1.5	10	24 ~ 36
状态变化所需修改时间/h	0.5	4 ~ 6	6 ~ 8

5　结　论

1) 基于三维几何的柔性工装运动学建模,通过设定相应的定位单元目标参数及利用数值方法计算电机位移参数对工装的适应性进行预判,提高了工装运动控制的安全性. 另外,在产品装配期间的工装运动控制策略规划过程中,利用此模型以数值方法实现了对运动目标设定合理性的评价和不考虑动力学效应的仿真运动.

2) 在分析基于仿真模型在线编程技术的基础上,研究了结构功能性结合工艺驱动构建仿真环境的方法,不仅为工装调形运动提供了必要的数据依据,而且大大提升了编程的可靠性、自主适应性、规范性和效率.

3) 成功进行了工程应用验证,证明了在线编程系统可精准、高效地生成柔性工装调形运动控制指令,数据可在平台的各系统间顺畅传递,能够从算法编程角度很好地响应柔性工装“一架多型”和“一架多态”的深度柔性需求,最大幅度地减少工装准备周期,具有良好的可维护性.

参考文献 (References)

[1] 陈建平. 数字化柔性装配的飞机结构设计[J]. 科技创新与应用,2013(23):12.
Chen J P. Aircraft structural design for digital flexible assembly [J]. Technology Innovation and Application,2013(23):12(in Chinese).

[2] Jonsson M,Ossbahr G. Aspects of reconfigurable and flexible fixtures[J]. Production Engineering,2010,4(4):333-339.

[3] Bi Y B,Yan W M,Ke Y L. Numerical study on predicting and correcting assembly deformation of a large fuselage panel during digital assembly [J]. Assembly Automation, 2014, 34 (2):

204-216.

[4] 王亮,李东升,罗红宇,等.飞机装配数控柔性多点工装技术及应用[J].北京航空航天大学学报,2010,36(5):540-544.
Wang L,Li D S,Luo H Y,et al. Numerical control reconfigurable compliant tooling technology and application in aircraft assembly [J]. Journal of Beijing University of Aeronautics and Astronautics,2010,36(5):540-544(in Chinese).

[5] 郭飞燕,王仲奇,康永刚,等.基于坐标孔的数字化柔性工装定位技术[J].计算机集成制造技术,2013,19(4):720-726.
Guo F Y,Wang Z Q,Kang Y G,et al. Locating technology in digital flexible tooling based on coordinate holes[J]. Computer Integrated Manufacturing Systems,2013,19(4):720-726(in Chinese).

[6] 郭志敏,蒋君侠,柯映林.基于 POGO 柱三点支撑的飞机大部件调姿方法[J].航空学报,2009,30(7):1319-1324.
Guo Z M,Jiang J X,Ke Y L. Posture alignment for large aircraft parts based on three POGO sticks distributed support[J]. Acta Aeronautica et Astronautica Sinica,2009,30(7):1319-1324(in Chinese).

[7] Zhang B,Yao B G,Ke Y L. A novel posture alignment system for aircraft wing assembly[J]. Journal of Zhejiang University:Science A,2009,10(11):1624-1630.

[8] Zhu Y G,Huang X,Fang W. Trajectory planning algorithm based on quaternion for 6-DOF aircraft wing automatic position and pose adjustment method[J]. Chinese Journal of Aeronautics,2010,23(6):707-714.

[9] 邱宝贵,蒋君侠,毕运波,等.大型飞机机身调姿与对接试验系统[J].航空学报,2011,32(5):908-919.
Qiu B G,Jiang J X,Bi Y B,et al. Posture alignment and joining test system for large aircraft fuselages[J]. Acta Aeronautica et Astronautica Sinica,2011,32(5):908-919(in Chinese).

[10] 王建华,欧阳佳,陈文亮.飞机柔性装配工装关键技术及发展趋势[J].航空制造技术,2013(17):49-52.
Wang J H,Ouyang J,Chen W L. Key technology and development of flexible assembly tool for aircraft[J]. Aeronautical Manufacturing Technology,2013(17):49-52(in Chinese).

[11] Kihlman H,Engström M. Flexible fixtures with low cost and short lead-time,2007-01-3797[R]. Warrendale:SAE,2007.

[12] Millar A,Kihlman H. Reconfigurable flexible tooling for aerospace wing assembly,2009-01-3243[R]. Warrendale:SAE,2009.

[13] Kihlman H,Engstrom M. Flexapods-flexible tooling at SAAB for building the NEURON aircraft,2010-01-1871[R]. Warrendale:SAE,2010.

[14] 郭洪杰,康晓峰,王亮,等.飞机部件装配数字化柔性工装技术研究[J].航空制造技术,2011(22):94-97.
Guo H J,Kang X F,Wang L,et al. Research on flexible tooling technology for digital assembly of aircraft fuselage[J]. Aeronautical Manufacturing Technology,2011(22):94-97(in Chinese).

[15] 袁立,郭洪杰.一代飞机一代技术[J].航空制造技术,2012(23-24):38-41.
Yuan L,Guo H J. New generation aircraft with new generation technology[J]. Aeronautical Manufacturing Technology,2012(23-24):38-41(in Chinese).

[16] Li Y G,Jian J B,Yan R J,et al. Aircraft tooling collaborative design based on multi-agent and PDM[J]. Concurrent Engineering:Research and Applications,2009,17(2):139-146.

[17] 王亮,李东升.飞机数字化装配柔性工装的低成本化[J].南京航空航天大学学报,2012,44(增刊 1):27-31.
Wang L,Li D S. Low cost of flexible tooling technology for aircraft digital assembly[J]. Journal of Nanjing University of Aeronautics & Astronautics,2012,44(Supple. 1):27-31(in Chinese).

[18] 王海宇.飞机装配工艺学[M].西安:西北工业大学出版社,2012:7-31.
Wang H Y. Aircraft assembly technology[M]. Xi'an:Northwestern Polytechnical University Press,2012:7-31(in Chinese).

Online programming technique for flexible assembly of fuselage

WU Fengfeng, LI Dongsheng*, WANG Liang

(School of Mechanical Engineering and Automation, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: In order to meet the deep flexible demands of "one fixture with more models" and "one fixture with more states" for digital assembly of fuselage, enable the "bridge-frame" flexible tooling to have better capacity for independent adaptability, its kinematic model must be set up and its movement controls strategy must be planned properly by online programming technique. For the locating styles of fuselage and the characteristics of flexible tooling, the kinematic model of flexible tooling was set up, then simulation model-based online programming disciplines, the way of building reference points for assembly and the technology of movement controls strategy planning were studied. The way of building a process-driven simulation environment based on structure function was proposed. In order to generate the codes quickly to control motion, the online programming system for flexible assembly of fuselage was developed. With the assembly of rear fuselage, for example, the accuracy, high efficiency and strong independent adaptability of this system were verified.

Key words: fuselage; flexible assembly; online programming; independent adaptability; kinematic model; simulation

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0272

基于单幅立式标靶图像的单目深度信息提取

黄小云，高峰*，徐国艳，丁能根，邢龙龙

（北京航空航天大学 交通科学与工程学院，北京 100191）

摘 要：针对在智能车中车载单目视觉系统已经检测到路面障碍物的情况下，计算障碍物相对于本车的距离问题，提出了一种仅利用单幅标靶图像且无需相机内部参数的图像深度信息提取方法. 该方法利用一种放置于相机前方的立式标靶，建立图像纵坐标像素与实际成像角度之间的映射关系，结合投影几何模型实现实时深度信息的提取. 依据立式标靶图像的特点，设计了包括标靶图像感兴趣区设置、模板匹配、候选点聚类、筛选及精确定位等处理的亚像素级角点检测及定位算法. 实验结果显示，该方法具有较高的测量精度及实时性. 相对于在路面摆放参照物的方法，该方法无需大标定场地且规避了数据拟合引起的误差. 同时该方法标定只需一幅图像，过程简单，便于实际应用.

关 键 词：机器视觉；单目测距；立式标靶；智能车；角点检测
中图分类号：U491
文献标识码：A **文章编号**：1001-5965(2015)04-0649-07

机器视觉作为智能车辆环境感知系统中最重要的组成部分，为决策层提供大量必要的环境信息，具有十分重要的意义. 其中，障碍物的深度信息为无人驾驶或辅助驾驶系统的主动避撞、防撞预警及路径规划等[1-2]提供重要的参数信息. 目前智能车的机器视觉测量一般分为双目视觉测量[3-4]和单目视觉测量两类. 双目视觉测量容易受到特征点误匹配的影响，且计算量大，难以满足实时性要求；而单目视觉测量结构简单，图像处理算法运算效率高，具有广泛的应用前景.

目前单目视觉系统一般采用对应点标定法（简称标定法）来获取待测物体的深度信息[5-9]. 该方法通常是利用棋盘格标定板标定出相机内外部参数后，结合投影模型，求解出图像坐标系与实际成像角度之间的对应关系，从而得到深度信息. 此方法需要多次采集不同方位的标定板图像，并且精确记录每个点在世界坐标系和图像坐标系中的对应坐标，标定精度对于测量精度影响相当大，某些参数的标定误差会在测量中被放大数十倍. 文献[10]则是在路面摆放参照物并测量其距离，选择合适的数学模型，拟合参照物距离和像素之间的对应关系，再利用此关系实时提取深度信息. 此方法在工程中也有广泛的运用，但需要较大的场地，且精度会受到长距离手动测量误差及拟合误差的影响. Wu 等[11]研究了一种激光与单目视觉相结合的测距方法；王荻等[12]则是利用光电经纬仪和 GPS 获得相机位姿信息，然后代入定位方程计算距离；另外，还有通过样本训练[13]或面积映射关系[14]等方法来实现单目深度信息提取的.

本文提出一种不受棋盘格标定板标定误差影响且不需要大场地及长距离手动测量的单目深度信息提取方法. 此方法通过检测立式标靶图像的

收稿日期：2014-05-14；**录用日期**：2014-07-25；**网络出版时间**：2014-08-11 14:01
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0272.html
基金项目：国家自然科学基金资助项目(51105021)；国家自然科学基金青年基金资助项目(51405008)；北京市自然科学基金资助项目(3133040)
作者简介：黄小云(1985—)，女，湖南邵阳人，博士生，hxly0820@126.com
*通讯作者：高峰(1955—)，男，河南禹州人，教授，gaof@buaa.edu.cn，主要研究方向为智能车辆.

引用格式：*黄小云，高峰，徐国艳，等. 基于单幅立式标靶图像的单目深度信息提取[J]. 北京航空航天大学学报，2015，41(4)：649-655. Huang X Y，Gao F，Xu G Y，et al. Depth information extraction of on-board monocular vision based on a single vertical target image[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：649-655 (in Chinese).*

角点并定位建立图像纵坐标像素与实际测量角度之间的映射关系,利用此映射关系结合投影几何模型得到深度信息.

1 深度信息提取原理

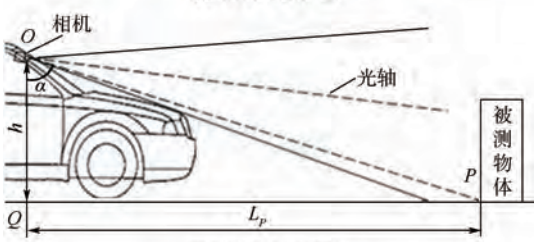
在智能车辆单目视觉系统中,相机一般安装于车内后视镜或车顶横向中间位置,而这两种安装位置的投影几何模型及后续图像处理方法并没有明显区别. 此外,城市化道路一般较平整,坡度不大,所以基于水平路面假设,以相机安装于车内后视镜处为例,相机安装位置及投影几何模型如图 1 所示. 当前可以通过测量及车辆检测、行人识别等算法得到的信息为:相机高度 h、被测物体在图像上的位置(一般为矩形区域,图像平面如图 2 所示,其中矩形 $ABCD$ 即为被测物体位置). 而被测物体在路平面上的实际成像角度 α 无法根据图像得到,所以也不能得到物体的距离 L_P(即 PQ 的长度).

如果可以建立图像平面中被测物体底部(即图 2 中 AB 所在位置)的像素 y_P 与投影几何模型中实际成像角度 α 之间的映射关系,即 $\alpha = F(y)$,由像素信息得到角度信息,那么图 1 中被测物体的距离 L_P 就可以通过边长 OQ 及角度 α 算得:

$$L_P = h \tan \alpha \tag{1}$$

立式标靶如图 3 所示,标靶以一定高度竖直放置于地面,靶面上共有 17 个共线角点,其中最低角点距离地面的竖直高度为 1 m,且相邻角点之间相差 0.05 m. 利用图 3 所示的立式标靶确定纵坐标像素与实际成像角度之间的映射关系 $\alpha = F(y)$.

(a) 相机安装位置



(b) 投影几何模型

图 1 相机安装位置及投影几何模型

Fig. 1 Installation position and projection geometry model of camera

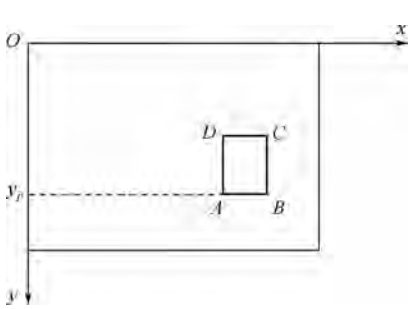


图 2 图像平面

Fig. 2 Image plane

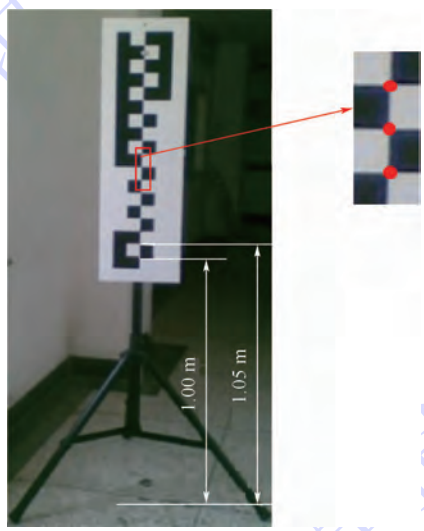


图 3 立式标靶

Fig. 3 Vertical target

图 4 为基于立式标靶的测量原理图. 将立式标靶放置于相机的正前方,采集标靶图像,通过角点检测算法得到各角点在图像上的坐标. 由于各角点的实际高度及标靶与相机的水平距离均已知,可得到其对应的成像角度. 以图 4 中的角点 S 为例,由 ΔOSG 可得到其对应的成像角度 $\alpha_S = \arctan \dfrac{D}{h - h_S}$,其中,$h_S$ 为角点 S 的高度,D 在放置标靶时测得. 据此建立图像纵坐标像素与实际成像角度之间的映射关系,再结合式(1)得到深度信息.

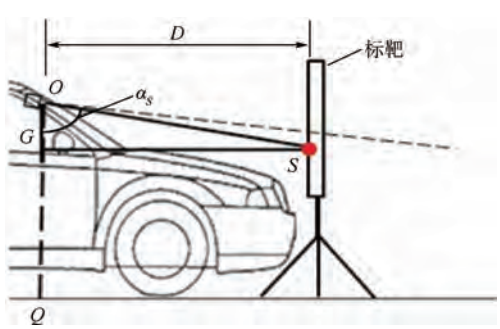


图 4 基于立式标靶的测量原理图

Fig. 4 Measuring principle diagram based on vertical target

北航学报 赠阅

2　立式标靶角点检测及定位

由于整个测量过程均是基于利用立式标靶建立图像纵坐标像素与实际成像角度之间的映射关系，所以立式标靶的角点检测精度及定位准确性成为影响测量精度的关键问题.

2.1　立式标靶图像

现有车型众多，车内后视镜位置离前保险杠的距离也各不相同，为了在方便定位的同时满足精度要求，立式标靶放置的位置在尽量靠近相机的同时，满足采集的图像中必须包含最低的角点且角点总数大于 8 个这一要求. 由于标靶尽量放置于相机正前方，所以设置标靶角点检测的感兴趣区为图像中间 1/3 的区域，如图 5 所示.

图 5　标靶图像

Fig. 5　Target image

2.2　初始角点检测

由于传统的角点检测算法会得到过多的非目标角点，容易受到环境的干扰，所以很难满足标靶图像的角点检测要求. 本文设计了一种满足小角度旋转的基于模板匹配的角点检测算法.

2.2.1　模板匹配

图 6 为 4 种模板匹配图像，其像素大小都是 20×20（此数值可依据图像分辨率的大小适当调整），上下左右平均分为 4 块，白色部分像素值为 255，黑色部分像素值为 0. 其中，模板 1 和 2 用来匹配检测标靶中目标角点的候选点，模板 3 和 4 用来匹配检测定位所需的参考点（参考点的定义将在第 2.3 节中介绍）.

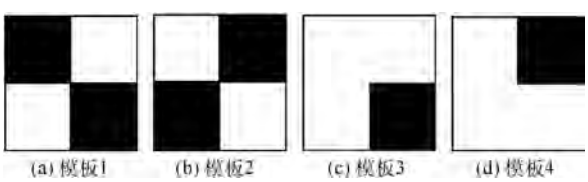


图 6　4 种模板匹配图像

Fig. 6　Four template images for matching

首先对感兴趣区分块处理，再对每个块运用最大类间方差法进行自适应二值化处理[15]，得到感兴趣区的二值图像. 在该区域内以逐点匹配的方式检测候选角点，匹配步骤如下：

步骤 1　图像中待匹配子图与模板进行差值处理（即像素值相同取 0，不同取 255），得到差值图像，差值处理如图 7 所示.

图 7　差值处理

Fig. 7　Differential treatment

步骤 2　以 $n \times n$ 个像素作为统计单元（本文中 n 取 7），对差值图像进行统计. 判定准则为

$$\frac{m}{n^2} > T \tag{2}$$

式中，m 为统计单元中像素值为 255 的点的个数；T 为密度阈值（本文中 T 取 0.32）.

如果存在满足式（2）的单元则认为此待匹配子图与模板不匹配，否则认为匹配. 利用模板 1 和 2 匹配检测出角点图像，二值化后目标角点检测结果图像如图 8 所示，共检测出 89 个点.

图 8　二值化后目标角点检测结果图像

Fig. 8　Images of target corners detection after binaryzation

2.2.2　候选点聚类及筛选

将所有认为匹配的点作为目标角点的候选点，因为存在相邻多个点都为匹配点或标靶外的区域也存在匹配点的情况，所以为了得到较准确的初始角点，将对候选点集进行聚类及筛选的操作.

聚类：遍历所有候选点，将所有点与点之间的横坐标及纵坐标之差同时小于阈值 T_1（本文中 T_1 取 10）的点记为同一小类，各小类的中心点坐标为

$$\begin{cases} x_c = \dfrac{\sum X}{k} \\ y_c = \dfrac{\sum Y}{k} \end{cases} \tag{3}$$

北航学报 赠阅

式中，$\sum X$ 和 $\sum Y$ 分别为同一类中所有点的横坐标和纵坐标之和；k 为此小类中点的个数.

筛选：遍历所有小类，将所有中心点横坐标之差小于 T_1 的小类记为同一个大类，然后保留包含小类数目最多的大类，剔除其他大类，并将保留下来的小类的中心点存为初始角点.

例如图 8 中共检测到 89 个点，聚类后得到了 8 个小类，经过筛选只存在 1 个大类，表 1 为保留的初始角点.

表 1 保留的初始角点

Table 1 Preserved initial corner points

编号	坐标
I_1	(439,379)
I_2	(438,335)
I_3	(439,247)
I_4	(439,203)
I_5	(439,158)
I_6	(439,113)
I_7	(438,68)
I_8	(438,23)

2.3 角点定位

角点定位是指判断检测到的各角点对应的实际高度. 3 个定位参考点如图 9 所示.

图 9 3 个定位参考点

Fig. 9 Three positioning reference points

由于存在光线等因素的影响，由图 9 与表 1 对比可以看出还有角点未检测到. 所以，当定位出最低角点后即可将其余角点全部补充完整，最终实现所有角点的定位，过程如下：

首先，搜索出初始角点中横坐标最大值 x_{max}，利用图 6 中的模板 3，以第 2.2.1 节中的匹配方法，在坐标小于 x_{max} 的感兴趣区域内从上至下、从右至左检测匹配子图，一旦匹配即停止检测，记录坐标 (x_a, y_a)，此点即为图 9 中参考点①或②. 计算 $x_{max} - x_a$，作为角点的间距参考值.

然后，在点 $(x_a - (x_{max} - x_a)/2, y_a)$ 的左下感兴趣区内利用图 6 中模板 4 从上至下、从右至左检测匹配的子图，一旦匹配即停止检测，点 (x_b, y_b) 即为参考点③.

最后，如果 $y_b - y_a$ 接近 $x_{max} - x_a$ 的 3 倍则认为 (x_a, y_a) 为参考点①，如果 $y_b - y_a$ 与 $x_{max} - x_a$ 相差很小，则认为 (x_a, y_a) 为参考点②. 如果没有检测到两个参考点，则需要重新放置标靶再次检测.

目前参考点③、初始角点中的最低点及角点的间距参考值均为已知，通过计算参考点③与初始角点中的最低点在 y 方向上的位置差值，再与角点间距参考值进行比对，即可判断出初始角点中的最低点是否为标靶最低的角点. 如果初始角点中的最低点判断为标靶最低的角点，则利用角点间距参考值将其余初始角点补充完整；如果不是最低的角点，则将点 $(x_{max}, y_b - 2(x_{max} - x_a))$ 定位为最低角点，再将其余初始角点补充完整. 表 2 为初始角点及补充角点的坐标及定位高度，角点 C 即为补充的定位角点.

表 2 初始角点及补充角点的坐标及定位高度

Table 2 Coordinates and positioning height of initial corner points and complementary corner point

编号	坐标	定位高度/m
I_1	(439,379)	1.00
I_2	(438,335)	1.05
C	(439,291)	1.10
I_3	(439,247)	1.15
I_4	(439,203)	1.20
I_5	(439,158)	1.25
I_6	(439,113)	1.30
I_7	(438,68)	1.35
I_8	(438,23)	1.40

2.4 亚像素级角点检测

为了更精确地提取角点位置，运用 openCV 中的 cvFindCornerSubPix( ) 函数，以初始角点为基准将其更新为亚像素级角点，亚像素级角点检测结果如表 3 所示. cvFindCornerSubPix( ) 函数的具体用法见文献[16].

表 3 亚像素级角点检测结果

Table 3 Subpixel corner point detection results

编号	初始坐标	亚像素级角点坐标
I_1	(439,379)	(438.387 91,378.317 99)
I_2	(438,335)	(438.592 53,334.891 36)
C	(439,291)	(438.714 80,290.474 18)
I_3	(439,247)	(438.826 14,246.784 79)
I_4	(439,203)	(438.821 87,202.199 31)
I_5	(439,158)	(438.652 92,157.331 95)
I_6	(439,113)	(438.432 56,112.366 84)
I_7	(438,68)	(438.042 79,67.456 29)
I_8	(438,23)	(437.650 18,22.745 09)

北航学报 赠阅

3　建立像素与成像角度的映射关系

分析图 4 可知,当相机和立式标靶安装或放置完毕后,距离 D 及高度 h 的值将固定不变,各角点的实际成像角度可通过计算得出. 此外,各角点的像素与实际标靶上的角点也已经通过定位一一对应. 所以各角点的实际成像角度值与像素值即可完全确定.

例如,采集图 8 时,距离 D = 1 500 mm,高度 h = 1 320 mm,计算得到像素值与实际成像角度的映射关系如图 10 所示. 可以看出,实际成像角度与 y 方向像素的映射关系可以近似为一条直线,但是为进一步提高精度,采用相邻点的分段直线来表示映射关系,从而减少拟合误差.

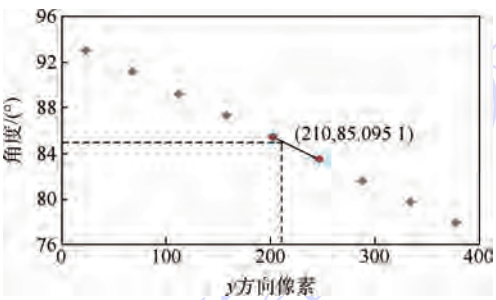


图 10　像素值与实际成像角度的映射关系

Fig. 10　Mapping relation between image row pixel values and actual imaging angles

在计算某像素点对应的成像角度时,需先判断此像素点的 y 值所在像素值范围,然后利用相邻两点构造的直线方程计算成像角度. 例如,当 y = 210 时,利用相邻两点建立的直线方程可求解对应的实际成像角度 α 为 85.095 1°,然后通过式(1)即可求得距离结果.

4　实验验证

为验证方法的可行性及精度,在 VS2010 + OpenCV2.2 环境下,按照上述方法编写程序并调试后,分别进行了静态及动态实验. 将 CCD (Charge-coupled Device)相机安装于车内后视镜处,俯仰角为 5°,距地面高度 h 为 1 320 mm,立式标靶的靶面与相机距离 D 为 1 800 mm. 图像分辨率为 752 × 480.

4.1　静态实验

首先将标靶放置于相机前方,使立式标靶的角点在图像的感兴趣区内. 运行编写好的程序,完成角点检测及定位,标靶角点检测及定位界面如图 11 所示,亚像素级角点坐标及对应实际成像角度如表 4 所示. 然后将实验被测物体按不同的距离放置于路面,并记录其实际距离值. 选取采集的被测物体图像,手动选取出物体底部位置,通过上述程序运算后得到被测物体的深度信息. 部分实验图像如图 12 所示,实际距离和计算距离对比如表 5 所示.

图 11　标靶角点检测及定位界面

Fig. 11　Target corner detection and positioning interface

表 4　亚像素级角点坐标及对应实际成像角度

Table 4　Coordinates and actual imaging angles of subpixel corner points

编号	亚像素级角点坐标	成像角度/(°)
1	(363.838 84,401.420 47)	79.919 47
2	(364.333 74,364.973 36)	81.469 30
3	(364.413 33,329.076 60)	83.031 81
4	(365.247 28,291.402 07)	84.604 81
5	(365.594 76,254.206 18)	86.186 00
6	(365.937 74,216.781 45)	87.773 03
7	(366.441 19,179.210 14)	89.363 48
8	(366.855 07,142.176 82)	90.954 77
9	(367.151 34,104.834 32)	92.544 73
10	(367.534 64,67.900 67)	94.130 78
11	(367.822 54,30.959 16)	95.710 52

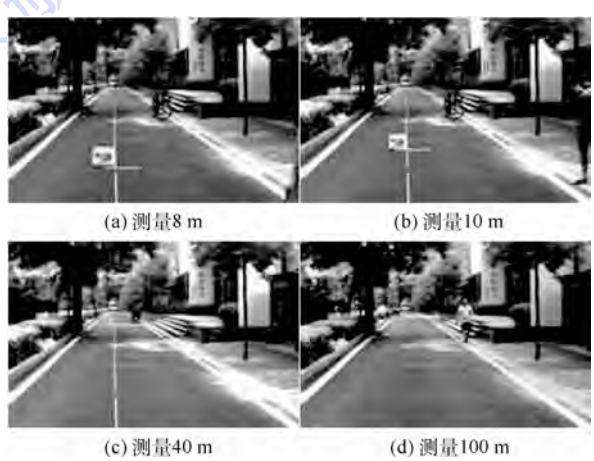


图 12　部分实验图像

Fig. 12　Part of experimental images

表 5 实际距离和计算距离对比

Table 5 Comparison between actual distance and calculated distance

实际距离/m	y 像素值	计算距离/m	误差/%
5.53	479	5.549 6	0.354 4
8	384	8.025 8	0.322 5
10	341	10.043 7	0.437 0
20	253	20.069 9	0.349 5
30	224	29.836 8	0.544 0
40	209	39.840 3	0.399 3
50	200	49.854 7	0.290 6
60	194	59.887 4	0.187 7
80	186	81.842 3	2.302 9
100	182	100.207 8	0.207 8
140	177	139.631 7	0.263 1

从测量结果中可以看出,该方法总体误差很小,一般小于 1%. 由于 80 m 时误差突然达到 2.302 9%,所以又进行了进一步分析. 将 y 像素值调整为 187,利用表 4 的映射关系,计算得到深度值为 78.262 5 m. 因此一个像素所占的距离为 3.579 8 m,而且这个数值会随着实际距离的增加而变大. 所以,距离越远对障碍物检测精度的要求越高,可能的误差也越大. 但对于智能车系统,离本车越远危险系数越低,所以测量的精度要求也可以有所降低. 总的来说,实验结果显示上述方法已经可以满足系统的要求.

4.2 动态实验

为检验此方法的实时性及在实际行驶状态下的效果,进行了动态实验,动态实验图像如图 13 所示. 实验中采集图像的帧率为 15 帧/s. 实验主要是利用上述方法结合基于机器学习的车辆识别技术实现前方车辆识别及深度信息提取. 实验表明,从采集图像到处理完成共消耗不到 50 ms,可以满足实时性要求. 而且同时与雷达测量结果进行对比发现,两种测量值相差保持在 ±0.5 m 以内,所以此方法也可以满足智能车单目视觉系统准确性的要求.

图 13 动态实验图像

Fig. 13 Dynamic experimental images

5 结 论

利用单幅立式标靶图像、焦点检测与定位算法建立了实际成像角度与 y 方向像素的映射关系. 再结合投影几何模型,实现了单目视觉图像的深度信息提取. 通过分析与比较,得到以下结论:

1) 所设计的依据二值化差值图像统计单元中不匹配点密度的模板匹配算法可快速检测目标角点. 结合 OpenCV 库函数,实现了亚像素级的角点检测及定位,进一步提高了检测精度.

2) 利用检测立式标靶中高度已知的角点可建立图像纵坐标像素与实际成像角度之间的映射关系,而且此关系近似于一条直线.

3) 上述定位方法只需测量相机高度及相机与标靶的距离,采集一副标靶图像即可,操作简单且测量精度及实时性都可满足智能车辆单目视觉系统要求.

本文暂未考虑相机抖动造成的俯仰角变化以及斜坡路面的影响,因此在不平路面及驶入或驶出斜坡之前的工况下测量精度略显不足,还有继续改进的空间.

参考文献 (References)

[1] Xu F F, Wang B, Zhou Z Q, et al. Real-time lane detection for intelligent vehicles based on monocular vision[C]//Proceedings of the 31st Chinese Control Conference. Washington D. C.: IEEE, 2012: 7332-7337.

[2] Sivaraman S, Trivedi M M. Combining monocular and stereo-vision for real-time vehicle ranging and tracking on multilane highways[C]//Proceedings of International IEEE Conference on Intelligent Transportation Systems. Washington D. C.: IEEE, 2011: 1249-1254.

[3] 王浩,许志闻,谢坤,等. 基于 OpenCV 的双目测距系统[J]. 吉林大学学报,2014,32(2):188-194.
Wang H, Xu Z W, Xie K, et al. Binocular measuring system based on OpenCV[J]. Journal of Jilin University, 2014, 32(2): 188-194 (in Chinese).

[4] Sun W, Chen L, Hu B, et al. Binocular vision-based position determination algorithm and system[C]//Proceedings of the 2012 International Conference on Computer Distributed Control and Intelligent Environmental Monitoring. Piscataway: IEEE Computer Society, 2012: 170-173.

[5] 于乃功,黄灿,林佳. 基于单目视觉的机器人目标定位测距方法研究[J]. 计算机测量与控制,2012,20(10):2654-2660.
Yu N G, Huang C, Lin J. Target location and ranging based on monocular vision robot[J]. Computer Measurement & Control, 2012, 20(10): 2654-2660 (in Chinese).

[6] 吴刚,唐振民. 单目式自主机器人视觉导航中的测距研究[J]. 机器人,2010,32(6):828-832.
Wu G, Tang Z M. Distance measurement in visual navigation of monocular autonomous robots[J]. Robot, 2010, 32(6): 828-832 (in Chinese).

[7] 鲁威威,肖志涛,雷美琳. 基于单目视觉的前方车辆检测与测距方法研究[J]. 视频应用与工程,2011,35(1):125-128.
Lu W W, Xiao Z T, Lei M L. Study on method of preceding vehicles detection and distance measurement based on monocular vi-

sion[J]. Video Application & Project,2011,35(1):125-128(in Chinese).

[8] 战强，吴佳. 未知环境下移动机器人单目视觉导航算法[J]. 北京航空航天大学学报，2008，34(6)：613-617.
Zhan Q, Wu J. Monocular vision-based navigation algorithm for mobile robots in unknown environments[J]. Journal of Beijing University of Aeronautics and Astronautics, 2008, 34(6): 613-617(in Chinese).

[9] 王青. 基于机器视觉的车辆距离测量技术[D]. 南京：南京理工大学，2008.
Wang Q. Vehicle distance measurement technology based on machine vision[D]. Nanjing: Nanjing University of Science and Technology, 2008(in Chinese).

[10] Wu C F, Lin C J, Lee C Y, et al. Applying a functional neurofuzzy network to real-time lane detection and front-vehicle distance measurement[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, 2012, 42(4): 577-589.

[11] Wu K, Di K C, Sun X, et al. Enhanced monocular visual odometry integrated with laser distance meter for astronaut navigation[J]. Sensors, 2014, 14(3): 4981-5003.

[12] 王荻，王洁，付小宁. 基于单目图像和方向的测距系统及性能分析[J]. 光子学报，2011，40(7)：1112-1116.
Wang D, Wang J, Fu X N. A passive ranging system based on image sequence from single lens and the respective imaging direction along with its performance[J]. Acta Photonica Sinica, 2011, 40(7): 1112-1116(in Chinese).

[13] 赵信宇，朱晓蕊，余锦全. 单幅图像深度信息的提取[J]. 制造业自动化，2010，32(3)：15-17.
Zhao X Y, Zhu X R, Yu J Q. Learning depth information from monocular image[J]. Manufacturing Automation, 2010, 32(3): 15-17(in Chinese).

[14] 韩延祥，张志胜，戴敏. 用于目标测距的单目视觉测量方法[J]. 光学精密工程，2011，19(5)：1110-1117.
Han Y X, Zhang Z S, Dai M. Monocular vision system for distance measurement based on feature points[J]. Optics and Precision Engineering, 2011, 19(5): 1110-1117(in Chinese).

[15] Otsu N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.

[16] 刘瑞祯，于仕琪. OpenCV 教材[M]. 北京：北京航空航天大学出版社，2007：249-250.
Liu R Z, Yu S Q. OpenCV teaching material[M]. Beijing: Beihang University Press, 2007: 249-250(in Chinese).

Depth information extraction of on-board monocular vision based on a single vertical target image

HUANG Xiaoyun, GAO Feng*, XU Guoyan, DING Nenggen, XING Longlong

(School of Transportation Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: Calculating obstacle distance for on-board monocular vision system on intelligent vehicle was investigated when the obstacle has been detected. An extraction method of depth information only using a single target image without any internal camera parameters was developed. The mapping relation between image row pixel values and the actual imaging angles was established with the image of vertical target, which was placed in the front of a camera. The obstacle depth information was extracted in real time by combining the projection geometry model. Given the characteristic of vertical target image, an algorithm of sub-pixel corner detection and location was designed, which includes region of interest setting, template matching, candidate points clustering and screening and precise location, etc. Experimental results show that the method has high precision and real-time performance. Compared with the method of putting reference on the road, it does not need large calibration site and could avoid the data fitting error. And the method also has a simple calibration procedure with a single image, which is suitable for practical application.

Key words: machine vision; monocular distance detection; vertical target; intelligent vehicle; corner detection

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0299

基于流动控制的水下航行体流体动力特性分析

张孝石，王聪*，曹伟，赵成功

（哈尔滨工业大学 航天学院，哈尔滨 150001）

摘 要：利用 CFX 对水下航行体流体动力特性进行了数值模拟研究. 求解了混合介质的剪切压力输运湍流模型控制方程、雷诺-平均奈维尔-斯托克斯方程以及各相间的质量输运方程，采用三维数值模拟方法对比分析了不同流动控制方案的水下航行体空泡形态特征、表面压力分布和阻力系数变化特性，讨论了不同参数对减阻效果的影响. 结果表明，对于有横流作用的水下垂直发射航行体，多相流动控制不仅可以降低通气空泡的不对称性和航行体阻力，同时可以均匀迎流面和背流面压力，从而实现航行体多相流空泡形态及表面压力特性等的控制. 此外，通气口的位置对减阻效果具有显著影响.

关 键 词：流动控制；通气口；空泡形态；流体动力；多相流；数值计算
中图分类号：TJ762.4
文献标识码：A **文章编号**：1001-5965(2015)04-0656-07

水下航行体在垂直发射过程中，由于受到牵连速度影响形成不对称空泡，会造成一个弯矩，弯矩过大将严重影响垂直发射航行体的弹道和载荷. 了解水下航行体这种特殊的受力情况对水下发射动力学分析有很大意义. 通过多相流动控制方式可以稳定空泡并控制空泡形态，受多相流动控制的航行体周围会产生空气、水、水蒸气三相混合介质，航行体周围的流体动力特性更加复杂. 国内外学者对空化问题做了大量的试验研究与数值计算. Reichardt[1] 于 1945 年首次提出通过人工通气的方法生成超空泡，后来的通气空泡研究均基于这一思想；Silberman 和 Song[2] 通过试验研究了通气空泡的振荡规律，得到了通气量与空泡形态和泡内压力之间的关系. Lindau 等[3] 建立了一套高雷诺数下的超空泡航行体多相流计算程序. 闵景新等[4] 通过数值模拟研究了垂直发射过程中局部空泡对流体动力的影响以及不同头型对空泡的抑制作用. Kawakami 和 Arndt[5] 通过水洞试验分析了通气空泡形态及空泡闭合处现象，并得到了水洞阻塞效应作用下的最小空化数. Matveev 和 Miller[6] 通过水洞试验研究了空泡在不同空化数下的稳定长度，并分析了空泡在没有泄气和有泄气情况下空泡的长度变化. Ramesh 等[7] 通过边界元法研究了不同通气孔直径和不同空化器形状条件下所产生的空泡波动问题. 黄彪等[8] 通过试验研究了轴对称体空化水动力脉动特性. 王一伟等[9] 通过数值模拟研究了航行体出水载荷问题，提出了出水溃灭压力物理模型. 目前为止，流动控制方法主要分为主动控制法和被动控制法. 微涡流发生器（MVG，Micro-Vortex Generator）被视为现阶段亚声速、超声速及高超声速进气道中最具有应用前景的被动流动控制装置，并成为近几年

收稿日期：2014-05-26；**录用日期**：2014-08-11；**网络出版时间**：2014-12-10 17:05
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0299.html
基金项目：国家自然科学基金资助项目(51149003)；中央高校基本科研业务费专项资金资助项目(HIT.NSRIF.2013033)
作者简介：张孝石(1987—)，男，黑龙江哈尔滨人，博士生，hitzxs@gmail.com
*通讯作者：王聪(1966—)，男，内蒙古呼和浩特人，教授，alanwang@hit.edu.cn，主要研究方向为运动体多相流动力学、水下发射技术、航天大型柔性结构动力学、振动与控制.

引用格式：张孝石，王聪，曹伟，等. 基于流动控制的水下航行体流体动力特性分析[J]. 北京航空航天大学学报，2015，41(4)：656-662. Zhang X S，Wang C，Cao W，et al. Hydrodynamic features of underwater vehicles based on flow control[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：656-662 (in Chinese).

边界层分离控制研究的热点.研究人员利用涡流发生器延缓边界层流体分离,从而进一步增加机翼的升力并减小阻力[10-11].Wendt 和 Hingst[12]模拟了二维翼型尾流发生器作用下尾涡的脱落过程,建立了尾涡的峰值涡量、涡流发生器尺寸以及来流条件之间的关系.Lu 等[13-14]通过实验和数值模拟研究了涡流发生器周围的漩涡特征和流场结构.于彦泽等[15]进行了大型飞机流动控制的减阻机理研究.季路成和林峰[16]通过数值仿真研究了绕凸包主动流动控制.虽然流动控制已经成为流体工程领域的研究热点,但水下运动体多相流动控制相关的文献仍然较少.

本文针对潜射航行体水下运动过程多相流场及流体动力特点,基于流动控制思想,通过改变通气口在航行体上的轴向位置和周向位置实现对航行体多相流空泡形态及表面压力特性等的控制,进而改善潜射航行体的载荷和弹道.

1　数学模型

1.1　流体动力学控制方程

由于数值计算时通气质量对应的气体流速较低,所以不考虑计算过程中气体可压缩性带来的影响,假设流体介质是不可压缩的.

连续性方程为

$$\frac{\partial \rho_{\mathrm{n}}}{\partial t}+\frac{\partial\left(\rho_{\mathrm{n}} u_j\right)}{\partial x_j}=0 \tag{1}$$

式中,$\rho_{\mathrm{n}}=(1-\alpha_2-\alpha_3)\rho_1+\alpha_2\rho_2+\alpha_3\rho_3$ 为混合物的密度,ρ_1,ρ_2 和 ρ_3 为空气、水和水蒸气的密度,α_1,α_2 和 α_3 为空气、水和水蒸气的体积分数,$\alpha_1+\alpha_2+\alpha_3=1$;$t$ 为时间;u_j 为速度分量;x_j 为坐标值.

动量守恒方程为

$$\frac{\partial\left(\rho_{\mathrm{n}} u_i\right)}{\partial t}+\frac{\partial\left(\rho_{\mathrm{n}} u_i u_j\right)}{\partial x_j}=-\frac{\partial p}{\partial x_i}+\frac{\partial}{\partial x_j}\left[\left(\mu_{\mathrm{n}}+\mu_{\mathrm{t}}\right)\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right)\right] \tag{2}$$

式中,$\mu_{\mathrm{n}}=(1-\alpha_2-\alpha_3)\mu_1+\alpha_2\mu_2+\alpha_3\mu_3$ 为混合物的动力黏性系数,μ_1,μ_2 和 μ_3 分别为空气、水和水蒸气的黏性系数;μ_{t} 为湍流黏性系数;p 为气泡周围流场压强.

1.2　湍流模型

湍流是流场结构中的一种强非线性流体状态,是空化流场的主要特征之一.在空泡的发展尤其是尾部闭合阶段均存在着湍流现象.

基于 Baseline k-ω(湍动能-湍动耗散率)的剪切压力输运(SST,Shear Stress Transport)湍流模型解决了湍流剪切应力的输运问题,综合了近壁面 k-ε(耗散率)模型的稳定性及边界层外部 k-ω 模型的独立性,不仅可以准确地预测各种来流,还可以在不同压力梯度情况下精确模拟分离现象.基于 Baseline k-ω 的 SST 湍流模型控制方程[17]为

$$\frac{\partial\left(\rho_{\mathrm{n}} k\right)}{\partial t}+\frac{\partial\left(\rho_{\mathrm{n}} u_j k\right)}{\partial x_j}=\frac{\partial}{\partial x_j}\left[\frac{\partial k}{\partial x_j}\left(\mu+\frac{\mu_{\mathrm{t}}}{\sigma_{k3}}\right)\right]+G_k+G_{\mathrm{b}}-\beta'\rho_{\mathrm{n}} k\omega \tag{3}$$

$$\frac{\partial\left(\rho_{\mathrm{n}}\omega\right)}{\partial t}+\frac{\partial\left(\rho_{\mathrm{n}} u_j\omega\right)}{\partial x_j}=\frac{\partial}{\partial x_j}\left[\left(\mu_{\mathrm{n}}+\frac{\mu_{\mathrm{t}}}{\sigma_{\omega3}}\right)\frac{\partial\omega}{\partial x_j}\right]+\alpha_3\frac{\omega}{k}P_k+P_{\omega\mathrm{b}}+2(1-Y_1)\rho_{\mathrm{n}}\frac{1}{\sigma_{\omega2}\omega}\frac{\partial k}{\partial x_j}\frac{\partial\omega}{\partial x_j}-\beta_3\rho_{\mathrm{n}}\omega \tag{4}$$

式中,σ_{k3},$\sigma_{\omega2}$,$\sigma_{\omega3}$ 和 β_3 为模型经验常数;μ 为流体的黏性系数;G_k 为黏性力产生的湍动能生成项;G_{b} 为浮力产生的湍动能生成项;β' 为系数,$\beta'=0.09$;P_k 为湍动能项;$P_{\omega\mathrm{b}}$ 为 ω 的产生项;Y_1 为混合函数.

SST 湍流模型中的涡黏度限制方程为

$$\nu_{\mathrm{t}}=\frac{b_1 k}{\max\left(b_1\omega,\Omega Y_2\right)} \tag{5}$$

式中,b_1 为系数,$b_1=0.31$;$\nu_{\mathrm{t}}=\mu_{\mathrm{t}}/\rho$,为采用的 k-ω 模型中的表达形式,其中,ρ 为液体密度;Ω 为应变率的不变测度;Y_2 为混合函数.

1.3　空化模型

当流场中发生空化现象时,水蒸气和水两相之间的质量转换输运方程为

$$\frac{\partial}{\partial t}\left(\alpha_i\rho_i\right)+\nabla\cdot\left(\alpha_i\rho_i u_i\right)=\Psi_i+\Gamma_i \tag{6}$$

式中,∇为 Nabla 算子;Ψ_i 为广义源;Γ_i 为进入第 i 相的单位体积质量源.

空化模型采用基于 Rayleigh-Plesset 方程的简化形式:

$$\frac{\mathrm{d}R_{\mathrm{g}}}{\mathrm{d}t}=\pm\sqrt{\frac{2(p_{\mathrm{c}}-p)}{3\rho}} \tag{7}$$

式中,R_{g} 为气泡半径;p_{c} 为饱和蒸气压.

若单位体积内气泡数量为 N_{g},则气体体积分数为

$$\gamma_{\mathrm{b}}=V_{\mathrm{g}}N_{\mathrm{g}}=\frac{4}{3}\pi R_{\mathrm{g}}^3 N_{\mathrm{g}} \tag{8}$$

式中,V_{g} 和 R_{g} 分别为单个气泡的体积和半径.

气泡生长和凝结的质量 $\dot{m}$ 转换表达式为

$$\dot{m}=C\frac{3\gamma_{\mathrm{b}}\rho_{\mathrm{g}}}{R_{\mathrm{g}}}\sqrt{\frac{2\left|p_{\mathrm{c}}-p_\infty\right|}{3\rho_{\mathrm{l}}}}\operatorname{sgn}(p_{\mathrm{c}}-p) \tag{9}$$

式中,ρ_{g} 和 ρ_{l} 分别为气体和液体密度;p_∞ 为相对

北航学报 赠阅

压力;C 为经验常数. 在一般的空化问题中,水蒸汽凝结速率远小于空化速率. 因此,在空化状态和凝结状态下的经验常数不同,在 CFX 中分别为 $C_{vap}=50$, $C_{cond}=0.01$.

2 数值计算模型

2.1 计算模型设置及网格划分

计算模型为半球头回转体,其直径为 D,总长度为 L;计算流域为长度为 $13L$,径向为 $10D$ 的三维圆柱流域,入口与回转体头部顶点的距离为 $2L$,出口与回转体底部的距离为 $10L$. 计算流域入口设置为速度入口,出口边界条件为压力出口,为减小壁面带来的阻塞效应,将流域壁面边界条件设置为速度入口,航行体表面通气口边界设置为质量流量入口,航行体剩余的表面均设置为不可滑移壁面,计算模型及流域边界条件如图 1 所示. 表 1为通气口位置分布. 通气口长度为 1/4 弹体柱段周长,宽度为 1/80 弹体柱段周长,按照通气口在迎流面轴向位置的不同,依次将试验方案命名为方案 1、方案 2 和方案 3. 方案 4 和方案 5 的通气口轴向位置与方案 2 的相同,通气口分别处于背流面和侧流面.

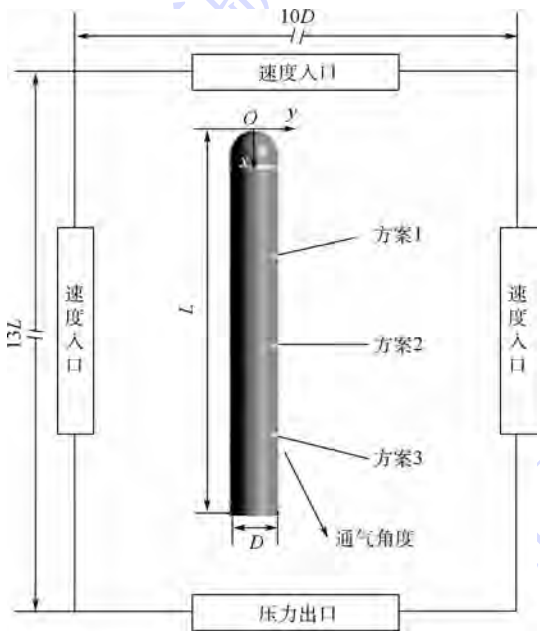


图 1 计算模型及流域边界条件

Fig. 1 Computational model and boundary conditions of flow

表 1 通气口位置

Table 1 Positions of ventilation holes

通气口	方案 1	方案 2	方案 3	方案 4	方案 5
轴向位置	$(3/10)L$	$(11/20)L$	$(4/5)L$	$(11/20)L$	$(11/20)L$
周向位置	迎流面	迎流面	迎流面	背流面	侧流面

计算网格划分如图 2 所示. 将航行体头部、通气口处以及航行体尾部区域等压力梯度变化较大的流场区域网格进行加密,航行体近壁面网格第 1 层厚度为 2×10^{-4} m,远场网格进行稀疏化处理,综合考虑计算效率和计算精度,选取计算网格数量为 60 万.

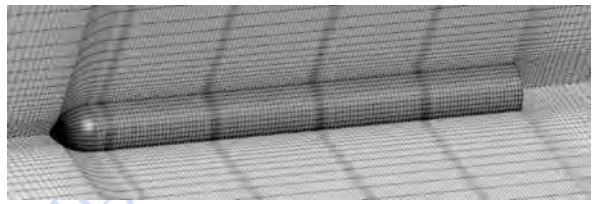


图 2 计算网格划分

Fig. 2 Mesh of model

2.2 通气角度的选择

文献[18]分析了通气角度对空泡长度和半径的影响. 文献[19]研究了通气角度对黏性阻力及压差阻力的影响,并发现不同通气角度时的阻力系数基本相同. 根据上述文献的研究结论,为了在保证通气方向尽量与流速方向相近的同时使气体从通气口顺利流出,计算过程中选取的通气角度为 20°.

3 数值计算方法验证与计算结果分析

3.1 数值计算方法验证

为验证计算方法的有效性,对图 1 中的半球头回转体在无重力条件下的空化问题进行数值计算,自然空化数选择为 0.3. 图 3 为水相体积分数为 α 时的等值云图,将数值计算得到的航行体表面压力分布曲线与文献[20]中试验测得的表面压力进行对比,数值计算结果与试验结果[20]对比如图 4所示. 从表面压力系数分布曲线中可以看出,数值计算结果与试验结果吻合较好,证实了计算方法的有效性.

图 3 水相体积分数为 α 时的等值云图

Fig. 3 Contour at liquid volume fraction α

3.2 通气空泡形态分析

在上述计算模型的基础上,开展轴向位置和周向位置水下垂直发射航行体空泡形态及流体动力问题的数值模拟研究. 水下航行体所在的笛卡

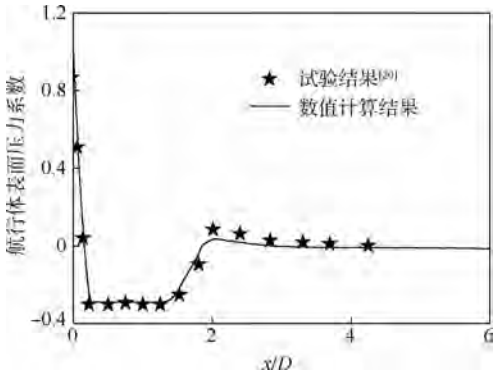


图 4　数值计算结果与试验结果[20]对比

Fig. 4　Comparison of numerical results and test results[20]

尔坐标系如图 1 所示,计算时航行体垂直于水面,因此重力方向为 x 轴正方向,设在航行体头部 $x=0$处的绝对压力值为 p_0,则 x 轴方向上航行体某一位置受到的绝对压力为 $p_0+\rho g x$,g 为重力加速度;数值计算过程中的饱和蒸汽压为 3 510 Pa,参考压力为 101 325 Pa.

图 5 为不同通气口轴向位置的航行体空泡形态. 从图 5(a)中可以看出,在有横流无通气口的情况下,航行体的迎流面和背流面空泡出现明显的不对称现象. 在迎流面加上通气口后,迎流面和背流面的泡长差明显减小. 由于图 5(b)中通气口比较靠前,在航行体尾部仍存在泡长差;图 5(c)中没有泡长差,形成一个包裹航行体的完整空泡;图 5(d)中通气口在迎流面空泡末端以下,所以对通气口上部分无空泡区域没有起到补充作用. 图 6 为不同通气口轴向位置的航行体气相体积分数,对比 3 种方案可知,方案 2 与另外2 种方案相比,空泡可以更好地包裹航行体.

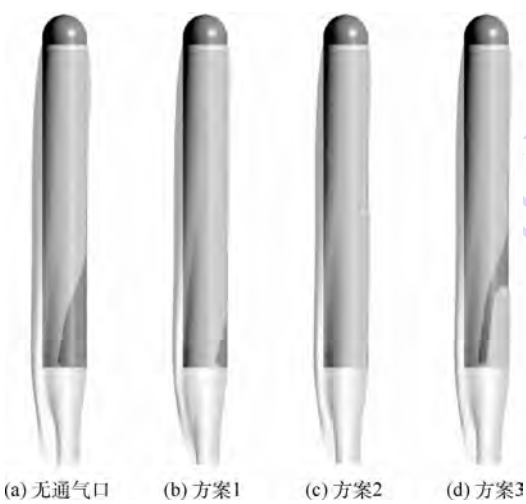


图 5　不同通气口轴向位置的航行体空泡形态

Fig. 5　Cavity shapes of vechicles with different ventilation hole axial positions

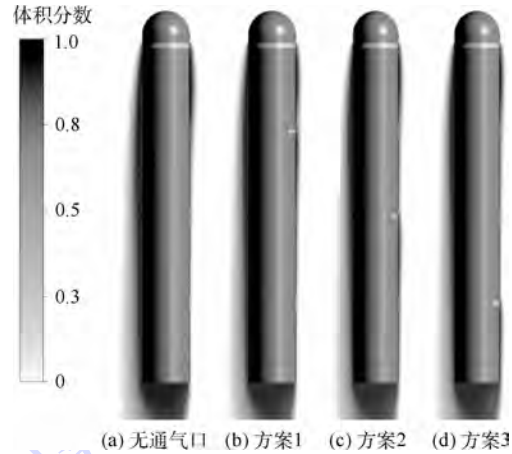


图 6　不同通气口轴向位置的航行体气相体积分数

Fig. 6　Air volume fraction of vehicles with different ventilation hole axial positions

3.3　航行体阻力变化分析

3.3.1　轴向位置的影响

在计算过程中改变通气口在航行体上的轴向位置,对比分析航行体的阻力系数. 计算结果表明,在航行体迎流面不同位置的通气口能够起到不同程度的减阻效果.

不同通气口轴向位置的航行体阻力系数对比如图 7 所示. 在相同通气率情况下,方案 1 由于通气口位置比较靠前,阻力系数较大,随着通气口位置的后移,减阻效果越来越明显. 在不同通气率情况下,随着通气率的增大,方案 1 的压差阻力系数呈现先增大后减小的趋势,黏性阻力系数单调减小,由于总阻力系数中压差阻力系数占主要地位,因此其总阻力系数也呈现先增大后减小的趋势;方案 2的压差阻力系数和黏性阻力系数呈单调减小趋势,且黏性阻力系数在相同通气率情况下比另外 2 个方案的小;方案 3 的黏性阻力系数在通气率较小时比方案 1 的小,随着通气率的增大逐渐大于方案 1. 这是由于经过方案 1 的流动控制后,迎流面的空泡长度增大,而方案 3 中通气孔的位置比较靠后,对迎流面空泡长度的影响较小. 综合来看,方案 2 和方案 3 均起到了减阻的效果,方案 2 的减阻效率更高.

3.3.2　周向位置的影响

通过对不同通气口周向位置航行体的分析可知,周向位置同样会对减阻控制效果产生影响,方案 2 流动控制效果较好,因此选择方案 2 中通气口所在的轴向位置对比分析不同通气口周向位置的减阻效果,方案 2、方案 4 和方案 5 中通气口分别处于迎流面、背流面和侧流面. 方案 4 和方案 5 中通气口的位置如图 8 所示,不同通气口周向位置的航行体阻力系数对比如图 9 所示. 可以看出,随着通气率的增加,方案 2 的压差阻力系数和黏

北航学报 赠阅

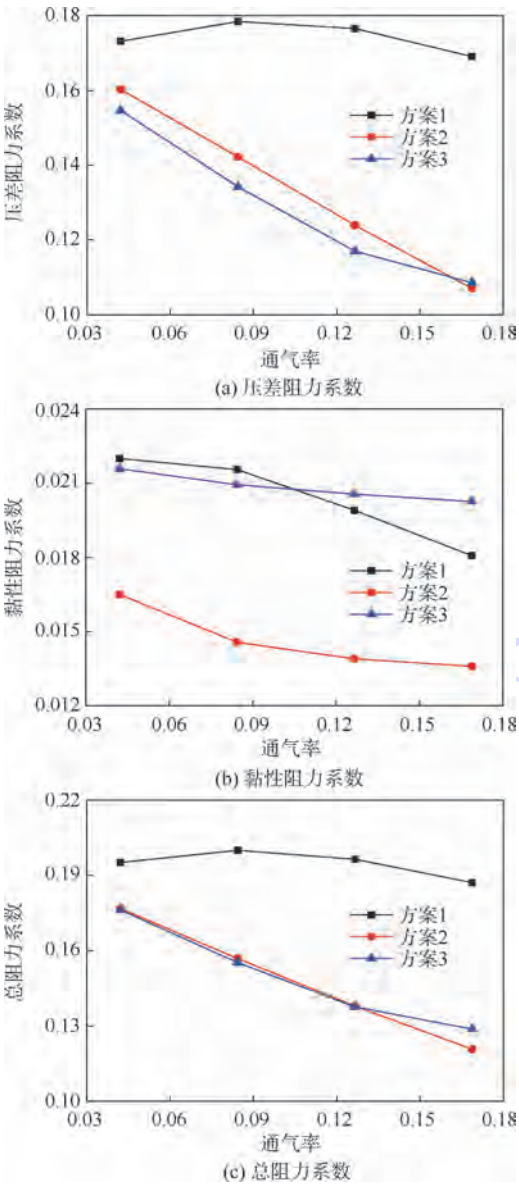


图 7 不同通气口轴向位置的航行体阻力系数对比

Fig. 7 Comparison of drag coefficients of vehicles with different ventilation hole axial positions

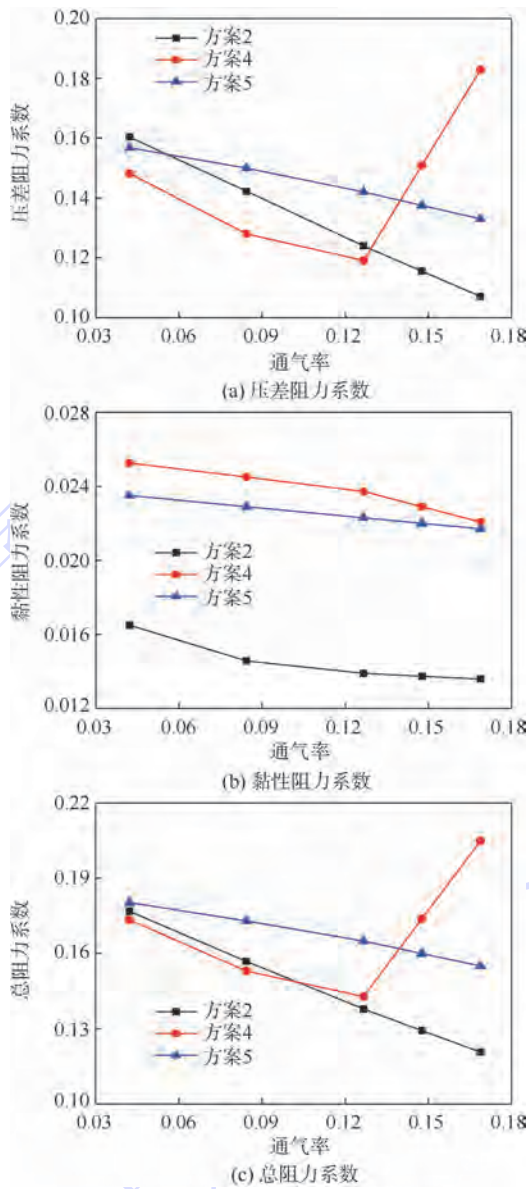


图 9 不同通气口周向位置的航行体阻力系数对比

Fig. 9 Comparison of drag coefficients of vehicles with different ventilation holecircumferential positions

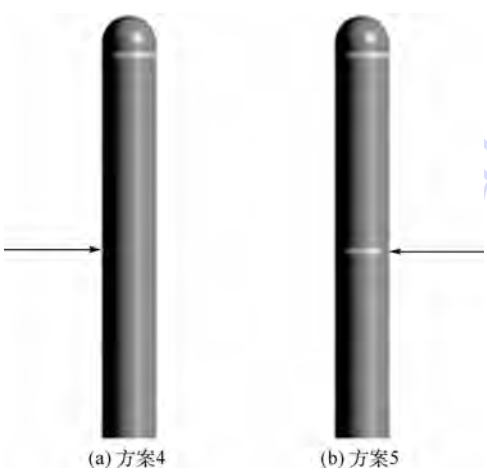


图 8 方案 4 和方案 5 中通气口的位置

Fig. 8 Positions of ventilation holes in Scheme 4 and Scheme 5

性阻力系数与另两个方案相比最小,方案 4 的压差阻力系数随着通气率的增加先降低后升高. 综合以上分析可知,方案 2 的总阻力系数变化稳定,并且当通气率最大时达到最小值.

3.3.3 航行体表面压力分析

对比分析采用方案 2 控制后和无控制情况下的表面压力分布,提取航行体轴向(9/10)L 处周向表面压力系数,航行体表面压力分布对比如图 10 所示,其中 z 轴负方向为迎流面,正方向为背流面. 可以看出,控制后航行体的迎流面和背流面最大表面压力系数均减小;压力系数曲线均存在波动现象,但控制后的曲线波动减小,这主要是因为控制后空泡分布对于压力分布产生了影响;控制后航行体表面压力系数经过侧流面低压区后,

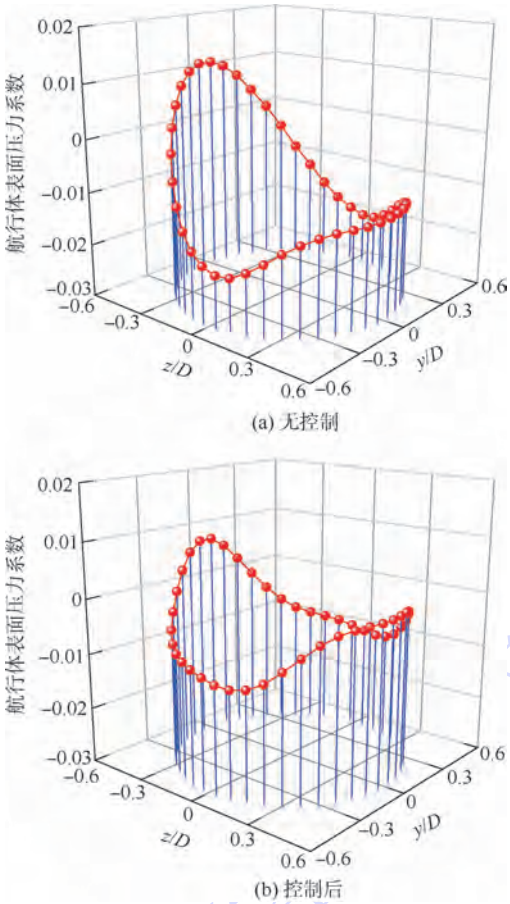


图 10　航行体表面压力分布对比

Fig. 10　Comparison of pressure distribution of vechicles

沿 z 轴正方向增大的幅度逐渐减小，且在背流面区域压力分布更加均匀. 可见，水下航行体多相流体控制不仅可以有效降低作用在航行体表面的压力，同时又有利于航行体表面压力的均匀分布.

4　结　论

采用三维数值模拟方法对比分析了水下航行体在不同流动控制方案下的通气空泡形态特征、阻力系数和表面压力分布变化特性，结果显示：

1）通过流动控制可有效减小在有横流作用下垂直发射航行体空泡的不对称性.

2）对于不同的通气口轴向位置，通气口在前部时，航行体尾部仍有较大面积的沾湿区，因此黏性阻力较大；通气口在中部时，可控制空泡完全覆盖航行体，黏性阻力最小. 前部通气口产生的压差阻力最大，中部和后部通气口的基本相同. 可见，中部通气口可以有效减小黏性阻力系数和压差阻力系数.

3）控制水下垂直发射航行体迎水面位置的通气口不仅可以减小水下航行体的表面压力，而且可以减小迎流面到背流面压力曲线的波动，从而有利于表面压力均匀分布.

本文仅对不同控制方法的水下航行体流体动力问题作了初步研究，但在通过流动控制方法控制水下航行体载荷和弹道这一问题上，仍需要进一步的探索.

参考文献（References）

[1] Reichardt H. The laws of cavitation bubbles at axially symmetrical bodies in a flow[M]. Gottingen: Kaiser Wilhelm Institute für Stromungsforschung, 1945: 322-326.

[2] Sliberman E, Song C S. Instability of ventilated cavities[J]. Journal of Ship Research, 1959, 5(1): 13-33.

[3] Lindau J W, Kunz R F, Mulherin J M, et al. Fully coupled, 6-DOF to URANS, modeling of cavitating flows around a supercavitating vehicle[C]//Fifth International Symposium on Cavitation. Osaka: [s. n.], 2003: 124.

[4] 闵景新，魏英杰，王聪，等. 潜射导弹垂直发射过程流体动力特性数值模拟[J]. 兵工学报，2010，31(10)：1303-1309.
Min J X, Wei Y J, Wang C, et al. Numerical simulation on hydrodynamic characteristics of submarine missile in the vertical launch process[J]. Acta Armamentarii, 2010, 31(10): 1303-1309(in Chinese).

[5] Kawakami E, Arndt R E. Investigation of the behavior of ventilated supercavities[J]. Journal of Fluids Engineering, 2011, 133(9): 091305.

[6] Matveev K, Miller M. Air cavity with variable length under a model hull[J]. Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment, 2011, 225(2): 161-169.

[7] Ramesh S S, Lim K, Khoo B. An axisymmetric hypersingular boundary integral formulation for simulating acoustic wave propagation in supercavitating flows[J]. Journal of Sound and Vibration, 2012, 331(19): 4313-4342.

[8] 黄彪，王国玉，权晓波，等. 轴对称体空化水动力脉动特性的实验研究[J]. 工程力学，2012，29(2)：239-244.
Huang B, Wang G Y, Quan X B, et al. Experimental study on fluctuating hydrodynamics around axisymmetric bodies[J]. Engineering Mechanics, 2012, 29(2): 239-244(in Chinese).

[9] 王一伟，黄晨光，杜特专，等. 航行体垂直出水载荷与空泡溃灭机理分析[J]. 力学学报，2012，44(1)：39-48.
Wang Y W, Huang C G, Du T Z, et al. Mechanism analysis about cavitation collapse load of undervater vehicles in a vertical launching process[J]. Chinese Journal of Theoretical and Applied Mechanics, 2012, 44(1): 39-48(in Chinese).

[10] Bragg M, Gregorek G. Experimental study of airfoil performance with vortex generators[J]. Journal of Aircraft, 1987, 24(5): 305-309.

[11] Calarese W, Crisler W. Afterbody drag reduction by vortex generators, AIM-85-0354[R]. Nevada: AIAA, 1985.

[12] Wendt B, Hingst W. Flow structure in the wake of a wishbone vortex generator[J]. AIAA Journal, 1994, 32(11): 2234-2240.

[13] Lu F K, Pierce A J, Shih Y, et al. Experimental and numerical study of flow topology past micro vortex generators[C]//40th Fluid Dynamics Conference and Exhibit. Chicago: [s. n.],

2010:4463.

[14] Lu F K, Li Q, Liu C. Microvortex generators in high-speed flow [J]. Progress in Aerospace Sciences, 2012, 53:30-45.

[15] 于彦泽,刘景飞,蒋增奠,等. 大型飞机后体流动控制及减阻机理研究[J]. 空气动力学学报,2011,29(5):640-644.
Yu Y Z, Liu J F, Jiang Z Y, et al. The investigation of flow control and drag reduction mechanism for transport airplane aftbody[J]. Acta Aerodynamica Sinica, 2011, 29(5):640-644(in Chinese).

[16] 季路成,林峰. 绕凸包主动流动控制的初步数值研究[J]. 工程热物理学报,2007,27(1):45-47.
Ji L C, Lin F. Numerical studies about active flow control over a hump flow[J]. Journal of Engineering Thermophysics, 2007, 27(1):45-47(in Chinese).

[17] Menter F R. Two-equation eddy-viscosity turbulence models for engineering applications [J]. AIAA Journal, 1994, 32(8): 1598-1605.

[18] 黄海龙,黄文虎,王聪,等. 数值模拟通气角度对超空泡形态特性影响分析[J]. 工程力学,2007,24(12):195-208.
Hung H L, Huang W H, Wang C, et al. Numerical analysis of the influence of angle of ventilation on the shapes of supercavity [J]. Engineering Mechanics, 2007, 24(12):195-208(in Chinese).

[19] 孙铁志,魏英杰,王聪,等. 孔状通气条件下潜射航行体流体动力特性研究[J]. 兵工学报,2013,34(11):1424-1430.
Sun T Z, Wei Y J, Wang C, et al. Research on hydrodynamic characteristics of submarine launched vehicle with ventilation holes[J]. Acta Armamentarii, 2013, 34(11):1424-1430(in Chinese).

[20] Acosta A. The effect of a longitudinal gravitational field on the supercavitating flow over a wedge[J]. Journal of Applied Mechanics, 1961, 28(2):188-192.

Hydrodynamic features of underwater vehicles based on flow control

ZHANG Xiaoshi, WANG Cong*, CAO Wei, ZHAO Chenggong

(School of Astronautics, Harbin Institute of Technology, Harbin 150001, China)

Abstract: The hydrodynamic characteristics of underwater vehicles were numerical modelling studied by CFX. Shear stress transport turbulence model control equation of mixed media, Reynolds-averaged Navier-Stokes equation and mass transport equation between each phase were solved, and three dimensional numerical simulation methods were adopted to contrast and analyze the variation characteristics of cavity shape, surface pressure distribution and drag coefficient of underwater vehicles with different flow controlschemes. The effects of different parameters on drag reduction effect were also discussed. The result shows that, for the vertical underwater launching vehicles under cross-flow effect, multiphase flow control can not only lower the asymmetry of ventilated cavity and drag force of vehicles, but also balance the pressure on the positive side and back surfaces to realize the control of cavity shape and surface pressure characteristics etc. of vehicle multiphase flow. Furthermore, the location of ventilation hole has a significant influence on drag reduction effect.

Key words: flow control; ventilation holes; cavity shape; hydrodynamics; multiphase flow; numerical simulation

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0308

Loop 细分曲面精加工刀具轨迹生成

陈甜甜，赵罡*
（北京航空航天大学 机械工程及自动化学院，北京 100191）

摘 要：细分曲面既能表示连续的几何设计模型也能表示离散的加工模型，避免了模型转换的复杂中间过程. 细分曲面除了对于构造具有任意拓扑结构的复杂零件具有巨大的优势外，对于数字化制造也极具发展潜力. 因此，对基于 Loop 细分曲面的精加工刀具轨迹生成算法进行了研究. 首先利用基于弦长误差的自适应插值 Loop 细分得到精加工模型；然后利用等斜率跟踪法将精加工模型分割为平坦区域和非平坦区域. 对各个区域依次进行处理，不同区域实施不同的刀具轨迹规划. 据此，既避免了因细分过程数据量过大而导致的曲面精度不足或表面质量降低的问题，又可顺利计算整张曲面相对均匀残留高度的刀具轨迹. 最后，为验证该算法的可行性进行铣削加工实验.

关 键 词：CAD/CAM；精加工；Loop 细分曲面；刀具轨迹；铣削
中图分类号：TP391.72
文献标识码：A **文章编号**：1001-5965(2015)04-0663-06

刀具轨迹的生成是复杂零件数控加工中最重要的内容. 它是根据零件几何模型，结合所选用的加工机床、刀具、走刀方式以及加工余量等工艺方法进行刀位计算并生成加工运动轨迹. 刀具轨迹的生成能力直接决定数控编程系统的功能及所生成加工程序的质量[1].

目前，基于 NURBS 自由曲面的数控加工刀具轨迹规划方法已较成熟. 但是，在数控加工等工程领域的实际应用中，由于 NURBS 构造的自由曲面模型是分片光滑连续的几何模型，故连续的 NURBS 曲面模型经常被离散为三角网格模型再生成刀具轨迹. 这是一个从连续到离散，再从离散到连续的过程. 这种过程往往会伴随着模型的复杂程度增大而变得更加困难.

与工业造型的标准工具 NURBS 方法相比，细分曲面是一种以网格细分为特征的离散造型方法. 该方法是在任意拓扑结构的多边形网格的基础上，按照某些规则对网格不断细化，使其趋向于一个满足一定光滑度的极限曲面. 因此，细分曲面既能表示连续的几何设计模型又能表示离散的加工模型，避免了模型转换的复杂中间过程，同时减少了计算量. 上述优点表明，细分曲面除了对于构造具有任意拓扑结构的复杂零件具有巨大的优势外，对于数字化制造也极具发展潜力.

Kurgano 等首先提出了基于 Catmull-Clark 四边形细分曲面的数控加工刀具轨迹生成算法[2]，该算法证实了细分曲面作为一种新的造型工具也可应用于数控加工. 钟大平、白杰等应用等距截平面法对 Loop 细分曲面的刀具轨迹生成算法进行了探讨[3-4]. 赵建民及唐虹对基于 Catmull-Clark 细分曲面等距面的生成及误差控制进行了分析与研究[5-6]. Lu 等则重点讨论了如何更好地控制

收稿日期：2014-05-28；**录用日期**：2014-08-05；**网络出版时间**：2014-09-05 16:40
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0308.html
基金项目：国家自然科学基金资助项目(61170198)；中央高校基本科研业务费资助项目(YWF-14-JXXY-19)
作者简介：陈甜甜(1982—)，女，上海人，实验师，candy_ctt@163.com
*通讯作者：赵罡(1972—)，男，河北文安人，教授，zhaog@buaa.edu.cn，研究方向为复杂曲面造型、加工新技术、飞机数字化装配、虚拟现实技术及应用.

引用格式：陈甜甜，赵罡. Loop 细分曲面精加工刀具轨迹生成[J]. 北京航空航天大学学报，2015，41(4)：663-668.
Chen T T，Zhao G. Tool path generation for Loop subdivision surface based finish machining[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：663-668. (in Chinese).

北航学报 赠阅

Catmull-Clark 细分曲面加工精度和保证零件的加工质量[7]. 孙殿柱等通过改进的 R*-tree 空间聚类索引结构实现刀轨截面与三角片的快速求交,利用局部偏置获取刀位点[8]. 原恩桃等探讨了多轴加工中刀轴矢量的计算方法,采用面积均值法计算 Catmull-Clark 细分曲面刀轨,该方法有效地保证残留高度均匀且加工零件表面光滑[9]. 上述利用截平面法生成刀具轨迹的算法,具有走刀路线灵活,生成的刀具轨迹分布均匀等特点,但有时会导致大量较短的加工路径出现,影响加工效率与精度[10]. 戴军富等提出了 Doo-Sabin 细分曲面 S-Buffer 算法生成 NC 刀具轨迹[11]. 陆云芳改进了利用 Z 坐标信息建立刀具路径的 Z-Buffer 模型,提出了根据加工精度对矩形区域进行网格划分来生成刀具轨迹[12]. 当待加工模型很复杂或者精度要求很高时,Z-Buffer 模型对内存要求较高,影响了计算效率.

与四边形网格相比,三角域的 Loop 细分模式在规则处能保持 C^2 连续,是目前应用最广泛的细分模式之一. 直接以三角网格为基础生成刀具轨迹,各三角片之间缺乏明显的拓扑关系,而 Loop 细分曲面模型本身就存储着三角片之间的拓扑关系,可以快速地生成刀具轨迹. 在精加工阶段,由于加工曲面残留高度保持恒定可以获得更高的加工质量,因此,本文重点研究 Loop 细分曲面的等残留高度精加工刀具轨迹生成算法.

1 精加工模型

在实际应用中,数控加工大致可分为粗加工、精加工和清根加工 3 个阶段. 精加工的最终目的是保证零件的尺寸精度满足设计要求. 精加工对曲面加工精度和表面质量要求均较高,因此,待加工模型的光滑极限曲面作为参考模型较为理想. 在利用控制网格多次 Loop 细分得到极限曲面时,三角片以指数倍增长的问题非常突出. 因此,本文直接将待加工模型的控制网格顶点由其对应的极限位置替代,以减少每次所需处理的三角片数据量. 事实上,大多数情况下这种极限网格比控制网格逼近极限曲面的效果更好[13].

自由曲面模型在精加工阶段主要有两个离散过程[7],其一是将连续的几何设计模型离散为精加工模型时产生的弦长误差;其二是精加工模型离散为一系列刀具轨迹时产生的残留高度误差.

Loop 极限曲面由极限网格逼近的过程,即为几何设计模型离散为精加工模型的过程. 故定义弦长误差为极限网格的三角片与其所对应的极限曲面片之间的误差. 通过不断 Loop 细分极限网格就可以得到满足弦长误差的精加工模型. 自适应细分模式可以有效地减缓细分中数据量增长过快的问题,节省存储资源并提高绘制速度. 因此,本文取三角片的弦长误差作为自适应 Loop 细分阈值,通过自适应插值 Loop 细分得到精加工模型. 给定弦长误差 c,具体步骤如下:

步骤 1 计算网格 M_0 每一个三角片的弦长误差,找到最小的一个弦长误差记为 $c_{\min}$.

步骤 2 不断 Loop 细分网格 M_0,直到细分 i 次后其最小弦长误差 $c_{\min} < c$,表明此时的网格 M_i 不需要均匀细分,并标记所有的三角片 F_i^j 为 false,$j = 1, 2, \cdots, \max F\{M_i\}$.

步骤 3 对于任一个三角片 F_i^j,如果其弦长误差 $c_i^j > c$,标记该三角片为 true.

步骤 4 对标记为 true 的三角片进行插值 Loop 细分[14],如果细分后三角片 F_{i+1}^j 的弦长误差 $c_{i+1}^j < c$,标记细分后的三角片为 false.

步骤 5 重复步骤 4 直到所有三角片为 false;消除存在裂缝的三角形并利用 Loop 极限位置更新细分曲面的顶点,输出精加工模型 M.

如果一个网格面已经停止细分,而其周围相邻面却未满足停止细分的准则,此时在相邻边处会产生缝隙. 为了避免视觉上的漏洞,同时也为了保持新网格的二维流形性质,需要消除不同分辨率层网格面间的裂缝[15].

2 等残留高度精加工刀具轨迹

2.1 等残留高度误差

残留高度 H 是由走刀行距 L 决定的,如图 1 给出走刀行距与残留高度之间的关系. 假设球头刀的半径是 R,则有

$$(R - H)^2 + \left(\frac{L}{2}\right)^2 = R^2 \tag{1}$$

由式(1)可以推出:

$$L = 2\sqrt{(2R - H)H} \tag{2}$$

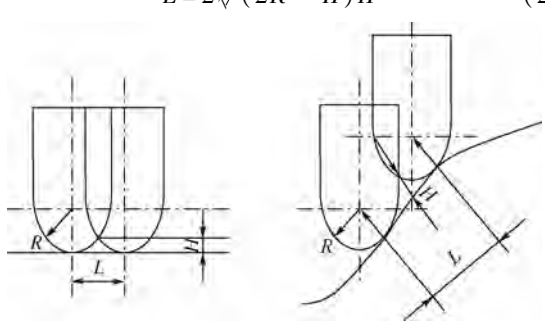


图 1 走刀行距 L 与残留高度 H 之间的关系

Fig. 1 Relationship between step-over L and cusp height H

北航学报 赠阅

在精加工阶段，理想的情况是均匀的残留高度，即走刀行距 L 是常量，等残留高度会最大限度地减少刀具轨迹并且得到最短的加工时间. 但是，等残留高度的计算复杂度较高，并且如果走刀行距 L 是常数，自由曲面模型不同的斜率会导致走刀步长 s 变化较大. 如果把走刀步长 s 设为常量，会导致加工后的残留高度不均匀，甚至出现零件表面残留高度大于指定的误差，影响零件的加工精度. 一个比较合理的方案就是把一个曲面分解为不同的区域，例如平坦和非平坦区域，然后采用不同的刀具轨迹规划. 这样，既可以控制残留高度误差又可以减少多余的刀具轨迹.

基于上述讨论，目前需要解决的主要问题是分割精加工模型，并对分割后的模型进行不同的刀轨规划. 由于参数曲面的几何性，分割参数曲面计算复杂不易做到. 但是，对于多面体网格模型而言，分割模型相对较容易完成[16]. 本文提出一种等斜率跟踪法来分割精加工模型.

给定倾斜角阈值 θ，如图 2 所示. 等斜率曲线就是由曲面上那些斜率为常数 θ 的点组成. 如果是闭合模型，那么等斜率曲线亦为闭合曲线环. 对于精加工模型上的任意顶点 v_i，其单位法向量为 $\boldsymbol{N}_i(n_x, n_y, n_z)$. 假设顶点 v_i 在极限曲面上的倾斜角为 α，那么 $n_z = \cos\alpha$. 所以，寻找精加工模型上等斜率曲线就是寻找所有的点具有相同的 n_z，即具有相同的倾斜角. 对于精加工模型 M，等斜率曲线可以通过三角片的边 v_iv_j 与等斜率曲线相交得到. 该边须满足：$n_{zi} \leqslant \cos\theta \leqslant n_{zj}$ 或者 $n_{zi} \geqslant \cos\theta \geqslant n_{zj}$. 这里 n_{zi} 是顶点 v_i 的法向量 z 轴坐标的分量，n_{zj} 是顶点 v_j 的法向量 z 轴坐标的分量. 通过线性插值 v_i 和 v_j 得到该曲面与等斜率曲线的相交点后，精加工模型即可以分割为平坦和非平坦区域. 算法具体步骤如下：

步骤 1　对于精加工模型 M 的每一个三角片 $v_iv_jv_k$，计算其 3 个顶点的法向量 z 轴坐标分量. 如果其法向量 z 轴坐标分量都小于 $\cos\theta$，标记该三角片为平坦；如果其法向量 z 轴坐标分量都大于 $\cos\theta$，标记该三角片为非平坦. 剩余的三角片为相交的三角面，标记为不可访问的面.

步骤 2　从标记为相交的三角片开始，利用线性插值找到等斜率曲线的第一个点，标记该面为可访问；下一个标记为相交的三角片与该三角片共享同一条边. 找到下一个相交的等斜率点并标记该面为可访问. 继续搜寻直到回到第一个相交点，此时产生一条等斜率曲线环.

步骤 3　继续搜寻标记为不可访问的面，按照步骤 2 生成另一条等斜率曲线环.

步骤 4　不断执行步骤 3 直到所有相交的三角片都为可访问为止.

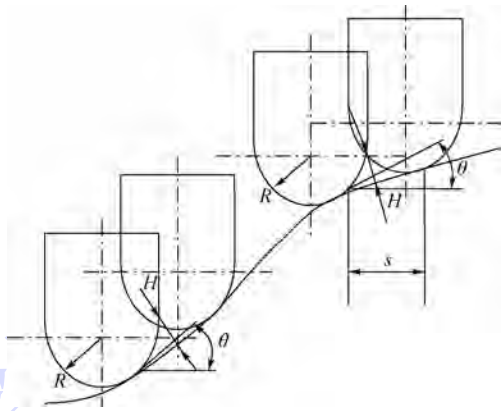


图 2　曲面倾斜角度对残留高度的影响

Fig. 2　Slope angle of surface and cusp height

上述算法利用了 Loop 细分曲面三角片之间的拓扑关系来简化等斜率曲线的跟踪，其计算复杂度为 $O(n)$，其中 n 是精加工模型三角片的个数. 这里，定义平坦和非平坦区域的交线，即等斜率曲线环，属于平坦或者非平坦区域两者之一. 如此一来，模型被分割为两个完全不同的区域. 需要指出的是，上述算法分割曲面所得到的区域不仅仅只有平坦和非平坦区域，实际上为两个以上不同的区域. 为了避免过多的抬刀和空切，所分解出的区域的个数亦不能太多. 对于给定的倾斜角阈值 θ，可以首先给出估计值. 例如，限制最短的加工时间所得到的阈值，或者根据经验不断测试得到的阈值.

图 3 是利用等斜率跟踪法对于圆盘模型进行分割的实例. 可以看出，给定不同的倾斜角阈值所得到的分割曲面亦不相同. 图 4 是分割 bishop 模型的实例. 其中，倾斜角阈值分别为 45°和 80°. 当阈值为 45°和 80°时，bishop 的正面都被分割为 4 个等斜率曲线环，即 4 个不同的区域. 很明显，当阈值为 80°时所分割的 4 个区域更为合理.

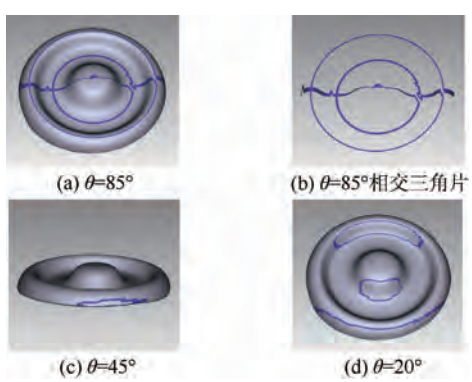


图 3　等斜率跟踪法分割圆盘模型

Fig. 3　Iso-slope curve tracing method for disc model decomposition

北航学报 赠阅

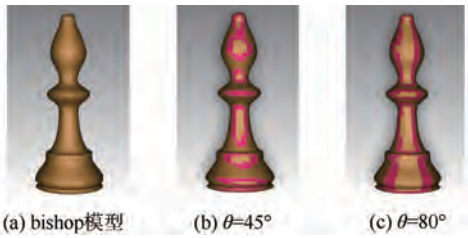


图 4 等斜率跟踪法分割 bishop 模型

Fig. 4 Iso-slope curve tracing method for bishop model decomposition

等斜率分割算法对于大数据量的细分曲面加工非常有利. 随着细分次数的递增,细分曲面数据量增长过快,此时依照等斜率跟踪法可以将精加工模型分割为不同的区域. 对于非平坦区域应用等高线法生成刀轨,对于平坦区域应用截面法生成刀轨,进而得到相对均匀残留高度的刀轨.

2.2 无过切刀具轨迹

计算多面体网格模型精确和完整的等距面比较复杂[17]. 本文采用等距精加工模型的每个顶点来逼近等距面. 利用等距面生成刀位点时会产生自相交从而导致零件过切. 在 3D 模型中较难检测和解决自相交问题,故选择转化为 2D 刀具轨迹之后进行处理.

对于等高线法生成的刀具轨迹,存在两种类型的自相交:局部自交和全局自交. 局部自交是与其中一条等高线有交点,全局自交是与很多条等高线有多个交点. 对于自相交,有如下两种方法可以解决:①等高线法. 利用等高线走向,开始于一个有效的点并且沿着一个方向寻找最简单的多边形等高线. ②多边形裁剪法. 标记在等高线内部的边,并且去除它们形成最终的等高线环. 对于截面法生成刀具轨迹产生的过切问题,三轴数控加工中可以利用扫描线法来寻找自相交点[18]. 通过检测自相交点去除无效部分保留剩余刀轨.

3 精加工刀具轨迹生成实例

为了检验基于 Loop 细分曲面的精加工刀具轨迹生成算法的有效性,结合 OpenGL 在 Visual C++6.0下实现本文算法.

对于图 5(a)的 knob 模型而言,首先利用算法自适应插值 Loop 细分生成精加工模型,弦长误差为 0.002. 当倾斜角阈值为 50.4° 时(如图 5(b)),knob 模型所分割的平坦区域和非平坦区域如图 5(c)和图 5(d)所示. 利用等高线法生成全局刀轨,然后对于残留高度较大的平坦区域利用截面法生成精加工刀具轨迹,如图 5(e)所示. 由图 5(f)的精加工仿真结果可以看出,对于 knob 模型不同的斜率采取走刀变步长以保持均匀残留高度.

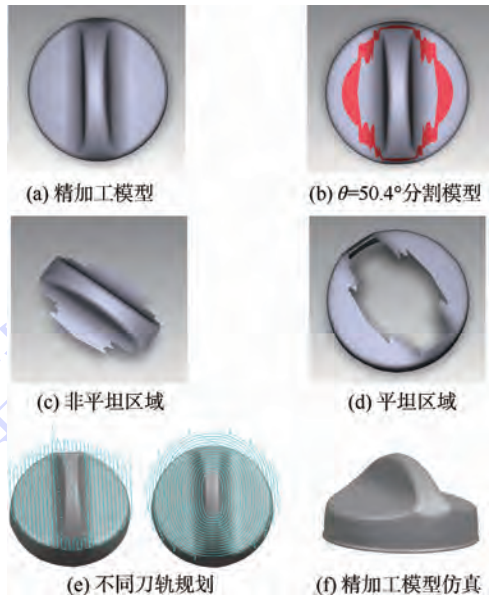


图 5 knob 模型精加工刀具轨迹生成

Fig. 5 Tool path generation for knob model finish machining

图 6(a)是运用自适应插值 Loop 细分生成的 face 模型精加工模型,弦长误差为 0.05. 利用等斜率跟踪法将精加工模型分割后的模型如图 6(b)所示,倾斜角阈值为 14°. 在非平坦区域运用等高线法生成刀具轨迹,对于平坦区域运用截面法生成刀具轨迹,如图 6(d)所示. 图 6(e)是最终的相对均匀等残留高度的加工仿真结果.

图 6 face 模型精加工刀具轨迹生成

Fig. 6 Tool path generation for finish machining of human face surface model

北航学报 赠阅

为验证本文算法的可行性，以 face 模型为例进行铣削加工. 实验机床为 VMC0850B 三轴立式加工中心. 工件材料为铝合金 7075，毛坯尺寸为 120 mm × 100 mm × 30 mm. 实验刀具为 Sandvik 硬质合金球头刀. 在本实验的精加工刀具路径规划中，对于非平坦区域（鼻尖、嘴处）采用等残留高度法；对于平坦区域（额头、脸处）采用截面法生成刀具轨迹. 精加工的切削参数见表 1.

表 1　切削参数

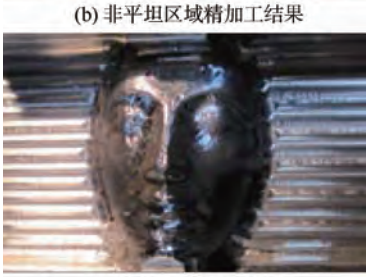
Table 1　Cutting parameters

参数	刀具直径/mm	刀齿数	主轴转速/(r·min^{-1})	进给速度/(mm·min^{-1})	行距/mm	切削余量/mm
数值	8	2	4 000	900	0.04	0.25

三轴铣削加工过程如图 7(a)所示；非平坦区域运用等高线法加工后的结果如图 7(b)所示，可以清晰地看出 face 模型被分割为非平坦区域和平坦区域；图 7(c)是精加工后零件的正面图. 由于本文重点讨论的是 Loop 细分曲面在精加工中刀具轨迹的生成，并未进行清根加工，所以精加工表面质量不够光滑. 此外，由于是三轴铣削加工，无法加工 face 模型鼻孔凹陷处，利用多轴加工规划 Loop 细分曲面刀具轨迹将是未来的研究方向.

(a) face模型精加工过程

(b) 非平坦区域精加工结果



(c) 精加工结果正面图

图 7　face 模型精加工实例

Fig. 7　Result of finish machining of human face surface model

4　结　论

本文重点探讨了 Loop 细分曲面在精加工阶段刀具轨迹生成算法.

1）针对精加工模型，通过自适应插值 Loop 细分算法来控制几何模型离散为精加工模型时所产生的弦长误差. 该算法在满足弦长误差的同时可以有效减缓 Loop 细分中三角片指数级增长的问题.

2）由于精加工模型离散为一系列刀具轨迹时，零件表面残留高度过大会影响零件的加工精度. 因此，本文提出等斜率跟踪法将精加工模型分解为平坦和非平坦区域，对于平坦区域应用截面法生成刀具轨迹，对于非平坦区域应用等高线法生成刀具轨迹. 该算法不仅在处理大数据量 Loop 细分曲面精加工方面具有优势，又能够得到整张曲面相对均匀残留高度的刀具轨迹.

3）最后以 face 模型为例对上述算法进行了三轴铣削加工实验验证. 后续将研究基于 Loop 细分曲面的多轴数控加工刀具轨迹生成.

参考文献（References）

[1] 周济，周艳红. 数控加工技术[M]. 北京：国防工业出版社，2002：78-147.
Zhou J, Zhou Y H. Numerical control machining technology [M]. Beijing: National Defence Industry Press, 2002: 78-147 (in Chinese).

[2] Kurgano J, Suziki H K F. Generation of NC tool path for subdivision surfaces [C]//Proceedings of CAD/Graphics. Kunming: IEEE Computer Society Press, 2001: 676-682.

[3] 钟大平，周来水，王占东，等. 细分曲面的 NC 导轨生成算法及实现[J]. 东南大学学报，2004，34(1)：1-4.
Zhong D P, Zhou L S, Wang Z D, et al. Research on NC tool path generation for subdivision surface [J]. Journal of Southeast University, 2004, 34(1): 1-4 (in Chinese).

[4] 白杰，赵罡，姚福生. 带折痕的 Loop 细分曲面等距面处理算法[J]. 计算机辅助设计与图形学学报，2008，20(10)：1261-1265.
Bai J, Zhao G, Yao F S. An offset algorithm for Loop subdivision surface with creases [J]. Journal of Computer-Aided Design & Computer Graphics, 2008, 20(10): 1261-1265 (in Chinese).

[5] 赵建民. 基于 Catmull-Clark 模式的细分曲面 NC 刀具轨迹生成技术[D]. 大连：大连理工大学，2008.
Zhao J M. Technology of generating NC tool path for subdivision surfaces based on Catmull-Clark scheme [D]. Dalian: Dalian University of Technology, 2008 (in Chinese).

[6] 唐虹. Catmull-Clark 细分曲面的 NC 刀具生成关键技术研究[D]. 南京：南京航空航天大学，2008.
Tang H. Research on key technologies of NC tool path generation based on Catmull-Clark subdivision surface [D]. Nanjing: Nan-

jing University of Aeronautics and Astronautics, 2008 (in Chinese).

[7] Lu C, Ting K. Subdivision surface-based finish machining[J]. International Journal of Production Research, 2006, 44(12): 2445-2463.

[8] 孙殿柱,范述鑫,李延瑞,等. 三角网格细分曲面数控加工刀轨快速生成算法[J]. 农业机械学报,2009,40(3):218-221.
Sun D Z, Fan S X, Li Y R, et al. Algorithm of NC tool path generation for triangular meshing[J]. Journal of Agricultural Machinery, 2009, 40(3): 218-221(in Chinese).

[9] Yuan N T, Liao W H, Tang H. Generation of NC tool path for Catmull-Clark subdivision surface[C]//Proceeding of First International Conference on Modeling and Simulation. Nanjing: IEEE Computer Society Press, 2008: 93-98.

[10] 徐金亭,刘伟军,卞宏友,等. 基于网格曲面模型的等残留刀位轨迹生成方法[J]. 机械工程学报,2010,46(11):193-198.
Xu J T, Liu W J, Bian H Y, et al. Constant scallop tool path for triangular surface machining[J]. Journal of Mechanical Engineering, 2010, 46(11): 193-198(in Chinese).

[11] 戴军富,秦开怀. Doo-Sabin 曲面 NC 刀具的并行轨迹计算[J]. 计算机辅助设计与图形学学报,2003,15(5):604-609.
Dai J F, Qin K H. A parallel method for generating NC tool paths of Doo-Sabin surfaces[J]. Journal of Computer-Aided Design & Computer Graphics, 2003, 15(5): 604-609(in Chinese).

[12] 陆云芳. 细分曲面在 CAM 中的关键技术研究[D]. 杭州:浙江大学,2007.
Lu Y F. Research on the key technology of subdivision surface in CAM[D]. Hangzhou: Zhejiang University, 2007(in Chinese).

[13] Huang Z J, Wang G P. Bounding the distance between a Loop subdivision surface and its limit mesh[C]//Proceedings of Geometric Modeling and Processing. Heidelberg: Springer Verlag, 2008: 33-46.

[14] 陈甜甜,赵罡. 半规则三角网格模型细分曲面重构[J]. 北京航空航天大学学报,2012,38(9):1245-1249.
Chen T T, Zhao G. Semi-regular triangular mesh subdivision surface reconstruction[J]. Journal of Beijing University of Aeronautics and Astronautics, 2012, 38(9): 1245-1249(in Chinese).

[15] 李桂清,吴壮志,马维银. 自适应细分技术研究进展[J]. 计算机辅助设计与图形学学报,2006,18(12):1789-1798.
Li G Q, Wu Z Z, Ma W Y. Research advances in adaptive subdivision techniques[J]. Journal of Computer-Aided Design & Computer Graphics, 2006, 18(12): 1789-1798(in Chinese).

[16] Flutter A, Todd J. A machining strategy for tool making[J]. Computer-Aided Design, 2001, 33(13): 1009-1022.

[17] Jun C S, Kim D S, Park S. A new curve-based approach to polyhedral machining[J]. Computer-Aided Design, 2002, 349(5): 379-389.

[18] Bentley J L, Ottmann T A. Algorithm for reporting and counting geometric intersections[J]. IEEE Transactions on Computers, 1979, C-28(9): 643-647.

Tool path generation for Loop subdivision surface based finish machining

CHEN Tiantian, ZHAO Gang*

(School of Mechanical Engineering and Automation, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: Both smooth design model and discrete machining model can be represented by the subdivision surface, which avoids the complex progress of model transformation. Subdivision surface can be used to generate a smooth limit surface on an arbitrary topology mesh, and thus has great advantage in modeling complex parts. In addition, it has great potential for applications in NC machining. The algorithm for tool path generation of finish machining based on Loop subdivision surface was investigated. The method of adaptive interpolation Loop subdivision surfaces was proposed to control the chord error between limit mesh and limit surface. The finish machining model was then decomposed into different regions by iso-slope curve tracing method, and different tool path patterns were calculated to obtain an even cusp height. Therefore, it is simple to generate the even cusp height tool path of the entire finish machining model, and the conflict between large-scale subdivision data and required high surface quality and precision is also solved. Milling test is implemented to verify the algorithm proposed.

Key words: CAD/CAM; finish machining; Loop subdivision surface; tool path; milling

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0309

基于显著特性的保持边缘滤波算法

冯策，戴树岭*，赵帅和

（北京航空航天大学 自动化科学与电气工程学院，北京 100191）

摘 要：针对传统保持边缘滤波算法中存在光晕的缺点，提出了基于显著特性的图像保持边缘滤波算法.该算法主要思想是利用原图的显著特性图具有边缘突出的特点，简化双边滤波法中灰度因子的设定工作.首先提取出原图像的显著图，然后根据图像显著值的大小自适应模糊图像中的细节和噪声部分：显著值较小的区域灰度因子设置较大，平滑此区域；显著值较大的区域灰度因子设置较小，保持边缘部分的清晰.实验结果表明，该算法在平滑了细节和噪声的同时，有效地保持了边缘信息的清晰，与传统方法相比，新方法不仅避免了光晕现象的产生，而且应用更为广泛.

关 键 词：保边滤波；图像平滑；显著图；双边滤波；引导滤波
中图分类号：TP751.1；V557+.4
文献标识码：A **文章编号**：1001-5965(2015)04-0669-06

图像滤波是计算机视觉、医学图像处理等图像工程领域中重要的预处理手段.保持边缘的滤波是图像滤波研究领域里一个重要的目标.理想的保持边缘滤波法，一方面要平滑细节和噪声，另一方面要保持图像重要的边缘信息.

近年来涌现出了许多优秀的保边滤波算法，本文将其分为如下2类：

1）非平均滤波法.最具代表性为经典的中值滤波法[1]，1992 年，Rudin 提出的 TV 滤波法[2]采用了梯度的 L1 范数作为正则化约束.Xu[3]后来提出通过最小化 L0 范数方法来平滑图像，实现了显著边缘的保持.2008 年，Farbman[4]提出了加权最小二乘滤波算法（WLS）应用于多尺度图像处理.2011 年，Paris 等[5]提出一种基于局部拉普拉斯金字塔的保边缘图像滤波算法.以上滤波法虽然效果比较理想，但计算量比较大，时间复杂性高，内存空间开销大，效率较低.

2）平均滤波算法.比较有代表性的为各项异性混合[6]、双边滤波[7]等，双边滤波是使用最为广泛的一种非线性滤波法，其加权因子是距离权重和灰度权重的乘积，后期衍生出了一系列双边滤波的优化算法[8-10]，2010 年，He 等[11]提出了引导滤波法，同样属于平均滤波范畴，引导滤波法是一种线性的保边滤波法，具有很好保边滤波特性.但是它们的缺点是在边缘区域会出现光晕现象.

一幅图像的信息通常包含2部分：

1）边缘区域.复杂度较大区域，包含显著边缘.

2）非边缘区域.复杂度较小区域，包含噪声，细节和平坦部分.

保持边缘滤波法的目标是平滑非边缘区域的同时保持边缘区域的清晰.目前滤波算法很难区分边缘区域和非边缘区域，无法明确地约束边缘的取舍，因此会导致边缘部分的模糊，在边缘处出现光

收稿日期：2014-05-29；**录用日期**：2014-08-25；**网络出版时间**：2014-10-14 09:23
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0309.html
基金项目：国家高技术研究发展计划资助项目(2009AA01Z333)；中央高校基本科研业务费专项资金资助项目(YWF-13-D2-HK-26)；北京市科委项目(Z111100074811001)
作者简介：冯策(1984—)，男，黑龙江牡丹江人，博士生，fengce1984@163.com
*通讯作者：戴树岭(1966—)，男，河北沧州人，教授，sldai@yeah.net，研究方向为虚拟现实技术.

引用格式：*冯策，戴树岭，赵帅和. 基于显著特性的保持边缘滤波算法[J]. 北京航空航天大学学报，2015，41(4)：669-674.*
Feng C，Dai S L，Zhao S H. Edge-preserving filtering based on saliency map[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：669-674 (in Chinese).

晕等失真现象. 所以如何寻找一个合适的区域间平衡点是目前保边滤波法主要研究的问题.

本文提出了一种基于显著特性的保边滤波法. 由于本文的显著特性算法具有边缘区域突出的特点,容易区分出边缘和非边缘区域,因此可利用这个特点来自适应地控制局部窗口的模糊程度,本文算法除了保持了传统滤波保边算法优点以外,还有效地避免了边缘区域的光晕现象.

1 双边滤波的问题分析

1.1 双边滤波法

双边滤波法是在高斯滤波法基础上提出的保边滤波法,滤波过程不仅考虑到了空间距离信息,也考虑到了灰度的相似性,但是双边滤波法是局部加权的平均滤波法,所以不可避免地会出现光晕现象. 双边滤波法如下:

$$P_i = \frac{1}{k}\sum_{j\in\Omega} I_j w_s(i,j) w_r(I_i,I_j) \tag{1}$$

式中

$$w_r(I_i,I_j) = \exp\left(-\frac{|I_i - I_j|^2}{2\sigma_r^2}\right) \tag{2}$$

$$w_s(i,j) = \exp\left(-\frac{|i-j|^2}{2\sigma_s^2}\right) \tag{3}$$

$$k = \sum_{j\in\Omega} w_s(i,j) w_r(I_i,I_j) \tag{4}$$

P 为双边滤波后的图像;I 为输入图像;w_s 为距离权重;w_r 为灰度权重;σ_s 为距离权重的标准差;σ_r 为灰度权重的标准差.

1.2 问题分析

双边滤波中灰度标准差 σ_r 的取值决定了其距离权重的高斯曲线的扩展程度,σ_r 越大,高斯曲线下降越慢,平滑效果更明显. σ_r 越小,高斯曲线下降越快,平滑效果变弱. 因此,如何选取合适的 σ_r 是双边滤波保持边缘特性的关键. 如果在边缘区域,σ_r 取值过大,会导致边缘部分的模糊,在处理图像时会出现光晕现象,如图 1 所示,左侧为图像增强后出现的光晕(halo)效果. 而 σ_r 取值过小时,虽然保持了边缘区域信息,但无法实现细节和噪声的有效平滑.

图 1 原图像与出现光晕的图像进行比较

Fig. 1 Comparison between original image and image with halo

根据以上问题的分析,本文提出一种基于显著图的保边滤波算法,利用显著性图边缘区域突出的特点以控制该点标准差 σ_r. 在边缘区域,σ_r 取较小值;在非边缘区域,σ_r 取较大值,从而更好地实现滤波算法的保边特性,避免传统保边滤波法中的光晕现象.

本文算法中假设 $\sigma_r = J(i)$,在不同区域使得 $J(i)$ 值不同:较为显著的边缘部分,$J(i)\approx 0$,保持边缘清晰. 而在非边缘部分,$J(i)$ 较大,模糊细节和噪声. 式(2)变为如下:

$$w_r(I_i,I_j) = \exp\left(-\frac{|I_i - I_j|^2}{2J(i)^2}\right) \tag{5}$$

2 基于频率域的显著图

显著图是一种模拟生物视觉注意机制的选择性模型,显著图直接反映了图像区域引起视觉注意的强弱程度. 显著图是由视觉特殊性以及图像中的颜色、亮度、边缘等特征所决定的,计算机视觉领域利用数学模型模拟了人类视觉的注意力机制,提出了自底向上和自顶向下的显著特征估计算法,显著特性估计目前是自适应压缩[12]、兴趣区域提取[13]、目标检测与识别[14-15]任务中重要的步骤.

根据算法对于效率以及噪声不敏感的要求,本文采用了基于频率域的显著性分析算法[16],此算法比传统的显著图算法保存了更多的高频信息,尤其是边缘部分,如图 2 所示. 该方法具有的特点:边缘显著;对噪声不敏感;输出完整分辨率的显著图;高效性.

图 2 突出显著边缘的显著图

Fig. 2 Original image and saliency map

本文重点关注的是其突出显著边缘的特点,利用这个特点以简化双边滤波中 σ_r 的选取工作. 在 CIELab 颜色空间上进行计算,每个像素取 $[L,a,b]^T$ 向量,首先对图像进行高斯滤波,然后

计算其平均灰度向量，并与原图像的高斯滤波结果作差得到每个像素的显著向量，最后计算其显著向量的欧式距离并进行灰度拉伸得到显著图，具体算式如下：

$$S(x,y)=\|\boldsymbol{I}_{\mathrm{u}}-\boldsymbol{I}_{\mathrm{Guass}}(x,y)\| \tag{6}$$

式中，$\boldsymbol{I}_{\mathrm{u}}$ 为图像 I 算术平均灰度向量；$\boldsymbol{I}_{\mathrm{Guass}}$ 为原始图像高斯模糊后的结果；$S(x,y)$ 为图像 I 的显著值；$\|\ \|$ 为欧式距离.

3　阈值分析

如图 3 所示，将一维信号显著值 S 通过均值滤波分为两部分：基本层和细节层，图 4 是细节层的放大图，细节层包含了噪声和细节高频信息.

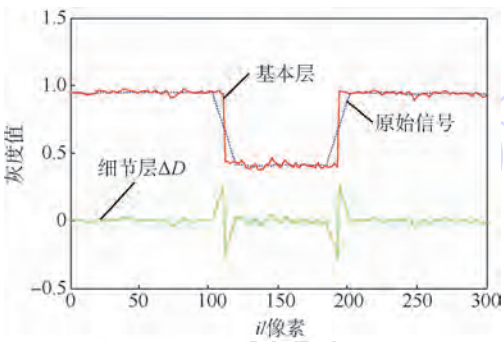


图 3　均值滤波获取的细节层和基本层

Fig. 3　Detail layer and base layer by average filtering

将细节层的 i 位置的灰度值定义为

$$\Delta D_i=S_i-\bar{u}_\omega \tag{7}$$

式中，ΔD_i 为 i 位置的高频信息；S_i 为位置 i 的显著值；$\bar{u}_\omega$ 为 i 像素所在窗口均值.

本文通过 ΔD_i 来区分边缘区域和非边缘区域，如图 4 所示，本文设定阈值 σ，认为当 $\Delta D_i>\sigma$ 时 i 位置像素属于边缘区域，当 $\Delta D_i<\sigma$ 时 i 位置像素属于非边缘区域. 本文取阈值 $\sigma=\bar{S}$，$\bar{S}$ 为整个显著图细节层灰度的平均值.

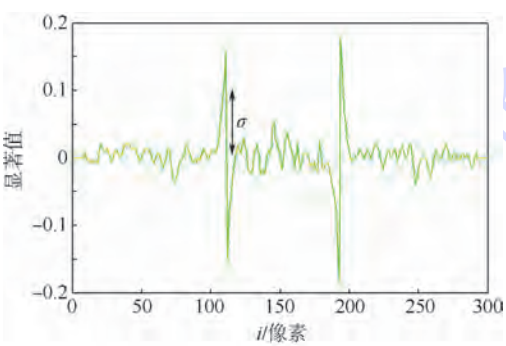


图 4　细节层 ΔD 放大

Fig. 4　Detail layer ΔD enlarger

4　自适应灰度因子

根据上文分析，本文建立 $J(i)$ 数学模型：

$$J(i)=0.5-0.5\cdot \mathrm{sgn}(|\Delta D_i|-\sigma)\cdot(||\Delta D_i|-\sigma|)^{\tau} \tag{8}$$

式中，$\Delta D_i=S_i-\bar{u}_\omega$；$\tau$ 为控制曲线陡度的参数，决定图像边缘和非边缘区域阈值处的模糊程度，如图 5 所示，当 $\tau=0.01$ 时图像处理较为生硬，而 $\tau=0.1$ 时部分边缘会被模糊. 本文经过几组图像的测试，选取 $\tau=0.05$，此时图像的边缘保留和细节平滑效果最佳.

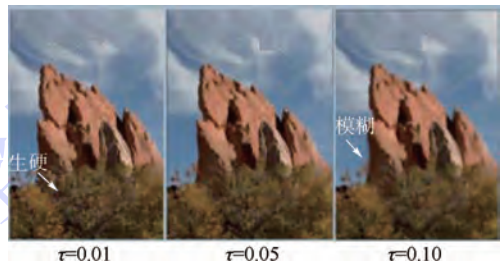


图 5　τ 取不同值的模糊结果

Fig. 5　Results with different τ

由图 6 中曲线可知，当横坐标 $\Delta D_i<\sigma$ 时，进入非边缘区域，随着横轴高频信息 ΔD_i 的减小，$J(i)$ 逐渐增大. 当 $\Delta D_i>\sigma$ 时进入边缘区域，并且随着高频信息 ΔD_i 增大，$J(i)$ 逐渐减小，并趋向于 0.

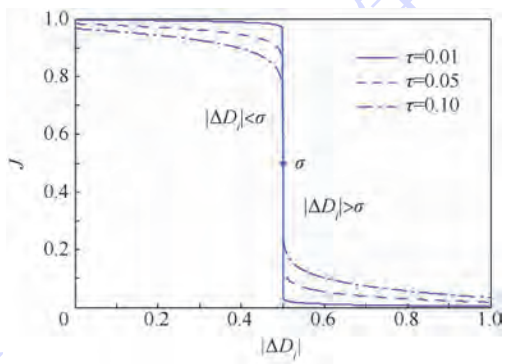


图 6　不同 τ 情况下，σ 取 0.5 的 $J(i)$

Fig. 6　$J(i)$ with $\sigma=0.5$ with different τ

5　保边滤波原理分析

针对“边缘区域”，存在 $\Delta D_i>\sigma$，此区域内边缘灰度复杂度较高，因此 $J(i)$ 较小，趋向于 0，除了中心像素以外的距离权重因子 $w_r(I_i,I_j)\approx 0$，此时窗口内不做任何处理，保留了此区域信息，实现保边特性.

针对“非边缘区域”，此区域包含了噪声细节和平坦部分，复杂度较小，存在 $\Delta D_i\leqslant\sigma$，因此 $J(i)$ 较大，区域内像素的 $w_r(I_i,I_j)\approx 1$，q_i 退化为高斯滤波结果，实现细节和噪声的有效平滑.

北航学报 赠阅

6 一维信号分析

由一维信号可见,引导滤波和双边滤波会模糊信号的高频部分,图 7(a)和图 7(b)中在像素索引为 100 和 200 左右位置的细节层出现了光晕的环状信息,而图 7(c)中,本文提出的算法在高频处不会出现光晕信息,由此可见本文算法具有较好保边特性.

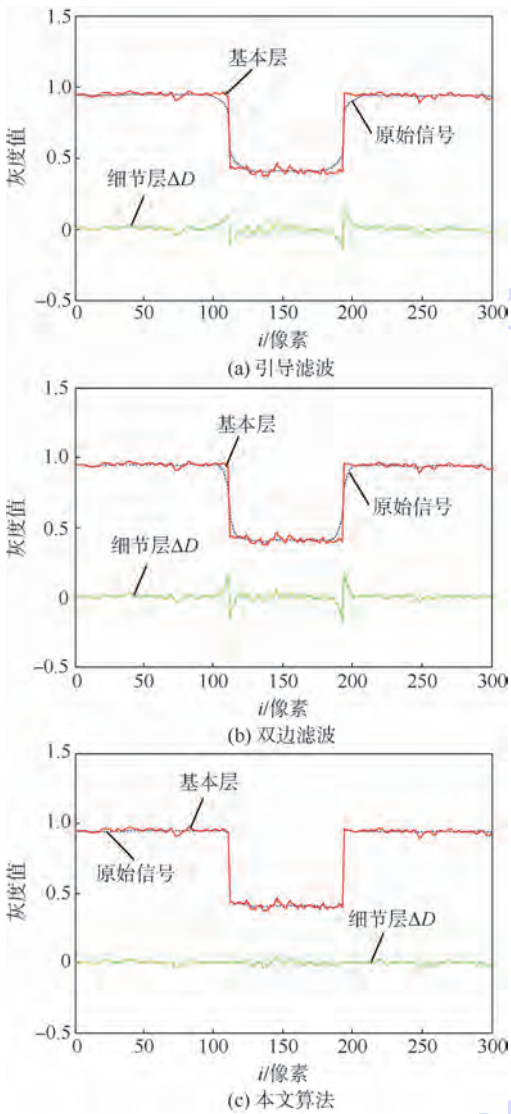


图 7 应用于一维信号的 3 种滤波结果比较

Fig. 7 Comparison of the three filtering results of one-dimensional signal

7 实验及其应用

7.1 风格化

图像风格化是计算机图形学中的非真实感绘制的热点问题之一,在图像处理、娱乐、绘画等领域有着较广泛的应用. 首先对图像进行多尺度分解,只保留基本层,便可产生抽象的效果,利用 Canny 算子求出边缘线条,添加到图像中得到风格化效果,如图 8 所示.

图 8 运用本文算法风格化的效果图

Fig. 8 Effect drawing of stylization using our method

7.2 多尺度图像增强

多尺度图像增强,是按照一定的规则把图像分成多个尺度,对不同尺度的细节进行灵活控制,最后结合不同尺度的层次,实现图像质量的有效改善. 将输入图像迭代,应用本文滤波算法,得到不同尺度图像 n_i,随着迭代次数增加,n_i 平滑程度逐级增大,即越来越多的细节被平滑掉. 首先将其分解成为一个光滑的基本层 n_k 和多个细节层 $d_1, d_2, \cdots, d_{k-1}$:

$$d_{i-1} = n_{i-1} - n_i \qquad i = 1, 2, \cdots, k \tag{9}$$

那么 $d_1, d_2, \cdots$ 表示尺度递增的细节层;n_k 表示基本层. 最后将每个细节层各自乘以相应的放大系数,叠加到基本层上,得到最后的增强图像:

$$I = n_k + \sum_{i=1}^{k-1} \delta_i \cdot d_i \tag{10}$$

式中,I 为输出图像;δ_i 为第 i 层细节层的放大系数. 本文实验过程中取 $k = 4$,即得到 3 个细节层 d_1, d_2, d_3 和一个基本层 n_4.

$$I = n_4 + 2d_3 + 3d_2 + 4d_1 \tag{11}$$

实验中与双边滤波法的结果进行比较,如图 9所示,可见本文算法避免了增强过程中光晕现象的出现,并且细节部分增强效果更为明显.

图 9 多尺度增强

Fig. 9 Multi-scale enhancement

北航学报 赠阅

7.3　图像降噪

将本文算法应用到两幅图片上,并与引导滤波和双边滤波的结果进行比较,从图 10 中可见,双边滤波和引导滤波导致边缘部分的模糊,而本文算法则会保持边缘部分清晰.

7.4　HDR 色调映射

将本文的滤波算法应用于 HDR 图像的色调映射. 图 11 显示了本文算法与引导滤波的结果比较,本文算法将图像的亮度进行了有效的压缩,同时保持了原图的细节,避免了光晕的现象.

图 10　平滑结果

Fig. 10　Smoothing results

图 11　HDR 色调映射

Fig. 11　HDR Tone-mapping

8　结　论

本文提出了基于显著特性的保边滤波算法,利用显著图边缘突出的特点,自适应地进行光滑程度的限制,在显著值较大的地方(边缘区域),双边滤波的灰度标准差较小,不作平滑处理;而显著值较小的地方(非边缘区域),双边滤波的灰度标准差较大,作平滑处理. 实验结果表明:

1) 本文算法有效地避免了平滑过程中显著边缘周围的光晕现象,抑制噪声的同时保持了图像中显著边缘的部分,与引导滤波以及双边滤波进行比较,本文算法在平滑和多尺度增强方面都更为理想;

2) 由于本文算法的保边特性,可以应用于图像细节的分解,在色调映射以及图像风格化等方面也有较好的表现;

3) 利用显著图的特点,简化了区分边缘和非边缘的阈值设定工作;

4) 本文算法是基于局部平均加权的滤波法,满足并行计算的要求,因此下一步计划在 GPU (Graphic Processing Unit)上实现本文算法.

参考文献 (References)

[1] Gonzalez R C, Woods R E. Digital image processing[M]. 2nd ed. New York: Prentice Hall, 2002: 97-98.

[2] Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms[J]. Physics D: Nonlinear Phenomena, 1992, 60(1): 259-268.

北航学报 赠阅

[3] Xu L, Lu C, Xu Y, et al. Image smoothing via L0 gradient minimization[J]. ACM Transactions on Graphics (TOG), 2011, 30(6):252-265.

[4] Farbman Z, Fattal R, Lischinski D, et al. Edge-preserving decompositions for multi-scale tone and detail manipulation[J]. ACM Transactions on Graphics(TOG), 2008, 27(3):67:1-67:10.

[5] Paris S, Hasinoff S W, Kautz J. Local Laplacian filters: edge-aware image processing with a Laplacian pyramid[J]. ACM Transactions on Graphics(TOG), 2011, 30(4):68:1-68:12.

[6] Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion[J]. Pattern Analysis and Machine Intelligence, 1990, 12(7):629-639.

[7] Tomasi C, Manduchi R. Bilateral filtering for gray and color images[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 1998:839-846.

[8] Porikli F. Constant Time $O(1)$ bilateral filtering[C]//26th IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2008:1-8.

[9] Yang Q, Tan K H, Ahuja N. Real-time $O(1)$ bilateral filtering[C]//2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2009:557-564.

[10] Adams A, Gelfand N, Dolson J, et al. Gaussian KD-trees for fast high-dimensional filtering[J]. ACM Transactions on Graphics (TOG), 2009, 28(3):21:1-21:12.

[11] He K, Sun J, Tang X O. Guided image filtering[C]//Computer Vision-ECCV. Heidelberg: Springer Verlag, 2010:1-14.

[12] Christopoulos C, Skodras A, Ebrahimi T. The JPEG2000 still image coding system: an overview[J]. Consumer Electronics, 2000, 46(4):1103-1127.

[13] Han J, Ngan K N, Li M, et al. Unsupervised extraction of visual attention objects in color images[J]. Circuits and Systems for Video Technology, 2006, 16(1):141-145.

[14] Rutishauser U, Walther D, Koch C, et al. Is bottom-up attention useful for object recognition? [C]//Computer Vision and Pattern Recognition, 2004. Washington, D. C.: IEEE, 2004, 2: II-37-II-44.

[15] 李志成, 秦世引, Itti Lauren. 遥感图像的显著-概要特征提取与目标检测[J]. 北京航空航天大学学报, 2010(6):659-662.
Li Z C, Qin S Y, Itti L. Extraction of saliency-gist features and target detection for remote sensing images[J]. Journal of Beijing University of Aeronautics and Astronautics, 2010(6):659-662(in Chinese).

[16] Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection[C]//2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2009:1597-1604.

Edge-preserving filtering based on saliency map

FENG Ce, DAI Shuling*, ZHAO Shuaihe

(School of Automation Science and Electrical Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: A new edge-preserving filtering algorithm based on the saliency map was proposed to avoid the halo effect along salient edges in traditional edge-preserving filtering. The saliency map was characterized by the salient edges in the image and insensibility to noise. The key idea of our filtering algorithm was to simplify the decision of gray standard deviation in bilateral filtering by using the characteristics. We first obtained a saliency map of the original image, and then blurred the image adaptively according to the local saliency value. If the local saliency value is high, the standard deviation would become small to retain the region. If the local saliency value is low, the standard deviation would become large to smooth the region. Therefore, the filtering proposed could avoid unwanted smoothing near salient edges. Experiments show that the proposed filtering performs better in comparison with guided filtering and bilateral filtering. A variety of applications including high dynamic range (HDR) Tone mapping and stylization by our filtering algorithm are also demonstrated.

Key words: edge-preserving filtering; image smoothing; saliency map; bilateral filtering; guided filtering

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0310

基于修正 M-K 模型的铝合金板材成形极限图预测

杨希英，郎利辉*，刘康宁，蔡高参，郭禅

（北京航空航天大学 机械工程及自动化学院，北京 100191）

摘 要：为准确预测 AA7075-O 铝合金板材的成形极限，将韧性断裂准则和传统M-K模型相结合，提出一种基于韧性断裂准则的修正 M-K 模型. 利用单向拉伸模拟和试验相结合的方法，提取危险单元的应力应变历史并代入 C-L 韧性断裂准则中，得到材料常数. 基于单向拉伸试验所得成形极限点，通过 MATLAB 编程得到修正 M-K 模型的初始厚度不均度. 在常温下，以单向拉伸、宽板弯曲、液压胀形试验获得 AA7075-O 铝合金板材的成形极限曲线，与修正 M-K 模型和传统 M-K 模型计算得到的理论成形极限曲线进行对比，验证了修正 M-K 模型的可行性和准确性.

关 键 词：铝合金板；M-K 模型；韧性断裂准则；初始厚度不均度；成形极限图
中图分类号：V260.1；TG394
文献标识码：A **文章编号**：1001-5965(2015)04-0675-05

成形极限是板材成形中十分重要的性能指标和工艺参数，反映了变形过程中板材在塑性失稳状态前的最大变形程度，对于航空航天飞行器上的薄壁零件成形具有重要的指导意义，也是进行模具设计和工艺设计的主要依据. 在板材成形性能评价方法中，成形极限曲线(FLC)应用最广泛且直观有效.

对成形极限进行的理论研究，概括起来分成以下两类：第 1 类是经典的分叉理论，如 Swift 分散性失稳理论[1]、Hill 集中性失稳理论[2]和 Stören 和 Rice 的方法[3-4]；第 2 类是基于材料局部初始不均匀性是导致颈缩的假设，如 M-K 理论[5]等. 其中，M-K 理论应用最普遍，目前关于成形极限的理论研究绝大多数是建立在 M-K 理论的基础之上[6-8]. 它是以试验观察为基础而提出的，但是初始厚度不均度究竟如何确定并无准确的方法. 因此，由于初始厚度不均值的不同，M-K 理论与试验结果的符合程度难免具有一定的主观随意性. 同时，M-K 模型中关于集中颈缩时刻判定的准则也需要进一步讨论.

与 FLC 相比，利用韧性断裂准则预测板料成形极限的优点是可以更加方便地考虑应变路径的变化[9]. 到目前为止，多种不同的韧性断裂准则[10]被提出. 近年来，韧性断裂准则被应用到双向拉伸[11]、液压成形[12]、拉深[13]等成形过程的成形极限预测. 一般而言，金属板料的韧性越好，其断裂成形极限曲线 FLCF(Forming Limit Curves at Fracture)越趋近于直线；韧性越差，则其 FLCF 形状越复杂[14]. 因此，目前所用的几种韧性断裂准则具有一定的局限性[15].

在本文中，为准确预测 AA7075-O 铝合金板材的成形极限图，将韧性断裂准则和传统 M-K 模型相结合，提出一种基于韧性断裂准则的修正 M-K模型. 利用单向拉伸模拟和试验相结合的方

收稿日期：2014-05-29；**录用日期**：2014-07-25；**网络出版时间**：2014-09-05 16:41
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0310.html
基金项目：国家自然科学基金资助项目(50975014)；中俄国际合作与交流 NSFC-RFBR 基金资助项目(51010166)
作者简介：杨希英(1985—)，男，河北衡水人，博士生，yxiyingbuaa@163.com
* **通讯作者**：郎利辉(1970—)，男，河南新乡人，教授，lang@buaa.edu.cn，主要研究方向为先进成形技术及装备.

引用格式：*杨希英，郎利辉，刘康宁，等. 基于修正 M-K 模型的铝合金板材成形极限图预测[J]. 北京航空航天大学学报，2015，41(4)：675-679. Yang X Y，Lang L H，Liu K N，et al. Prediction of forming limit diagram of AA7075-O aluminum alloy sheet based on modified M-K model[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：675-679 (in Chinese).*

法,提取危险单元的应力应变历史并代入 C-L 韧性断裂准则中,得到材料常数. 通过 MATLAB 编程得到修正 M-K 模型的初始厚度不均度. 在常温下,以单向拉伸、宽板弯曲、液压胀形获得 AA7075-O 铝合金板材的成形极限试验点,与修正 M-K 模型和传统 M-K 模型计算得到的理论成形极限曲线进行对比,验证了修正 M-K 模型的可行性和准确性.

1 修正 M-K 模型建立

1.1 传统 M-K 假设

传统 M-K 模型以凹槽假设为理论核心,认为板材具有不可避免的厚度不均性,这种不均性用凹槽表示,如图 1 所示. 根据假设,板材的集中性失稳是由板材表面初始存在的缺陷引起的. 该模型广泛用于预测平面应力条件下的板材成形极限,理论假设包括以下几点[5].

1) 简单加载条件:A 区主应力及主应变均成比例增加且在整个变形过程中比值为常数;

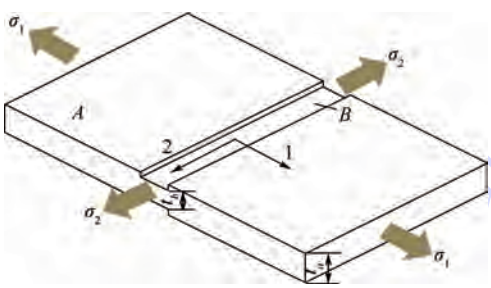
2) 变形协调条件:B 区第二主应变增量 $\mathrm{d}\varepsilon_{2b}$ 与 A 区第二主应变增量 $\mathrm{d}\varepsilon_{2a}$ 相等,即 $\mathrm{d}\varepsilon_{2a}=\mathrm{d}\varepsilon_{2b}=\mathrm{d}\varepsilon_2$;

3) 力平衡条件:A 区与 B 区第一主方向力始终平衡,即 $F_{1a}=F_{1b}$.

为考虑板材的厚向异性,在模型推导中采用 Hill'48 塑性屈服准则[16]:

$$\bar{\sigma}^2=\frac{\xi(\sigma_1-\sigma_2)^2+(\sigma_2-\sigma_3)^2+(\sigma_3-\sigma_1)^2}{1+\xi} \tag{1}$$

式中,ξ 为厚向异性指数;σ_1,σ_2,σ_3 分别为 1,2,3 主应力分量.



A 区—变形安全区;B 区—不均匀变形区;

t_a,t_b—变形过程中 A,B 区板材的厚度

图 1 M-K 模型理论框架图

Fig. 1 Theoretical frame diagram of M-K model

1.2 韧性断裂准则

根据应变能理论,材料的塑性变形程度可以用总塑性功来体现,其表达式[17]为

$$U=\int_0^{\bar{\varepsilon}}\bar{\sigma}\,\mathrm{d}\bar{\varepsilon} \tag{2}$$

式中,$\bar{\sigma}$ 为等效应力;$\bar{\varepsilon}$ 为等效应变;U 为总塑性功. 在采用积分函数的基础上,该判据具有考虑非线性应变路径变化的优点.

Cockcroft 和 Latham(C-L)等[17]考虑最大拉伸主应力在材料变形至断裂过程中的作用,提出如下公式:

$$\int_0^{\bar{\varepsilon}_f}\sigma^*\,\mathrm{d}\bar{\varepsilon}=C \tag{3}$$

式中,σ^* 为最大拉伸主应力. 这个准则应用最为广泛,在本文中,将以此(C-L 准则)为基础与传统 M-K 模型假设结合,进行板材成形极限曲线的预测. 同时,为保证积分值始终为正值,对 C-L 准则的左半部分取绝对值.

1.3 推导过程

假设图 1 中平面内应力主轴方向的主应力比值为 $\alpha=\sigma_2/\sigma_1(0\leqslant\alpha\leqslant1)$,$G_1=2\xi/(1+\xi)$,在平面应力状态下,则式(1)可表示为

$$\bar{\sigma}=\sigma_1(1+\alpha^2-G_1\alpha)^{1/2} \tag{4}$$

等效应力与最大主应力方向的比值为

$$\phi=\bar{\sigma}/\sigma_1=(1+\alpha^2-G_1\alpha)^{1/2} \tag{5}$$

由 Levy-Mises 增量理论关系式,有

$$\frac{\mathrm{d}\bar{\varepsilon}}{(1+\xi)\phi}=\frac{\mathrm{d}\varepsilon_1}{(1+\xi)-\xi\alpha}=\frac{\mathrm{d}\varepsilon_2}{(1+\xi)\alpha-\xi}=-\frac{\mathrm{d}\varepsilon_3}{1+\alpha} \tag{6}$$

记 ρ 为面内应变比值,β 为等效应变与第一主应变比值. 分别表示如下:

$$\rho=\frac{\mathrm{d}\varepsilon_2}{\mathrm{d}\varepsilon_1}=\frac{(1+\xi)\alpha-\xi}{(1+\xi)-\xi\alpha} \tag{7}$$

$$\beta=\frac{\mathrm{d}\bar{\varepsilon}}{\mathrm{d}\varepsilon_1}=\frac{(1+\xi)\phi}{(1+\xi)-\xi\alpha} \tag{8}$$

根据力平衡条件的理论假设,可得:

$$\sigma_{1a}t_a=\sigma_{1b}t_b \tag{9}$$

板材厚度法向方向的应变,即

$$\varepsilon_3=\ln(t/t_0) \tag{10}$$

可表示为

$$t=t_0\exp(\varepsilon_3) \tag{11}$$

式中,t 为变形过程中的板材厚度;t_0 为初始的板材厚度.

已知 $f_0=t_{b0}/t_{a0}$,可得变形过程中板材厚度不均度为

$$f=\frac{t_b}{t_a}=\frac{t_{b0}}{t_{a0}}\exp(\varepsilon_{3b}-\varepsilon_{3a})=f_0\exp(\varepsilon_{3b}-\varepsilon_{3a}) \tag{12}$$

结合式(1)、式(9)、式(12),得到以下平衡方程:

北航学报 赠阅

$$(\bar{\varepsilon}_a+\Delta\bar{\varepsilon}_a)^n(\Delta\bar{\varepsilon}_a)^m\phi_b=(\bar{\varepsilon}_b+\Delta\bar{\varepsilon}_b)^n(\Delta\bar{\varepsilon}_b)^m\phi_a f_0\exp(\varepsilon_{3b}-\varepsilon_{3a}) \tag{13}$$

对式(13)采用 Newton 迭代法进行计算,分析板材集中性失稳行为,得到板材变形过程中 A 区和 B 区的各应力应变值.此时,引入韧性断裂准则的通式:

$$\int_0^{\bar{\varepsilon}_f} f(\bar{\sigma},\sigma^*,\cdots)\,\mathrm{d}\bar{\varepsilon}=C \tag{14}$$

对 B 区进行判断,当 $\int_0^{\bar{\varepsilon}_f} f(\bar{\sigma},\sigma^*,\cdots)\,\mathrm{d}\bar{\varepsilon}\geqslant C$ 时,得到对应时刻的 A 区面内主应变 ε_{1a},ε_{2a},作为相应 $\alpha=\sigma_2/\sigma_1$ 值下的成形极限点.对 $\alpha=\sigma_2/\sigma_1$ 在 $0\leqslant\alpha\leqslant1$ 范围内进行遍历,得到板材的成形极限曲线.

2　韧性断裂准则中材料常数的确定

本文所用材料为 1 mm 厚的 AA7075-O 铝合金板材,对其进行单向拉伸试验以确定应力应变曲线,试件的几何尺寸如图 2 所示.试验在北京航空航天大学的 WDW-100 电子万能试验机上进行(如图 3 所示).

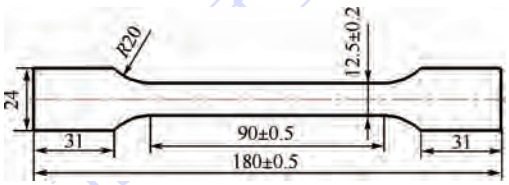


图 2　AA7075-O 铝合金单拉试样的几何形状及尺寸

Fig. 2　Uniaxial tensile specimen geometry and dimension of AA7075-O aluminum alloy sheet

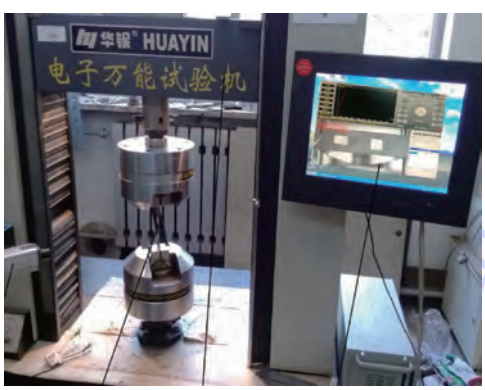


图 3　单向拉伸试验设备

Fig. 3　Uniaxial tensile experimental equipment

针对单拉试验数据,考虑各参数对 AA7075-O 铝合金板材力学性能的影响,对 Swift 模型进行修正,得到本构方程:

$$\bar{\sigma}=K\bar{\varepsilon}^n\dot{\bar{\varepsilon}}^m\exp(q\dot{\bar{\varepsilon}}+p\bar{\varepsilon}) \tag{15}$$

本构方程中的材料参数 K,n,m,q,p 通过拟合确定,同时计算得到板材厚向异性指数和弹性模量 E,如表 1 所示.在拉伸速率 $v=15$ mm/min 情况下,拟合得到应力应变曲线与试验结果对比如图 4 所示,结果符合较好.

表 1　材料性能参数

Table 1　Material performance parameters

参数	E/GPa	n	m	K/MPa	ξ	q	p
数值	71	0.458	0.023	804	0.713	35.48	2.19

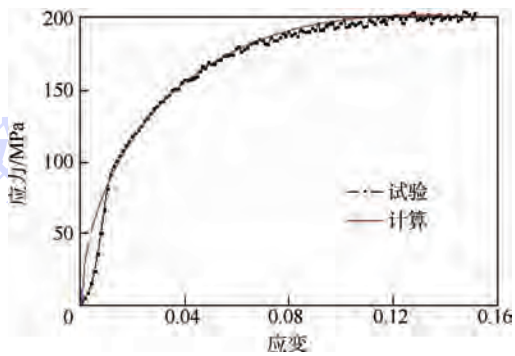


图 4　拟合应力应变曲线与试验数据比较

Fig. 4　Comparison of stress-strain curve between fitting and experimental data

为确定韧性断裂准则中的材料常数,本文采用单向拉伸数值模拟和试验对比的方法.利用 Abaqus/Explicit 进行有限元分析,材料参数采用表 1 中的数据.数值模拟得到试件的载荷-位移曲线,与试验得到的曲线对比结果如图 5 所示,二者符合较好,该条件下的模拟参数可信.图 6 为试件

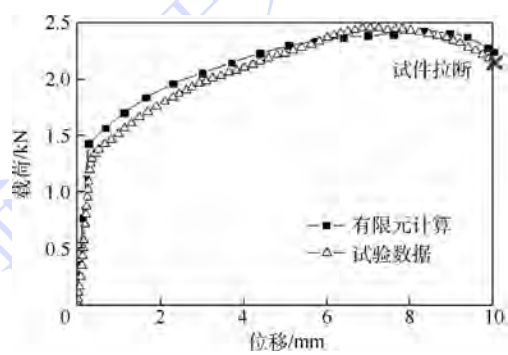


图 5　有限元计算与试验得到的载荷-位移曲线对比

Fig. 5　Comparison of loading-displacement curve between FEM and experimental data

图 6　临近拉断时刻等效塑性应变分布云图

Fig. 6　Equivalent plastic strain distribution close to failure

拉断时等效应变分布云图，将危险单元在变形过程中的应力应变历史代入 C-L 准则 $\int_0^{\bar{\varepsilon}_f}\sigma^*\mathrm{d}\bar{\varepsilon}=C$，可以得到材料常数 $C=192.43$.

3 成形极限理论预测与试验验证

在修正 M-K 模型中，为确定初始厚度不均度值，利用 MATLAB 编程获得. 流程如下：

1）设定 f_0 初值，遍历 $\alpha=\sigma_2/\sigma_1(0\leqslant\alpha\leqslant1)$，得到板材的理论成形极限曲线；

2）比较单向拉伸试验所得成形极限点与理论成形极限曲线差值，差值为正，则增大 f_0 值；差值为负，减小 f_0 值；直至差值小于 0.005 为止；

利用该方法确定 $f_0=0.955$，并得到理论成形极限曲线. 同时，利用单向拉伸、宽板弯曲、液压胀形试验得到 AA7075-O 铝合金板材在常温下的成形极限，试验件如图 7 所示.

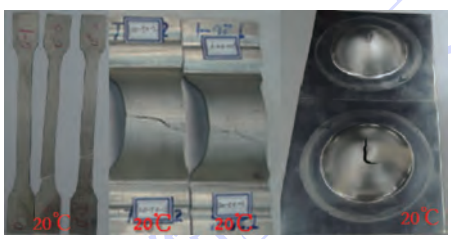


图 7 单向拉伸、宽板弯曲、胀形试验件

Fig. 7 Test pieces for uniaxial tensile, wide plate bending and hydrobulging

图 8 为利用修正的 M-K 模型及传统 M-K 模型分别得到的理论成形极限曲线与成形极限试验点的对比图. 由图可知，与传统 M-K 模型相比，修正的 M-K 模型能更好地预测 AA7075-O 铝合金板材的成形极限曲线，与成形极限试验点贴合度更高.

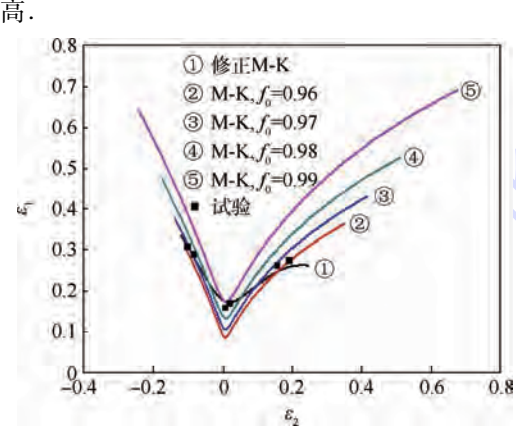


图 8 成形极限对比图

Fig. 8 Comparison of forming limit

4 对比分析

在传统 M-K 模型中，板材破裂失效时的判断准则为 B 区第一主应变增量 $\Delta\varepsilon_{1b}$ 与 A 区第一主应变增量 $\Delta\varepsilon_{1a}$ 的比值是常数，即，$\Delta\varepsilon_{1b}/\Delta\varepsilon_{1a}=10$. 该判断准则表明 $\Delta\varepsilon_{1b}/\Delta\varepsilon_{1a}$ 值与加载路径无关. 同时，由修正 M-K 模型所得到的 $\Delta\varepsilon_{1b}/\Delta\varepsilon_{1a}$ 值如图 9 所示. 从中可以看出，随着 α 从 0 到 1 遍历，$\Delta\varepsilon_{1b}/\Delta\varepsilon_{1a}$ 值先增大后减小. 在 $\alpha=0.5$ 附近，该值达到极值. $\Delta\varepsilon_{1b}/\Delta\varepsilon_{1a}$ 值的变化趋势同样可以在成形极限图中得到反映. 如图 8 所示，$f_0=0.96$ 时的传统 M-K 模型预测的成形极限曲线，在单向拉伸和双向拉伸点处与试验点接近；在平面应变处，远低于试验值. 由此可知，只有通过提高平面应变加载路径下的 $\Delta\varepsilon_{1b}/\Delta\varepsilon_{1a}$ 值作为判断准则，才能提高该处的成形极限值. 修正 M-K 模型中 $\Delta\varepsilon_{1b}/\Delta\varepsilon_{1a}$ 值的变化趋势与实际相符，表明以 C-L 韧性断裂准则能更加准确地反应 A 区凹槽的破坏，用于预测板材的成形极限曲线.

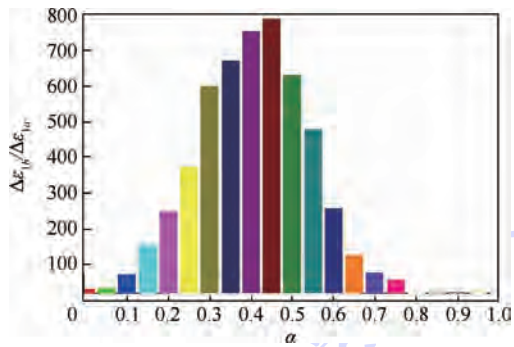


图 9 α 从 0 到 1 遍历下 $\Delta\varepsilon_{1b}/\Delta\varepsilon_{1a}$ 值分布图

Fig. 9 Distribution of $\Delta\varepsilon_{1b}/\Delta\varepsilon_{1a}$ when α ranges from 0 to 1

5 结 论

本文将传统 M-K 模型框架和基于应变能的韧性断裂准则相结合，提出一种修正的 M-K 模型预测板材的成形极限曲线，并进行了数值计算和试验对比验证.

1）利用单向拉伸数值模拟和试验相结合的方法，得到 C-L 韧性断裂准则中的材料常数. 基于单向拉伸试验所得成形极限点，通过 MATLAB 编程得到修正 M-K 模型的初始厚度不均度. 结果表明，针对同种规格板材及试验条件，材料常数及初始厚度不均度为定值.

2）试验得到 AA7075-O 铝合金板材的成形极限，对比表明，M-K 模型中 A 区凹槽破裂失效时的判断准则应修正为 C-L 韧性断裂准则，与试验符合程度更高，验证了本文提出的修正 M-K 模型的正确性.

参考文献（References）

[1] Swift H W. Plastic instability under plane stress[J]. Journal of

北航学报 赠阅

the Mechanics and Physics of Solid,1952,1(1):1-18.
[2] Hill R. On discontinuous plastic states with special reference to localized necking in thin sheets[J]. Journal of Mechanics and Physics of Solids,1952,1(1):19-31.
[3] Stören S,Rice J. Localized necking in thin sheets[J]. Journal of the Mechanics and Physics of Solid,1975,23(6):421-441.
[4] Zhu X H,Weinmann K,Chandra A. A unified bifurcation analysis of sheet metal forming limits[J]. Journal of Engineering Materials and Technology,2001,123(3):329-333.
[5] Marciniak Z, Kuczynski K. Limit strains in the processes of stretch-forming sheet metal[J]. International Journal of Mechanical Science,1967,9(3):609-620.
[6] 杜平梅,郎利辉,刘宝胜,等. 基于 M-K 模型的成形极限预测及参数影响[J]. 塑性工程学报,2011,18(5):84-89.
Du P M,Lang L H,Liu B S,et al. Theoretical prediction and parameter influence of FLDs based on M-K model[J]. Journal of Plasticity Engineering,2011,18(5):84-89(in Chinese).
[7] 马高山,万敏,吴向东. 基于 M-K 模型的铝锂合金热态下成形极限预测[J]. 中国有色金属学报,2008,18(6):980-984.
Ma G S,Wan M,Wu X D. Theoretical prediction of FLDs for Al-Li alloy at elevated temperature based on M-K model[J]. The Chinese Journal of Nonferrous Metals,2008,18(6):980-984(in Chinese).
[8] Assempour A,Nejadkhaki H K,Hashemi R. Forming limit diagrams with the existence of through-thickness normal stress[J]. Computational Materials Science,2010,48(3):504-508.
[9] 陈劼实,周贤宾. 成形极限预测韧性断裂准则及屈服准则的影响[J]. 北京航空航天大学学报,2006,32(8):969-973.
Chen J S,Zhou X B. Suitability of some ductile fracture criteria and yield criteria in forming limit prediction[J]. Journal of Beijing University of Aeronautics and Astronautics,2006,32(8):969-973(in Chinese).
[10] 谢延敏,于沪平,陈军,等. 板料成形中韧性断裂准则应用研究进展[J]. 工程设计学报,2007,14(1):6-10.
Xie Y M,Yu H P,Chen J,et al. Recent advances of research on application of ductile fracture criteria in sheet metal forming process[J]. Journal of Engineering Design,2007,14(1):6-10(in Chinese).
[11] Takuda H,Mori K. Finite element analysis of limit strains in biaxial stretching of sheet metals allowing for ductile fracture[J]. International Journal of Machine Tools and Manufacturing,2000,42(4):785-798.
[12] Lei L P,Kim J. Bursting failure prediction in tube hydroforming process by using rigid-plastic FEM combined with ductile fracture criterion[J]. International Journal of Mechanical Sciences,2002,44(7):1411-1428.
[13] 高付海,桂良进,范子杰. 基于韧性准则的金属板料冲压成形断裂模拟[J]. 工程力学,2010,27(2):204-208.
Gao F H,Gui L J,Fan Z J. Numerical simulation of the fracture in sheet metal stamping based on ductile criterion[J]. Engineering Mechanics,2010,27(2):204-208(in Chinese).
[14] Vallellano C,Morales D,Garcia-Lomas F J. A study to predict failure in biaxially stretched sheets of aluminum alloy 2024-T3[J]. Materials and Manufacturing Processes,2008,23(3):303-310.
[15] Jain M,Allin J,Lloyd D J. Fracture limit prediction using ductile fracture criteria for forming of an automotive aluminum sheet[J]. International Journal of Mechanical Sciences,1999,41(10):273-288.
[16] Hill R. A theory of the yielding and plastic flow of anisotropic metals[C]//Proceedings of the Royal Society of London Series A. London:The Royal Society,1948:281-297.
[17] Takuda H,Mori K,Hatta N. The application of some criteria for ductile fracture to the prediction of the forming limit of sheet metals[J]. Journal of Materials Processing Technology,1999,95(1-3):116-121.

Prediction of forming limit diagram of AA7075-O aluminum alloy sheet based on modified M-K model

YANG Xiying, LANG Lihui*, LIU Kangning, CAI Gaocan, GUO Chan

(School of Mechanical Engineering and Automation, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: To accurately predict the forming limit diagram of AA7075-O aluminum alloy sheet, a modified M-K model was put forward based on the ductile fracture criterion and the traditional M-K model. By using the method which combined the uniaxial tension simulation with experiment, the history of stress-strain of the dangerous element was substituted into the C-L ductile fracture criterion which was implemented into Abaqus/Explicit to get the material constant. Based on the forming limit data obtained by the uniaxial tensile, the initial thickness imperfection in the modified M-K model was obtained by the method which was programed in MATLAB. At room temperature, the experimental forming limit data of AA7075-O aluminum alloy sheet were tested through the uniaxial tensile, wide plate bending and hydraulic bulging experiments. Meanwhile, the theoretical forming limit curves obtained by the modified M-K model and the traditional M-K model were computed respectively. By comparison, the modified M-K model is proved to be feasible and more accurate.

Key words: aluminum alloy sheet; M-K model; ductile fracture criterion; initial thickness imperfection; forming limit diagram

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0407

基于负载特征聚类的节能资源调度算法

夏庆新，兰雨晴*，唐甜，肖利民

（北京航空航天大学 计算机学院，北京 100191）

摘 要：基础设施即服务(IaaS,Infrastructure as a Service)平台提供商为用户提供高性能服务的同时,必须考虑如何在不违反服务级别协议(SLA,Service Level Agreement)的前提下,节约云平台的能耗成本.采用基于负载特征聚类的方法,提出一种 IaaS 云平台上保证 SLA 的资源调度算法,最终实现降低 SLA 违反率和节约能耗的目标.具体采用能耗相关的负载特征提取和改进 K-means 聚类分析的研究方法,进行资源调度算法研究,对物理资源进行有效分配,以保证 IaaS 平台节约能耗的要求.实验验证方面,通过扩展 CloudSim 模拟实验平台,对本研究算法与改进 BFD(Best Fit Decreasing)算法进行比较,得出本研究算法在 SLA 违反率和节能方面更优.

关 键 词：基础设施即服务；负载；聚类；节能；服务级别协议
中图分类号：V221+.3；TB553
文献标识码：A **文章编号**：1001-5965(2015)04-0680-06

云计算的重要优势体现在能够根据用户按需即付的模式部署资源进行支撑计算资源虚拟化.近年来,云计算的发展和扩张,导致了数以千计的大规模数据中心的建立,造成了巨大的能源消耗[1].在大规模计算设施中,如集群和数据中心,问题将更加严重.而能够像电力网一样使得用户按需缴纳使用费用是本文的目标,这是云计算出现时就具有的优势,但是,在全世界云计算基础设施即服务实施过程中,如 Amazon EC2 和 Microsoft Azure等,能耗度量和成本结算等方面,并没有为用户提供真正的像电力网一样便利而透明的模式.而本研究是以节能为前提,在 SLA 违反率控制方面,进行一种基于负载特征聚类分析的资源调度方法的探索.

1 负载及特征描述

1.1 IaaS 云平台负载

IaaS 云平台资源调度能耗研究领域中,从资源的提供者和消费者 2 个视角进行考虑,一方面,物理主机和虚拟机是资源的提供者,物理主机的过载和欠载 2 个状态,直接影响了 IaaS 云平台的 SLA 违反率和能耗成本;另一方面,应用负载是 IaaS 云平台消耗物理资源的直接反映,是资源消耗者的直接反映.因此,可以应用负载分析方法进行 IaaS 云平台 SLA 违反率和能耗成本方面的研究.

当前,由于商业竞争等原因,可供研究分析的大规模的云计算数据集非常少,造成使用负载分析手段的局限.幸运的是,2009 年 Yahoo 公开了 M45 生产集群的日志数据集[2].同年,Google 公布 Google Cloud trace 日志数据集[3],尤其在该数据集的第 2 版中,包含的信息更加丰富,且数据更加规范,本研究将以此数据集作为研究负载.其中包括 12 532 台物理主机的 29 天的大约 250 GB 的虚拟机负载数据.负载数据包括 CPU、内存、网络等负载数据,系统采样时间为 300 s,相关数据格

收稿日期：2014-07-09；**录用日期**：2014-10-10；**网络出版时间**：2014-12-19 16:26
网络出版地址：www.cnki.net/kcms/detail/11.2625.V.20141219.1626.002.html
基金项目：国家自然科学基金重点资助项目（61232009）
作者简介：夏庆新(1978—)，男，辽宁辽阳人，博士生，xiaqingxin@cse.buaa.edu.cn
*通讯作者：兰雨晴(1969—)，男，副教授，内蒙古呼和浩特人，lanyuqing@buaa.edu.cn，研究方向为操作系统、软件工程.

引用格式：夏庆新，兰雨晴，唐甜，等. 基于负载特征聚类的节能资源调度算法[J]. 北京航空航天大学学报，2015，41(4)：680-685. Xia Q X，Lan Y Q，Tang T，et al. Energy-saving resource scheduling algorithm based on workload characteristic clustering [J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：680-685(in Chinese).

式和细节参见文献[4].

1.2　负载特征

应用负载是 IaaS 云平台消耗物理资源的直接反映,而 Google Cloud trace 提供了可供研究的数据基础.从大量真实的系统数据集中分析出可供使用的负载特征,是进行本研究的基础.

进行分析研究前,首先进行负载模型定义.负载模型具体由式(1)~式(7)确定:

$$V=\{v_1,v_2,\cdots,v_i,\cdots\} \tag{1}$$

$$v_i=\{\boldsymbol{\alpha}_i,\boldsymbol{\beta}_i,\boldsymbol{\gamma}_i,\cdots\} \tag{2}$$

$$\begin{cases}\boldsymbol{\varepsilon}_{\boldsymbol{\beta}_i}=\dfrac{\sum\boldsymbol{\beta}_i}{\boldsymbol{\alpha}_i}\\ \boldsymbol{\varepsilon}_{\boldsymbol{\gamma}_i}=\dfrac{\sum\boldsymbol{\gamma}_i}{\boldsymbol{\alpha}_i}\\ \vdots\end{cases} \tag{3}$$

$$\boldsymbol{\rho}_{\mathrm{m}}=\frac{\sum(\boldsymbol{\eta}_i-\boldsymbol{\varepsilon}_{\boldsymbol{\eta}_i})(\boldsymbol{\eta}_j-\boldsymbol{\varepsilon}_{\boldsymbol{\eta}_j})}{\sqrt{\sum(\boldsymbol{\eta}_i-\boldsymbol{\varepsilon}_{\boldsymbol{\eta}_i})^2\cdot\sum(\boldsymbol{\eta}_j-\boldsymbol{\varepsilon}_{\boldsymbol{\eta}_j})^2}} \tag{4}$$

$$\begin{cases}\mathrm{CDF}(\boldsymbol{v}_{\mathrm{down}})=P(X\leqslant\boldsymbol{\eta}_{\mathrm{down}})=\mathrm{down}\\ \mathrm{CDF}(\boldsymbol{v}_{\mathrm{up}})=P(X\leqslant\boldsymbol{\eta}_{\mathrm{up}})=\mathrm{up}\end{cases} \tag{5}$$

$$V_k=\{(\boldsymbol{v}_i,\boldsymbol{v}_j,\cdots)\mid\mathrm{Kmeans}(\boldsymbol{\varepsilon},\boldsymbol{\rho}_{\mathrm{m}},\boldsymbol{\eta}_{\mathrm{down}},\boldsymbol{\eta}_{\mathrm{up}})\} \tag{6}$$

$$\bar{V}=\{V_1,V_2,\cdots,V_k\} \tag{7}$$

式中,V 为虚拟机应用负载的集合;$\boldsymbol{v}_i$ 为具体每一个负载,具体表示为式(2),由多个维度确定,包括负载长度、CPU 使用率、内存使用率、网络使用率等,分别用 $\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}$ 和 ω 表示.负载长度($\boldsymbol{\alpha}$)使用工作量总量的 MI 值计量,CPU($\boldsymbol{\beta}$)和内存($\boldsymbol{\gamma}$)使用对该负载使用量的采样计量.进而,云负载可以被描述成一个虚拟机负载的集合 V.而负载集合 V 是由每一个由 $\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}$ 的采样向量定义的负载 $\boldsymbol{v}_i$ 组成的.

式(3)~式(5)给出了 4 个负载特征 $\boldsymbol{\varepsilon},\boldsymbol{\rho}_{\mathrm{m}},\boldsymbol{\eta}_{\mathrm{down}}$ 和 $\boldsymbol{\eta}_{\mathrm{up}}$.其中 $\boldsymbol{\varepsilon}$ 为负载采样样本均值(负载使用率的平均波动幅度),$\boldsymbol{\eta}_{\mathrm{up}}$ 和 $\boldsymbol{\eta}_{\mathrm{down}}$ 分别为负载累积分布在对应概率阈值下的负载阈值,而定义 $\boldsymbol{\rho}_{\mathrm{m}}$ 为负载采样样本的互相关系数,计算方法如式(8)所示:

$$\boldsymbol{\rho}_{\mathrm{m}}=\begin{bmatrix}\rho_{11}&\rho_{12}&\cdots&\rho_{1\alpha}\\ \rho_{21}&\rho_{22}&\cdots&\rho_{2\alpha}\\ \vdots&\vdots&\cdots&\vdots\\ \rho_{\alpha1}&\rho_{\alpha2}&\cdots&\rho_{\alpha\alpha}\end{bmatrix} \tag{8}$$

式(6)给出了负载特征的聚类分析过程,通过使用 K-means 聚类算法依据负载特征对负载进行分类.其中 V_k 就是 K 类负载集合 V 的子集之一,如式(7)所示.这样该负载模型的组成和关系可以如下形式化关系表示:$\forall v_1,v_2,\cdots,v_i,\cdots\in V$,$\exists\boldsymbol{\varepsilon},\boldsymbol{\rho}_{\mathrm{m}},\boldsymbol{\eta}_{\mathrm{down}},\boldsymbol{\eta}_{\mathrm{up}}$,有 $\mathrm{Kmeans}(\boldsymbol{\varepsilon},\boldsymbol{\rho}_{\mathrm{m}},\boldsymbol{\eta}_{\mathrm{down}},\boldsymbol{\eta}_{\mathrm{up}})\mapsto\{V_1,V_2,\cdots,V_k\}$,其中 $V_k\subseteq V$.这就是负载模型的数学表示,模型的正确性和优劣通过实验验证.

2　聚类分析

2.1　K-means 聚类算法

K-means 聚类算法被提出已经超过 50 年历史,但目前为止,其仍然是应用最广泛的划分聚类算法之一[5].K-means 容易实施、简单、高效等优点成为其广泛流行的主要原因[6].

K-means 聚类算法有数据子集数目 K、初始聚类质心选取、相似性度量 3 个参数[7].通常选取欧氏距离作为相似性和距离判断准则,计算聚类内各点与聚类质心 c_i 的距离平方和,目标函数为

$$J(V_k)=\sum_{x_i\in V_k}\|x_i-c_i\|^2 \tag{9}$$

对于一般的数据维数和类别个数,K-means 聚类算法是一 NP 难优化问题[8].K-means 聚类算法是一个贪心算法,在多项式时间内,仅能获得局部最优,而不可能获得全局最优.即使这样,对于一个启发式算法已经足够了.此外,基于网格聚类[9]和模型聚类[10]也被众多学者所研究,同时给本文以很大启示.

2.2　改进 K-means 聚类分析

2.2.1　层次 K 值选取

K-means 算法中 K 值的选取尤为重要,其改进研究中,K 值的选取研究占有了很大比重,尤其是 K 值自动选择成为新的研究热点[11].K 值有 2 种极限情况:当 $K=1$ 时,聚类划分为 1 类,为对象全体;当 $K=n$(n 为对象总数)时,聚类划分为 n 类,每个对象元素都是划分类的中心.

本研究根据 1.2 节中分析的负载样本特征,使用层次划分的方法进行 K 值选取,具体过程如图 1 所示.图 1 中树的每一个节点表示一个聚类划分,每一个箭头表示划分为新的聚类划分.本研究采用递增 K 值逐步推演的方法,随着 K 值的增加,欧式距离 D 缩小的比率没有明显增加的时候,作为终止条件.本例中,选取 $K=6$.

2.2.2　特征聚类初始质心选取

聚类初值的选取是 K-means 聚类算法的关键步骤.常见的方法是随机选取初值,但是这样簇集合的质量常常很差[12].

本文的研究采用了 3 种方法进行了实验比较得出适合负载特征聚类的方法.处理选取初始质

北航学报 赠阅

心(Centroids)问题的第 1 种常用技术是增加运行迭代次数,每次使用随机初始质心. 这种策略简单,但是效果不好,取决于数据集和聚类的个数. 第 2 种选择初始质心的方法,随机地选择第 1 个点,然后,对于每个后继初始质心,选择离已经选取过的初始质心最远的点. 使用这种方法的缺点是得到的聚类不够稳定. 第 3 种方法是,取一个样本,并使用层次聚类技术对它聚类. 从层次聚类中提取 K 个簇集合,并用这些簇的中心作为初始值. 这样得到的聚类比较稳定. 本研究分别对此 3 种方法进行实验,如图 2 所示.

图 2 是采用了 200 个样本点,依次使用 3 种方法进行 K 值为 3 的聚类(聚类分别为 Cluster1, Cluster2 和 Cluster3,质心为 Centroids). 可以清楚地看到经过图 2(c)的初始质心选择方法聚类结果最佳.

因此,综合 3 个方法的优缺点,层次聚类选取初始质心的方法是针对云负载样本特征进行聚类分析的方法中最合适的. 本研究选用第 3 种方法作为 K-means 聚类算法的初始质心选取方法.

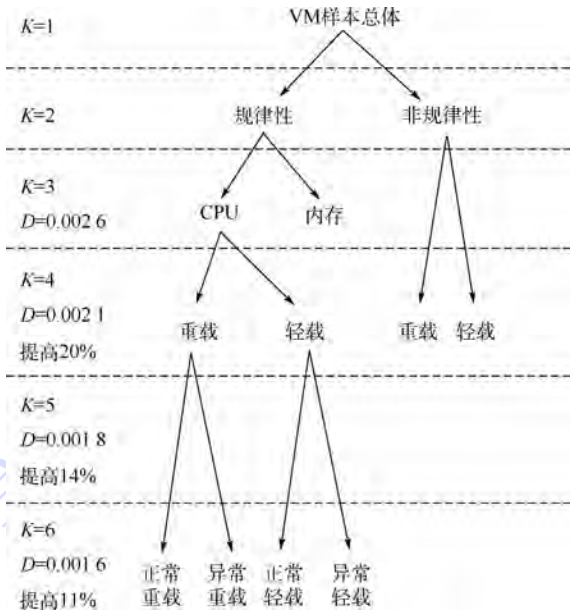


图 1 负载特征分层聚类分析及 K 值选取

Fig. 1 Hierarchical cluster analysis of workload characterisitic and K value choosing

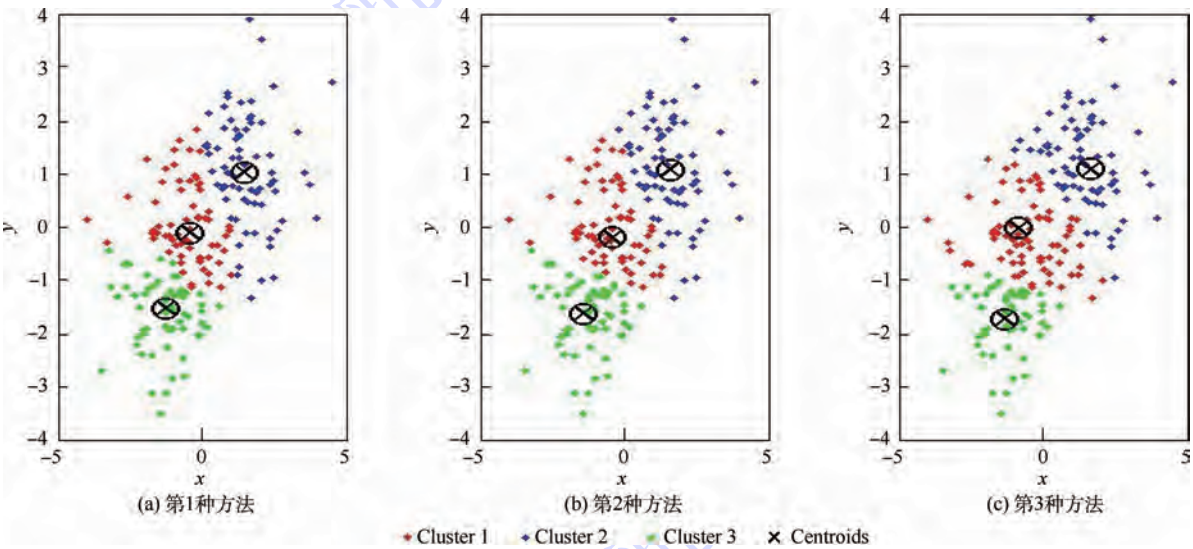


图 2 选择初始质心 3 种方法比较结果

Fig. 2 Comparison results of three methods for initial centroid selection

2.2.3 特征层次聚类

通过使用 Google Cloud Trace 第 2 版数据集作为研究对象,选取 29 天 1 000 个虚拟机的 CPU 和内存使用率负载数据,每个负载长度 α 取288 × 300 s(86 400 s 为 1 天). 经过 CDF 的特征提取,得到 4 个负载特征 $\boldsymbol{\varepsilon}$,$\boldsymbol{\rho}_{\mathrm{m}}$,$\boldsymbol{\eta}_{\mathrm{down}}$和 $\boldsymbol{\eta}_{\mathrm{up}}$.

首先,依据负载的互相关系数 $\boldsymbol{\rho}_{\mathrm{m}}$ 把 VM 样本总体分为规律性和不规律性 2 类(如图 1 所示). 不规律性负载体现在负载波动规律性差,$\boldsymbol{\rho}_{\mathrm{m}}$ 中相关系数 ρ 值偏小,一般小于 0.3. 从而,不容易判断下一时间段的负载. 互相关系数 $\boldsymbol{\rho}_{\mathrm{m}}$ 体现在不同天的负载相关性.

其次,依据负载 $\boldsymbol{\eta}$(CPU($\boldsymbol{\beta}$)或者内存($\boldsymbol{\gamma}$)等)的累积均值 $\boldsymbol{\varepsilon}$,把规律性负载和不规律性负载分别分为 2 类,如图 1 中 K-means 聚类分析 $K=4$ 的情形.

图 3 即是使用负载特征 $\boldsymbol{\varepsilon}$ 和 $\boldsymbol{\rho}_{\mathrm{m}}$ 进行分层 K-means 聚类分析 $K=4$ 的聚类分析结果. 其中,选取 100 个虚拟机的 14 天的 CPU 使用率负载采样样本作为负载,累积负载均值是 14 天中每天 CPU

负载均值求和得到的；匹配的 P-value 值表示求同一虚拟机负载在不同 14 天中对应的互相关系数后满足 P-value 小于 0.5 时，匹配 P-value 的对数. Cluster1 和 Cluster2 对应规律性负载的聚类，而 Cluster3 和 Cluster4 对应非规律性负载的聚类，聚类之间的区别是对应负载的高低.

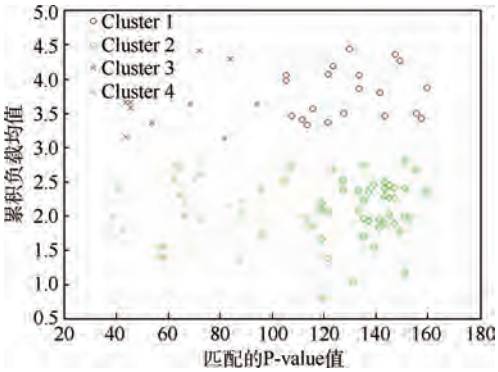


图 3　依据 $\boldsymbol{\varepsilon}$ 和 $\boldsymbol{\rho}_{\mathrm{m}}$ 的 $K=4$ 聚类分析结果

Fig. 3　Results of clustering analysis using $\boldsymbol{\varepsilon}$ and $\boldsymbol{\rho}_{\mathrm{m}}$ when $K=4$

最后针对规律性负载进行 $\boldsymbol{\varepsilon}$，$\boldsymbol{\eta}_{\mathrm{down}}$ 和 $\boldsymbol{\eta}_{\mathrm{up}}$ 负载特征进行更细粒度的 K-means 聚类分析（如图 1 中 $K=4$ 进一步聚类的情形所示）.

图 4 中所示为规律性负载进一步聚类成正常重载（Cluster4）、异常重载（Cluster3）、正常轻载（Cluster2）和异常轻载（Cluster1）4 个簇集. 正常与异常的区别体现在依据 CDF 累积分布特征 $\boldsymbol{\eta}_{\mathrm{up}}$ 和 $\boldsymbol{\eta}_{\mathrm{down}}$ 的分布. 因此，这个聚类分析结果为应用负载资源调度提供了基础.

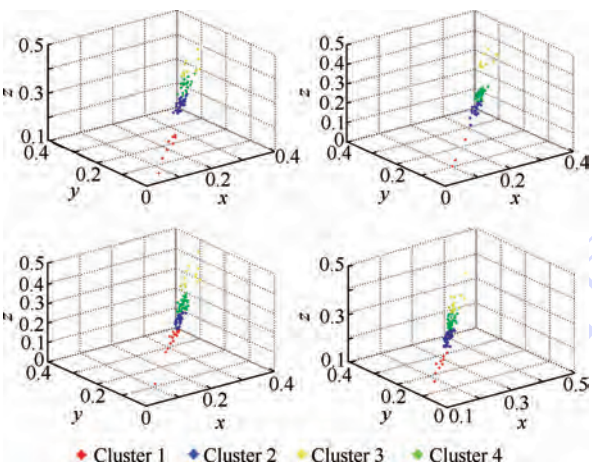


图 4　4 组负载特征聚类结果

Fig. 4　Clustering results of four groups of workload characteristics

3　资源调度算法

经过 IaaS 平台上的负载特征聚类分析，把负载消耗的资源量分析清楚，有的放矢地把虚拟机部署到合适的物理主机上，以达到降低违反 SLA 的可能性和节约能耗的目标. 目前，改进 BFD 算法[13]是应用最普遍的资源调度算法，而本文研究的基于负载特征聚类分析的资源调度算法将与文献[14]中的改进 BFD 算法进行比较. 下面给出本文研究的负载特征聚类资源调度算法伪代码和步骤描述.

```
1.  Input: hostList, vmList
2.  Output: allocation of Vms
3.  vmList. workloadClusteringVmPower( )
4.  foreach vm in vmList do
5.      foreach host in hostList do
6.          If host has enough resources for vm then
7.              vm. power←
        getPower( avg(U_workload) · TotalMips_VM / TotalMips_PM )
8.  foreach vm in vmList do
9.      minPower ← MAX
10.     allocatedHost ← NULL
11.     foreach host in hostList do
12.         If host has enough resources for vm then
13.             power ← getpower(host; vm)
14.             If power < minPower then
15.                 allocatedHost ← host
16.                 minPower ← power
17.     If allocatedHost, NULL then
18.         allocation. add(vm, allocatedHost)
19.  return allocation
```

负载特征聚类资源调度算法步骤第 1 和第 2 行为算法输入与输出，第 3 行为 VM 列表得到负载分析的聚类结果，第 4 ~ 7 行为每个 VM 尝试找到最合适的物理主机资源，第 8 ~ 16 行实现聚类的虚拟机分配到最优能耗的物理主机上. 由于聚类时考虑到使用率和阈值的关系（具体参见第 2 节聚类分析内容），所以保证了降低违反 SLA 的可能性.

4　实验及结果

4.1　实验平台

本实验平台选用 CloudSim3. 03 工具包作为实验模拟平台，它支持按需资源配置，易于扩展，并广泛为研究者认可[15]. 首先，创建一个数据中心，包括 800 个异构物理节点，其中，惠普 ProLiant 服务器 ml110 G4，惠普 ProLiant ml110 G5 服务器，惠普 ProLiant dl120 G5 服务器，惠普 ProLiant

dl180 G5 服务器各占总数的 1/4. 服务器的功耗数据由 SPEC 官方给出[16]. 其次,每个服务器运行在 1 GB/s 的网络带宽下,虚拟机的特点与类型为亚马逊 EC2 实例类型:高 CPU 实例(2 500 MIPS,0.85 GB);大内存实例(2 000 MIPS,3.75 GB);小实例(1 000 MIPS, 1.7 GB);微小实例(500 MIPS, 613 MB). 本实验负载特征样本采取在 Hadoop 集群上应用 Map-Reduce 并行计算的方法,处理近 20 亿条记录,相比传统数据库的效率提高 30 倍.

4.2 结果与分析

实验统一采用 MAD-MMT 迁移策略[17],对聚类调度算法和改进 BFD 算法进行比较. 实验场景选择第 2 节中 2 周(14 天)负载数据之外的 15 天 1 000 个虚拟机的负载数据集,进行参数为 SLA 违反率、能耗和虚拟机迁移数比较实验.

图 5 所示,15 天 1 000 个 VM 负载数据集在本实验场景下运行的每天违反 SLA 的百分率. 其中,聚类调度算法与改进 BFD 算法相比较在降低 SLA 违反率方面,平均有 38.6% 的提高,也就是说用户可以避免超过 1/3 性能下降的风险.

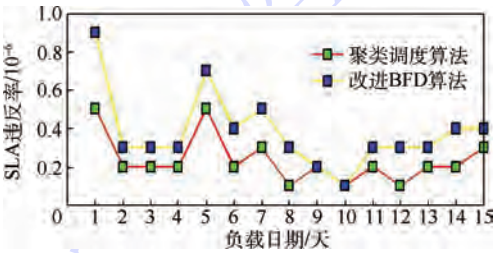


图 5 SLA 违反率实验结果

Fig. 5 Experimental results of SLA violation rate

经过 SLA 违反率参数的比较,可以看到聚类调度算法在用户 SLA 方面的优势. 而在节约能耗方面,通过同场景实验得到图 6 所示的结果. 结果显示,聚类调度算法比改进 BFD 算法在每一天都有明显的能耗节省,平均每天节省 13.72%.

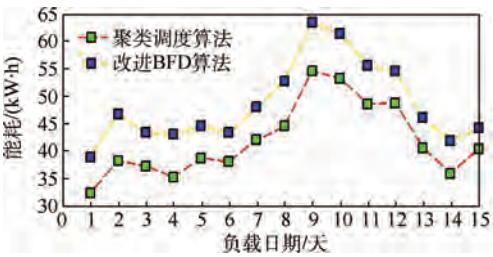


图 6 物理主机能耗实验结果

Fig. 6 Experimental results of energy consumption of physical hosts

VM 的迁移会造成额外的能耗,所以降低 VM 的迁移次数也是节能的关键因素之一. 通过同场景实验,2 种算法的 VM 迁移次数如图 7 所示. 通过比较可以知道聚类调度算法的明显优势.

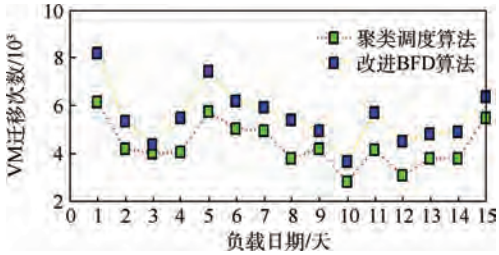


图 7 虚拟机迁移次数实验结果

Fig. 7 Experimental results of migration times of virtual machines

通过以上对聚类调度算法与改进 BFD 算法的比较实验分析,本文提出的资源调度算法在保证性能的前提下,能够大幅度减少虚拟机迁移次数,减少了因虚拟机过度迁移产生的额外能耗,因此,可以判断聚类调度算法更为节能,在降低 SLA 违反率方面更优. 这项研究可以与作者联系,获得完整的算法和数据下载,为在研究中实验可以复现.

5 结 论

本文的贡献是依据大规模云环境负载数据集的特点,对云环境负载进行特征提取并进行聚类分析,把最优的资源给最亟需的资源需求方. 依据这个思路提出基于负载特征聚类的资源调度算法,对 IaaS 云平台资源进行合理调度,达到降低 SLA 违反率和节能的目的. 经实验与改进 BFD 算法比较验证表明:

1) 在降低 SLA 违反率方面,平均有 38.6% 的提高,即用户可以避免超过 1/3 性能下降的风险.

2) 在降低能耗方面,平均每天节省 13.72% 的能耗.

3) 在减少虚拟机迁移次数方面,每天减少 1 000 ~ 2 000 次之间,大大减少了 IaaS 云平台额外的开销.

未来研究将继续进行负载特征提取和预测的研究,解决 IaaS 云平台上的能耗问题是最为重要的努力方向.

参考文献 (References)

[1] Wang X L, Liu Z H. An energy-aware VMs placement algorithm in cloud computing environment[C]//Intelligent System Design and Engineering Application(ISDEA), 2012 Second International Conference on. Piscataway, NJ: IEEE, 2012: 627-630.

[2] Yahoo. M45 supercomputing project[EB/OL]. Pittsburgh: Car-

negie Mellon University, 2009 [2014-07-04]. http://research.yahoo.com/node/1884.

[3] Google. Google Cluster data V2 [EB/OL]. California: Google-code, 2011 (2014-11-17) [2014-07-04]. http://code.google.com/p/googleclusterdata/wiki/ClusterData2011.

[4] Brown R. Report to congress on server and data center energy efficiency: public law 109-431 [J]. Lawrence Berkeley National Laboratory, 2008: 1-137.

[5] Anil K J. Data clustering: 50 years beyond K-means [J]. Pattern Recognition Letters, 2010, 31(8): 651-666.

[6] Zhang W, Yang H I, Jiang H Y, et al. Automatic data clustering analysis of arbitrary shape with K-means and enhanced ant-based template mechanism [C]//36th IEEE Annual International Computer Software and Applications Conference. Washington, D. C.: IEEE Computer Society, 2012: 452-455.

[7] 孙吉贵,刘杰,赵连宇. 聚类算法研究[J]. 软件学报,2008, 19(1): 48-61.
Sun J G, Liu J, Zhao L Y. Research on clustering algorithm [J]. Journal of Software, 2008, 19(1): 48-61 (in Chinese).

[8] Mahajan M, Nimborkor P, Varadarajan K. The planar K-means problem is NP-hard [J]. Lecture Notes in Computer Science, 2009, 5431: 274-285.

[9] 陈宁,陈安,周龙骧. 基于密度的增量式网格聚类算法[J]. 软件学报,2002, 13(1): 1-7.
Chen N, Chen A, Zhou L X. Incremental clustering algorithm based on density [J]. Journal of Software, 2002, 13(1): 1-7 (in Chinese).

[10] Gupta H, Srivastava R. k-means based document clustering with automatic "k" selection and cluster refinement [J]. International Journal of Computer Science and Mobile Applications, 2014, 2(5): 7-13.

[11] 李凯,李昆仑,崔丽娟. 模型聚类及在集成学习中的应用研究[J]. 计算机研究与发展,2007, 44(z2): 203-207.
Li K, Li K L, Cui L J. Study of model clustering and its application to ensemble learning [J]. Journal of Computer Research and Development, 2007, 44(z2): 203-207 (in Chinese).

[12] Polczynski M, Polczynski M. Using the k-means clustering algorithm to classify features for choropleth maps [J]. Cartographica: The International Journal for Geographic Information and Geovisualization, 2014, 49(1): 69-75.

[13] Coffman Jr E G, Csirik J, Galambos G, et al. Bin packing approximation algorithms: survey and classification [M]. Handbook of Combinatorial Optimization. New York: Springer, 2013: 455-531.

[14] Hong S, Zhang D, Lau H C, et al. A hybrid heuristic algorithm for the 2D variable-sized bin packing problem [J]. European Journal of Operational Research, 2014, 238(1): 95-103.

[15] Wang L, Lan Y Q, Xia Q X. Using CloudSim to model and simulate cloud computing environment [C]//Computational Intelligence and Security (CIS), 2013 9th International Conference on. Emeishan: IEEE, 2013: 323-328.

[16] SPEC. SPECpower_ssj2008 [EB/OL]. Texas: Standard Performance Evaluation Corporation, 2008 (2013-3-21) [2014-07-04]. http://www.spec.org/power/docs/SPECpower-Device_List.html.

[17] Beloglazov A, Buyya R. Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers [J]. Concurrency and Computation: Practice and Experience, 2012, 24(13): 1397-1420.

Energy-saving resource scheduling algorithm based on workload characteristic clustering

XIA Qingxin, LAN Yuqing*, TANG Tian, XIAO Limin

(School of Computer Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: When infrastructure as a service (IaaS) providers offer high performance services for users, they must think about how to reduce the energy cost of the cloud platform without violating the service level agreement (SLA). A resource scheduling algorithm to ensure SLA was proposed based on clustering analysis of the load characteristic. Ultimately, the targets of reducing SLA violation rate and saving energy were realized. The resource scheduling algorithm was analyzed based on improved K-means clustering analysis and extraction of workload characteristic according to energy consumption. Physical resources were effectively allocated to ensure the requirement of energy saving of IaaS platform. Based on the extension of the CloudSim simulation platform, the algorithm proposed was compared with the optimized best fit decreasing (BFD) to show lower SLA violation rate and energy consumption.

Key words: infrastructure as a service (IaaS); workload; clustering; energy-saving; service level agreement (SLA)

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0603

力约束管材自由胀形试验研究与材料性能测试

程鹏志[1], 郎利辉*[1], 葛宇龙[1], 阮尚文[1], 王韬[2], 武海[2]
(1. 北京航空航天大学 机械工程及自动化学院, 北京 100191; 2. 上海汇众汽车制造有限公司, 上海 200122)

摘 要: 相较于单向拉伸试验,通过管材胀形试验(TBT,Tube Bulging Test)获得的材料性能参数能够更准确地反映材料在高压流体作用下的塑性成形性能,不同的管端边界将会严重影响管材胀形试验的测试结果.针对国际上现有试验方法和设备存在的不足,研制出了一套约束边界清晰、加载精确的管材自由胀形试验系统.在管材测试过程中,基于位移随动力主动加载的控制策略和比例伺服油缸,实现实时的轴向力、轴向位移和内压力的精确加载.端部约束的测试管材通过特殊设计的工装保证了其轴向自由滑动.实时内压力和胀形管材顶点处材料的壁厚和胀形高度信息通过超高压压力传感器、超声测厚仪和磁致伸缩位移传感器采集,进而基于 Swift 材料本构模型和采集到的数据拟合出材料应力应变曲线和材料性能参数.试验结果表明,管材两端侧推力与内压力对管材内腔端面的作用力和管材轴向自由对称收缩的平衡条件始终处于动态稳定中,试验设备能够准确获得实时胀形高度、顶点厚度、轴向收缩长度和内压力的信息,能够为材料性能测试和工艺设计提供可信的材料参数.

关 键 词: 充液成形; 胀形试验; 管端边界; 自由滑动; 轴向力; 内压力; 胀形高度; 材料参数
中图分类号: V261.2+8
文献标识码: A **文章编号**: 1001-5965(2015)04-0686-07

管材充液成形技术(tube hydroforming)是指管材在内部液压力和轴向推力作用下充满模具型腔并贴模,进而成形具有一定复杂型面的空心薄壁零件高压柔性成形工艺,也被称为内高压成形或液压成形工艺,鉴于其特别适用于成形整体、复杂、薄壁的空心零件,及其成形精度高、材料利用率高、生产成本低的特点,该技术在 21 世纪初得到了快速发展,并开始广泛应用于航空航天、汽车制造等产业[1-3].随着工艺应用的发展,研究人员发现由于高压均匀面力等因素的作用,原始板材单向拉伸试验获得的材料性能参数不能很好地适用于管材充液成形工艺分析,实际生产对材料准确度的要求也越来越高,目前管材充液成形性能测试方法和装置尚不够完善和统一[4-8].需要采用更为准确、可靠的胀形方法获得材料性能数据.He 等通过比较现有的解析模型,针对 AA6061 材料建立了基于不同边界条件的管材胀形统一解析模型,认为胀形轮廓为椭圆,并考虑了模具圆角的影响[7-8].Velasco 和 Zribi 等通过有限元模拟和试验研究了材料性能参数对管材胀形过程中材料分布和胀形轮廓的影响[9-10];Imaninejad 等采用挤出铝管材对充液胀形中不同的管端约束条件及对管

收稿日期: 2014-09-28; **录用日期**: 2014-11-19; **网络出版时间**: 2014-12-22 10:57
网络出版地址: www.cnki.net/kcms/detail/11.2625.V.20141222.1057.003.html
基金项目: 国家"973"计划资助项目(613152); 国家自然科学基金国际(地区)合作与交流资助项目(51111120088)
作者简介: 程鹏志(1985—),男,河北石家庄人,博士生,pengzhi80h@163.com
* **通讯作者**: 郎利辉(1970—),男,河南新乡人,教授,lang@buaa.edu.cn,主要研究方向为先进成形技术及装备.

引用格式: 程鹏志, 郎利辉, 葛宇龙, 等. 力约束管材自由胀形试验研究与材料性能测试[J]. 北京航空航天大学学报, 2015, 41(4): 686-692. Cheng P Z, Lang L H, Ge Y L, et al. Tube free bulging experiment with force-end and material properties testing[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(4): 686-692 (in Chinese).

北航学报 赠阅

材应变状态的影响进行了试验研究，并对试验中的摩擦力进行了定性分析和定量测量. 试验结果表明，不同的管端约束和摩擦状态对材料变形有非常大的影响，以至于可以得到成形极限图上大范围内的点，材料厚向异性指数与管材的轴向收缩量呈显著正相关[11-12].

现有的管材胀形试验方法通常采用模具将管材两端完全固定，边界条件清晰，但不能监控材料的轴向收缩量，因而无法得到反映材料厚向异性指数. 另外，管端固定的边界导致管材胀形时管材顶点处轴向应力分量很大，不能很好地反映材料径向性能，因而与材料单向拉伸试验数据相差很大. 当采用放开管端轴向自由度的试验方法时，往往由于管材的收缩导致侧推冲头与管材间的过盈密封状态不良，过度推料又容易对管材两端施加推力而影响试验结果，摩擦状态随着胀形压力和侧推冲头的作用而急剧变化，试验干扰因素非常多. 管材胀形实时厚度的测量问题一直没有解决，间接理论推导导致误差增大.

本文根据大量工程经验，针对现有 TBT(Tube Bulging Test)方法存在的不足之处，提出了一种基于轴向自由胀形的管材充液成形性能测试试验方法及装置，建立了与之对应的理论解析模型，以获得实时胀形高度、轴向收缩量和壁厚变化等数据，进而获得应力应变曲线和材料性能参数. 试验管材受力边界简单、清晰，摩擦干扰误差小，易于检测、采集和处理数据.

1　试验原理与系统设计

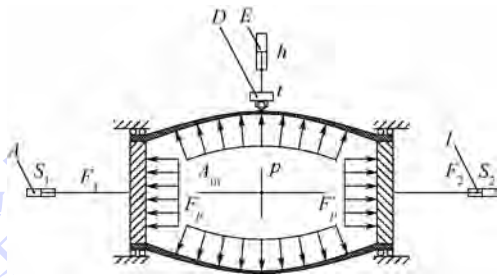
1.1　试验原理

力约束管材自由胀形试验方法原理如图 1 所示，将试验管材的两端完全刚性约束，仅放开其轴向移动的自由度，在管材内部施加均匀液压力 p，同时在管材左右两端分别通过伺服侧推油缸施加推力 F_1，F_2，并满足下式：

$$k_1F_1 = k_2F_2 = A_{in}p = F_p \tag{1}$$

$$S_1 = S_2 \tag{2}$$

式中，F_p 为内压力 p 对管材内腔端面 A_{in} 的作用力；A_{in} 为试验管材的管端内腔截面积；k_1，k_2 为加载系数；S_1，S_2 为左右两侧伺服油缸上安装的左侧位移传感器和右侧位移传感器的位移反馈值. 内压力 p 均匀上升直至管材破裂过程中保持前述关系不变，同时通过超声测厚仪及探头、胀形高度位移传感器实时监控管材胀形顶点的实时壁厚 t 和胀形高度 h. 当 $k_1 = k_2 = 1$ 时，侧推力 F_1，F_2 正好抵消内压力 p 对管材内腔端面 A_{in} 的作用力 F_p，通过获得的胀形过程数据 p，t，h 即可按照塑性理论推导出试验管材的轴向自由胀形应力应变曲线及材料参数. 此种工况下，试验管材仅受到内压力 p 对管腔内壁的法向作用和管端节点的径向约束，同时保持胀形最高点轴向不窜动，而不受任何他它外力、约束或摩擦的影响.



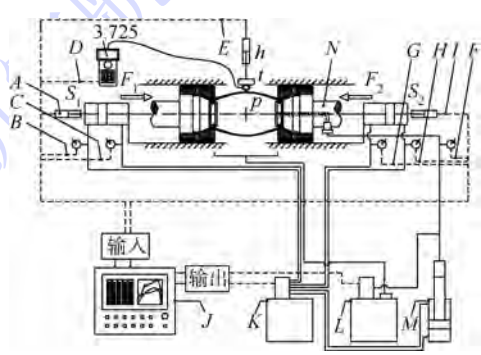
A—左侧位移传感器；D—超声测厚仪及探头；
E—胀形高度位移传感器；I—右侧位移传感器.

图 1　力约束管材自由胀形试验原理

Fig. 1　Principle diagram of TBT system

1.2　试验系统组成

如图 2 所示，试验系统分为控制系统、液压系统、水系统、增压器和试验工装 5 部分. 液压系统工作介质为液压油，水系统工作介质为乳化液，增压器为液压力转换和放大装置. 控制系统同时监测并记录左侧位移传感器、左侧油缸无杆腔压力传感器、左侧油缸有杆腔压力传感器、超声测厚仪及探头、胀形高度位移传感器、超高压压力传感器、右侧油缸有杆腔压力传感器、右侧油缸无杆腔压力传感器、右侧位移传感器 9 个传感器的实时数据.



A—左侧位移传感器；B—左侧油缸无杆腔压力传感器；
C—左侧油缸有杆腔压力传感器；D—超声测厚仪及探头；
E—胀形高度位移传感器；F—超高压压力传感器；
G—右侧油缸有杆腔压力传感器；
H—右侧油缸无杆腔压力传感器；I—右侧位移传感器；
J—控制系统；K—液压系统；
L—水系统；M—增压器；N—试验工装.

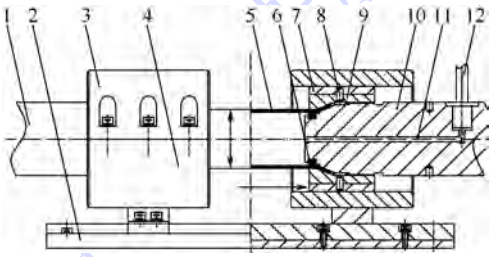
图 2　试验系统组成

Fig. 2　Composition of TBT system

2 关键技术实现与验证

2.1 管端约束、高压密封与导向结构

如图 3 所示,试验管材两端的约束和高压密封由左推杆、试验管材、对开螺母卡套部件、右推杆、超高压水管、高压密封圈 6 部分组成. 对开螺母卡套部件包括卡套、内螺纹圆柱销、对开螺母. 对开螺母上下瓣和卡套通过内螺纹圆柱销连接为一整体,并可反复拆装. 左、右推杆与对开螺母卡套部件采用螺纹连接,并可反复拆卸,试验管材经过扩口的两端被左、右推杆和对开螺母卡套部件夹紧在中间,通过螺纹实现预紧. 超高压水管通过螺纹连接在右推杆上,并可通过右推杆上的高压水通道向试验管材内注入高压液体. 卡套可以在左右两侧上压半圆形导套和下支撑半圆形导套组成的圆形导向结构内沿轴向自由窜动,配合界面采用大间隙配合,并涂油润滑,左右两侧的下支撑半圆形导套通过螺栓固定在底板上,并通过导向键保证同轴度,从而保证了管材沿轴向的自由移动.



1—左推杆;2—底板;3—上压半圆形导套;4—下支撑半圆形导套;5—试验管材;6—高压密封圈;7—卡套;8—内螺纹圆柱销;9—对开螺母;10—右推杆;11—高压水通道;12—超高压水管.

图 3 试验工装

Fig. 3 Tool set

2.2 侧推位移与力的比例伺服控制策略

对于比例伺服控制实现式(1)要求的力平衡条件仅需要对两侧伺服油缸进行力控制,当管材收缩时侧推油缸跟随运动. 但通过试验发现管材因胀形而收缩过程中两侧伺服油缸的位移是不稳定的,由于系统的不完全对称会导致管材偏移. 为了便于超声测厚仪稳定采集实时厚度数据,就需要对侧推油缸的位移进行控制,以确保管材顶点始终处于中间位置. 经过反复试验发现,精确控制左侧伺服油缸压力和右侧伺服油缸位移,左侧伺服油缸位移传感器采集的位移数据作为右侧伺服油缸位移控制数据,这种控制策略能够同时保证侧推力和侧推位移的对称,并且侧推油缸的位移量体现的是管材胀形的自由收缩量.

如图 4 所示,试验采集的数据表明了上述控制策略的有效性,在管材胀形至破裂的过程中,式(1)和式(2)所描述的试验平衡状态始终处于动态稳定中.

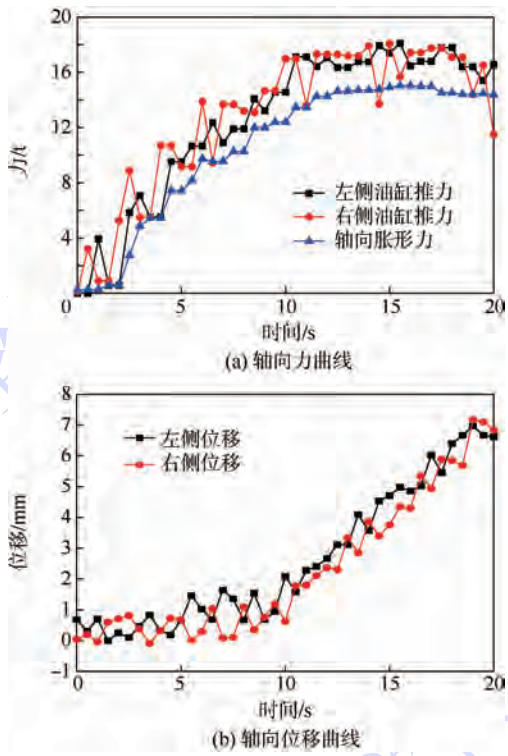


图 4 轴向力与位移曲线

Fig. 4 Force and displacement curve

2.3 关键技术特征

根据管端约束条件不同,传统的管材胀形试验方法分为管端固定、管端力加载和管端自由胀形[11]. 相较于现有试验方法,本文所描述的管材自由胀形试验方法具有如下特征:①在胀形试验之前,管材两端通过翻边模具进行扩口,对开螺母卡套部件将管材的翻边特征完全约束,因而阻滞了管材夹持段的材料向胀形区的流动,保证了管材变形区材料长度的一致性. ②由于管材两端的翻边特征完全被对开螺母卡套部件夹持,现有试验方法中管材与夹持模具之间的相对运动边界被转换到了对开螺母卡套部件与导套之间,原有的剧烈摩擦边界被间隙滑动配合所代替,因而在施加大的轴向力之前管材可以在导套内自由滑动. 在胀形阶段,夹持段管材的变形很小,对开螺母卡套部件与导套之间的滑动间隙仍旧稳定,滑动摩擦力很小,因而管材可以自由收缩. ③除了轴向移动的自由度,管材两端节点的其他自由度被完全约束,加上良好的滑动配合,伺服油缸提供的轴向力加载精确地作用在管材两端. 由于翻边特征的特殊固定方式,轴向拉力也可以施加在管材上,这为研究管材在双拉状态

下的变形行为提供了非常好的试验条件.

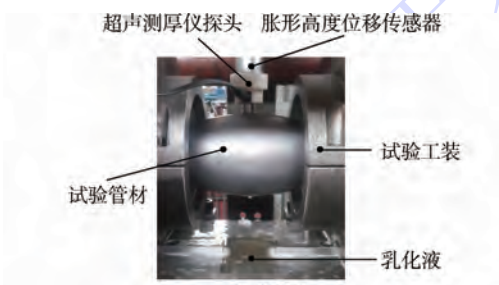
3　试验条件

3.1　试验材料

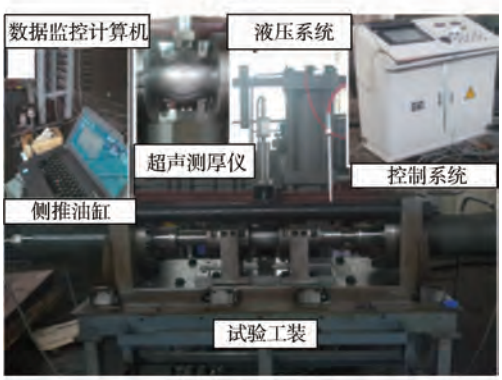
试验管材材料为 QSTE340 TM 高频焊管，长度 210 mm（含管材两端夹持段长度各 32.5 mm），外径 D_0 = 72.5 mm，长径比 L_0/D_0 = 2.0，变形区长度 L_0 = 145 mm，实测原始壁厚 t_0 = 3.665 mm. 单向拉伸试验测得的屈服强度 σ_s = 400 MPa，抗拉强度 σ_b = 450 MPa，延伸率 ≥25%. 试验前，在管材表面印上直径 2.5 mm 的圆形网格，用于测量管材破裂后的面内应变.

3.2　试验设备

试验设备为北京航空航天大学飞行器制造工程实验室为完成本试验自主研发的 FTBT 试验机，如图 5 所示. 设备最大胀形力 250 MPa，压力控制精度 0.5 MPa，侧推力 0 ~ 30 t，位移控制精度0.2 mm.



(a) 胀形管材



(b) 胀形试验机系统

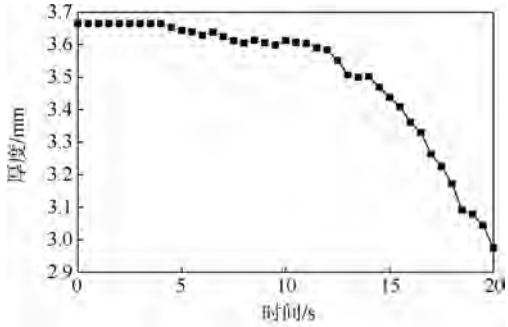
图 5　FTBT 管材自由胀形试验机

Fig. 5　Free-end tube bulging test machine

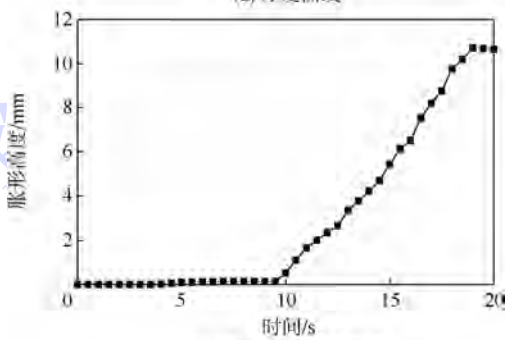
4　试验数据采集与分析

4.1　数据采集

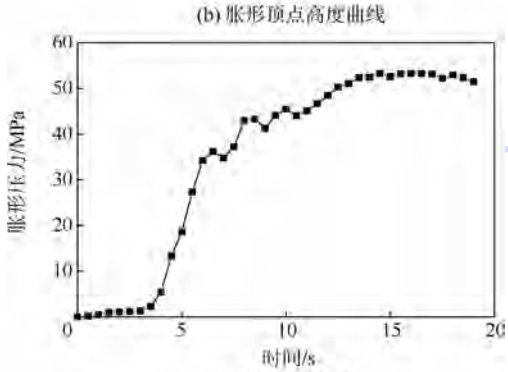
试验采用 PD-T1 精密型超声波测厚仪测量胀形零件顶点厚度 t，Novotechnik TMI 系列位移传感器测得胀形高度 h，及左右油缸的位移 S_1，S_2，KAVLICO PTE5000 系列压力传感器测得胀形压力 p. 控制系统每秒钟输出 2 组数据，增压时间 20 s，系统输出的顶点壁厚、胀形高度和压力曲线如图 6 所示.



(a) 厚度曲线



(b) 胀形顶点高度曲线



(c) 内压力曲线

图 6　试验数据采集曲线

Fig. 6　Experimental data

4.2　理论解析模型

如图 7 所示，力约束管材自由胀形试验中，在式(1)和式(2)满足的情况下，管材仅受到内压力 p 的作用，轴向推力 F_1，F_2 和内压力 p 对管材内腔端面 A_{in} 的作用力 F_p 相互抵消.

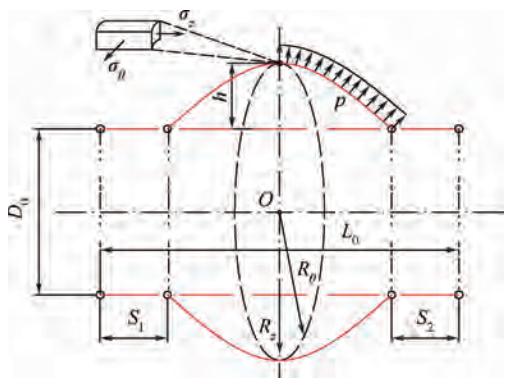


图 7　力约束管材自由胀形受力平衡示意图

Fig. 7　Force equilibrium diagram of free-bulging

北航学报 赠阅

根据前述边界条件和几何关系,列出下列平衡方程[12-13],可以得到胀形管材顶点的应力应变状态. 同时需要做出如下假设:

1) 薄壁管成形,忽略管材厚向应力,但考虑内压力 p 的作用;

2) 管材胀形过程中,轴向剖面轮廓曲线始终为余弦曲线;

3) 塑性变形过程中,材料体积不变;

4) 忽略材料各向异性.

首先分析管材顶点的应力状态,在塑性变形阶段,管材轴向仅受到内压力 p 在管材轴向轮廓的投影,轴向应力可以通过下式得出:

$$\sigma_z = \frac{p\pi[(D_0/2 + h)^2 - (D_0/2)^2]}{2\pi(D_0/2 + h)t} = \frac{p(R_\theta^2 - R_0^2)}{2(R_0 + h)t} \tag{3}$$

式中,R_0 为管材原始半径,当管材长径比 $L_0/D_0 = 2.0$ 时,端头效应的影响需要考虑在内,根据薄膜理论及静力平衡条件,此时有

$$\frac{\sigma_\theta}{R_\theta} + \frac{\sigma_z}{R_z} = \frac{p}{t} \tag{4}$$

式中,σ_θ 为顶点的环向应力分量;σ_z 为顶点的轴向应力分量;R_θ 为顶点的环向曲率半径;R_z 为顶点的轴向曲率半径;t 为顶点的实时壁厚.

假设管材轴向剖面轮廓曲线为余弦曲线:

$$y = a\cos(bx) + R_0 \tag{5}$$

根据几何关系可得:

$$R_z = \frac{1}{ab^2} \tag{6}$$

式中,有 $a = h, b = \pi/l$,则

$$R_z = \frac{l^2}{h\pi^2} = \frac{(L_0 - 2S_1)^2}{h\pi^2} \tag{7}$$

管材两端自由收缩量 $S_1 = S_2$ 通过图 4(b)可得,将式(3)、式(7)代入式(4)中即可推导出环向应力 σ_θ.

管材顶点的真实厚向应变可以通过超声测厚仪采集的实时厚度 t 得出:

$$\varepsilon_t = \ln\frac{t}{t_0} \tag{8}$$

管材环向真实应变为

$$\varepsilon_\theta = \ln\frac{D_0/2 + h}{D_0/2} \tag{9}$$

根据体积不变关系可得轴向应变:

$$\varepsilon_z = -\varepsilon_\theta - \varepsilon_t \tag{10}$$

平面应力状态下,根据 Von Mises 屈服准则推导出等效应力:

$$\bar{\sigma} = \sqrt{\sigma_\theta^2 - \sigma_\theta\sigma_z + \sigma_z^2} \tag{11}$$

根据下式推导出等效应变:

$$\bar{\varepsilon} = \sqrt{\frac{4}{3}(\varepsilon_\theta^2 + \varepsilon_\theta\varepsilon_z + \varepsilon_z^2)} \tag{12}$$

4.3 数据处理与分析

将图 4(b)和图 6 采集的数据代入前述公式即可获得材料塑性阶段的应力分量(图 8(a))、应变分量(图 8(b)). 采用式(13)所述的 Swift 材料本构拟合材料塑性阶段的等效应力应变曲线如图 8(c)所示.

$$\bar{\sigma} = K(\varepsilon_0 + \bar{\varepsilon})^n \tag{13}$$

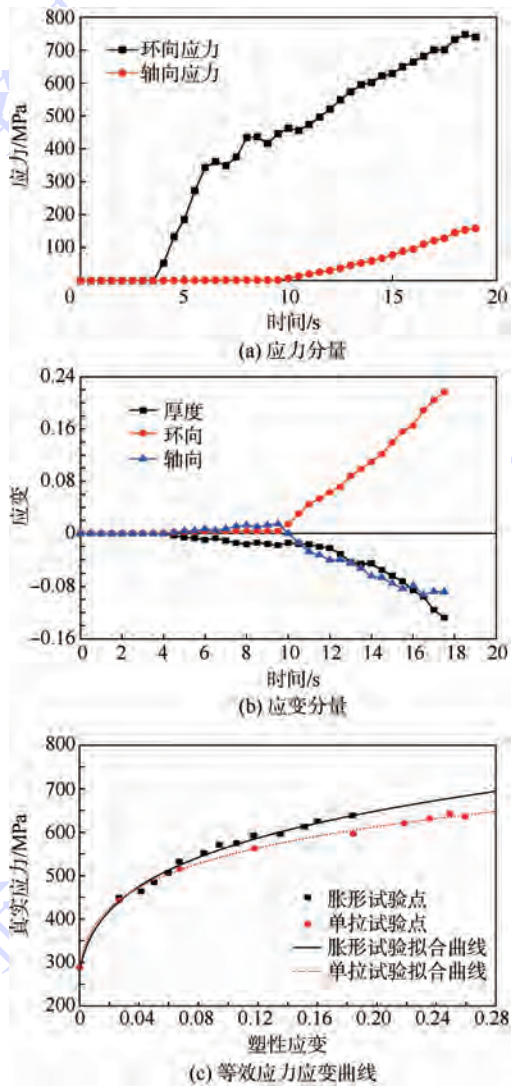


图 8 应力应变曲线及拟合

Fig. 8 Fit of strain stress curve

如图 8(a)所示的应力分量图,环向应力 σ_θ 远大于轴向应力 σ_z,受力状态接近于单向拉伸,轴向力的抵消起到了很好的效果,测试结果能够很好地反映管材的环向成形性能. 这个结果也是工艺设计所需要的,在实际工程应用中,管材充液成形零件的破裂主要发生在环向.

图 8(b)所示的应变分量图反映了管材胀形

过程中厚度逐渐减薄、环向逐渐拉长的过程，对应的轴向应变则经历了先压缩、后拉伸的过程，轴向力的抵消起到了明显的效果[14]. 在材料发生弹性变形和塑性变形的中前期阶段，胀形高度很低，内压力对管材轴向的作用很小，管材近似于仅受环向主应力的作用，引起明显的轴向和厚向收缩，因而该阶段的轴向应变为负值. 该阶段的应力应变状态非常近似于单向拉伸试验，从图 8(b)中也可以看出在该阶段轴向应变与厚向应变近似相等，与单向拉伸试验类似定义该阶段的厚向异性指数 r，则 r 近似为 1.0. 在胀形后期至破裂，由于胀形高度的增加，内压力对管材轴向的作用逐渐显著，材料应力状态向着双向拉伸快速转变，此时轴向应变趋势出现了明显的扭转上翘趋势，但最终轴向拉应力的作用还是远小于环向拉应力，因而轴向应变始终为负值.

进一步根据 Swift 材料本构拟合材料塑性阶段的等效应力-应变曲线如图 8(c)所示. 拟合所得的塑性强度系数 $K = 895.096$，硬化指数 $n = 0.202$，初始应变 $\varepsilon_0 = 0.003\,29$. 与材料单向拉伸试验获得的数据相比，试验数据结果基本符合材料在高压流体作用下表现出的特性[15-16]. 进一步的数据分析、数值模拟和试验工作需要继续完成，以便相互验证，为材料性能测试和工艺设计提供可靠手段和依据.

图 9 所示的为国产未退火材料和进口退火两种材料进行多次测试的管材胀形高度与内压力关系曲线，图 4 和图 6 中的数据均由进口退火材料测试获得. 通过图 9 可以看出试验系统具有较好的重复性，退火材料会有明显的屈服平台，而未退火材料可以获得更高的胀形高度.

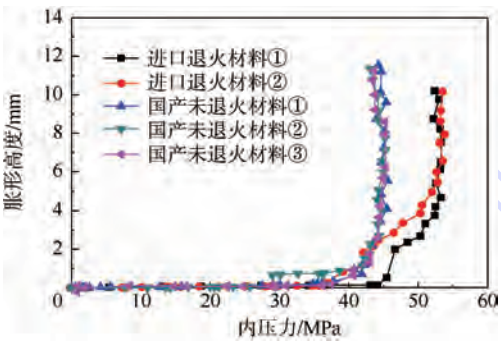


图 9　胀形高度与内压力关系曲线

Fig. 9　Bulging height vs internal pressure curve

5　结　论

1）相对于国际上现有的管材胀形试验机，本文所描述的试验方法和设备解决了 3 个固有问题：①将实现管端约束的对开螺母卡套部件和密封组件附着在管材两端，从而保证了管材胀形过程中卡套与半圆形导套的滑动间隙，即保证了管端的自由滑动，排除摩擦干扰的同时准确获得了管材胀形时的轴向自由收缩数据；②鉴于翻边特征的良好试验效果，比例伺服油缸可以对管材两端施加精确的推力(或拉力)，并能有效阻止管端被夹持材料向变形区的流动；③采用超声测厚仪实时采集管材顶点的厚度变化信息，直接准确测量推导出厚向应变. 试验数据的采集具有较好的可重复性精度.

2）在管材胀形过程中，环向应力 σ_θ 远大于轴向应力 σ_z，受力状态接近于单向拉伸，轴向力的抵消起到了很好的效果，测试结果能够很好地反映管材的环向成形性能.

3）胀形过程中发生弹性变形和塑性变形的中前期阶段，管材近似于仅受环向主应力的作用，引起明显的轴向和厚向收缩，该阶段的轴向应变为负值，轴向应变与厚向应变近似相等，对应单向拉伸试验可以看出厚向异性指数 r 近似为 1.0. 在胀形后期至破裂，材料应力状态向着双向拉伸快速转变，轴向应变趋势出现了明显的扭转上翘趋势，但轴向应变始终为负值.

参考文献（References）

[1] Yuan S J, Liu G. 3.04-Tube hydroforming (internal high-pressure forming)[J]. Comprehensive Materials Processing, 2014, 3: 55-80.

[2] Ahmetoglu M, Altan T. Tube hydroforming: state-of-the-art and future trends[J]. Journal of Materials Processing Technology, 2000, 98(1): 25-33.

[3] Lang L H, Wang Z R, Kang D C, et al. Hydroforming highlights: sheet hydroforming and tube hydroforming[J]. Journal of Materials Processing Technology, 2004, 151(1): 165-177.

[4] Ouirane A H B, Boudeau N, Velasco R, et al. Error evaluation on experimental stress-strain curve obtained from tube bulging test [J]. Thin-Walled Structures, 2011, 49(10): 1217-1224.

[5] Thiruvarudchelvan S, Seet G L, Ang H E. Computer-monitored hydraulic bulging of tubes[J]. Journal of Materials Processing Technology, 1996, 57(1-2): 182-188.

[6] Yang L F, Guo C. A simple experimental tooling with internal pressure source used for evaluation of material formability in tube hydroforming[J]. Journal of Materials Processing Technology, 2006, 180(1-3): 310-317.

[7] He Z B, Yuan S J, Lin Y L. Analytical model for tube hydro-bulging test, part I: models for stress components and bulging zone profile[J]. International Journal of Mechanical Sciences, 2014, 87: 297-306.

[8] He Z B, Yuan S J, Lin Y L. Analytical model for tube hydro-bulging test, part II: linear model for pole thickness and its applica-

tion[J]. International Journal of Mechanical Sciences,2014,87:307-315.

[9] Velasco R, Boudeau N. Tube bulging test: theoretical analysis and numerical validation [J]. Journal of Materials Processing Technology,2008,205(1-3):51-59.

[10] Zribi T, Khalfallah A, BelHadjSalah H. Experimental characterization and inverse constitutive parameters identification of tubular materials for tube hydroforming process[J]. Materials and Design,2013,49:866-877.

[11] Imaninejad M, Subhash G, Loukus A. Influence of end-conditions during tube hydroforming of aluminum extrusions[J]. International Journal of Mechanical Sciences,2004,46(8):1195-1212.

[12] Imaninejad M, Subhash G, Loukus A. Experimental and numerical investigation of free-bulge formation during hydroforming of aluminum extrusions[J]. Journal of Materials Processing Technology,2004,147(2):247-254.

[13] Boudeau N, Malecot P. A simplified analytical model for post-processing experimental results from tube bulging test: theory, experimentations, simulations[J]. International Journal of Mechanical Sciences,2012,65(1):1-11.

[14] Varma N S P, Narasimhan R. A numerical study of the effect of loading conditions on tubular hydroforming[J]. Journal of Materials Processing Technology,2008,196(1-3):174-183.

[15] Liu B S, Lang L H, Zeng Y S, et al. Forming characteristic of sheet hydroforming under the influence of through-thickness normal stress[J]. Journal of Materials Processing Technology, 2012,212(9):1875-1884.

[16] Strano M, Altan T. An inverse energy approach to determine the flow stress of tubular materials for hydroforming applications [J]. Journal of Materials Processing Technology, 2004, 146(1):92-96.

Tube free bulging experiment with force-end and material properties testing

CHENG Pengzhi[1], LANG Lihui[*1], GE Yulong[1], RUAN Shangwen[1], WANG Tao[2], WU Hai[2]

(1. School of Mechanical Engineering and Automation, Beijing University of Aeronautics and Astronautics, Beijing 100191, China;
2. Shanghai Huizhong Automotive Manufacturing Co., Ltd, Shanghai 200122, China)

Abstract: Compared with uniaxial tension, the material property parameters obtained by tube bulging test (TBT) can accurately reflect the plastic forming properties of materials under high-pressure fluid condition. Different end-conditions will seriously affect the experiment results of TBT. According to the defects of existed experimental method and equipment internationally, a special TBT system, with accurate boundary conditions and loading units, was successfully designed and developed. The axial force, axial displacement, and the internal pressure are the key points of the system which could be accurately controlled in real-time bycontrol strategy of displacement under dynamic active loading state and proportional servo valves. The free sliding of fixed tube ends was implemented by a special designed fixture. During the experimental process, the real-time thickness and bulging height in the pole, and the real-time internal pressure were monitored by the ultrasonic thickness gauge, the magnetostrictive displacement sensor and the ultra-high pressure sensor. The stress-strain curve and material properties were derived by the Swift material constitutive model and the monitor data. The experimental data demonstrate that the two load balance conditions, which are the balance of two axial forces with internal pressure and the balance of two axial contractive displacements, are always satisfied during all the test process. The test data processing results show that the monitor data have good repeatability and can be used to get the material properties.

Key words: tube hydroforming; bulging test; end-condition; free sliding; axial force; internal pressure; bulging height; material properties

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2015.0072

氯盐对碱激发矿渣净浆强度影响试验

程寅，黄新*
（北京航空航天大学 交通科学与工程学院，北京 100191）

摘 要：通过对掺加氯化钠(NaCl)、氯化钙($CaCl_2$)的碱激发矿渣(KC)净浆的无侧限抗压强度试验，研究两种氯盐对KC净浆强度的影响规律，利用X射线衍射(XRD)、热重分析(TG-DTG)、液相离子测定、水化程度测定等测试手段，分析两种氯盐对KC净浆强度的作用机理. 结果表明，随着NaCl掺量增加，KC净浆抗压强度提高；随着$CaCl_2$掺量增加，KC净浆抗压强度基本不变. 掺加NaCl与$CaCl_2$的KC净浆中均生成含氯水化物水化氯铝酸钙(Fs，Friedel's salt)，但Fs的生成对KC净浆强度无影响作用. 掺加NaCl的KC净浆中氢氧化钠(NaOH)的生成是导致其强度增强的主要原因，混合物中NaOH的存在提高了KC净浆的液相碱度，促进矿渣水化的进一步发生，进而生成更多的水化硅酸钙(CSH)使KC净浆强度得到提高.

关 键 词：氯化钠；氯化钙；碱激发矿渣；作用机理；水化氯铝酸钙；氢氧化钠
中图分类号：TU528.2
文献标识码：A **文章编号**：1001-5965(2015)04-0693-08

许多近海沿岸的混凝土构筑物，由于长期受海水浸渍和海潮冲刷，腐蚀严重. 海水中大量存在的氯盐通过渗透扩散对混凝土中钢筋有强腐蚀性[1]，同时其对混凝土本身也有强侵蚀性[2-3]，这都严重影响混凝土结构的耐久性和安全性，影响混凝土在这类地区的使用. 因此亟需寻求一种耐氯盐侵蚀的材料并可应用于高氯盐环境下的海工混凝土中. 矿渣作为可提高混凝土抗盐蚀性能的材料可用于高氯盐环境[4]. 首先矿渣作为矿物掺合料可有效提高混凝土物理抗渗性能，黎鹏平和余红发等[5-6]的研究显示以矿渣作为掺合料可改善混凝土孔隙率和平均孔径，有效提高混凝土抗氯离子渗透的能力，防止氯离子渗透对混凝土中钢筋的锈蚀. 同时矿渣作为胶凝材料可有效结合氯离子实现抗氯盐侵蚀[7-8]，Deja 和 Shi 等[9-10]的研究表明碱激发矿渣具有的抗氯盐侵蚀能力与激发剂种类及激发反应过程有关；罗睿等[11-12]通过对碱激发矿渣净浆结合外渗氯离子能力的研究得到矿渣通过化学结合游离氯离子形式实现对氯离子侵蚀的抵抗；孙国文等[13]通过矿渣-水泥体系对氯离子结合能力和结合机理的研究表明矿渣通过化学结合氯离子形式生成一种含氯水化物水化氯铝酸钙(Fs，Friedel's salt)；余红发等[14]通过等温吸附法研究得到矿渣化学结合氯离子生成Fs的能力较强；Talero 和 Zhu 等[15-16]通过矿渣净浆液相离子浓度变化研究提出Fs在矿渣中可能的生成机理.

上述研究多为单就矿渣结合氯离子能力或结合机理的研究，对矿渣在氯盐环境下的强度变化规律、水化作用机理及两者的内在联系未有系统的研究，亦未见有其他的相关报道. 海水环境中氯

收稿日期：2015-02-03；**录用日期**：2015-03-01；**网络出版时间**：2015-03-03 10:41
网络出版地址：www.cnki.net/kcms/detail/11.2625.V.20150303.1041.003.html
基金项目：国家自然科学基金资助项目(50278007)
作者简介：程寅(1984—)，男，浙江衢州人，博士生，306876662@qq.com
*通讯作者：黄新(1956—)，男，北京人，教授，hx2000@buaa.edu.cn，主要研究方向为岩土工程化学与软土固化技术、软土地基处理原理预技术、水泥基复合材料.

引用格式：程寅，黄新. 氯盐对碱激发矿渣净浆强度影响试验[J]. 北京航空航天大学学报，2015，41(4)：693-700.
Cheng Y，Huang X. Experiment of chloride effect on strength of alkali-activated slag paste[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：693-700 (in Chinese).

盐主要包括 NaCl 和 $CaCl_2$ 等,通过矿渣净浆浆体内掺加 NaCl 和 $CaCl_2$ 的方法研究碱激发矿渣净浆在氯盐环境下的强度变化规律,同时借助 X 射线衍射(XRD,X-Ray Diffraction)、热重分析(TG-DTG, Thermal Gravity-Differential Thermal Gravity analysis)、液相离子测定、水化程度测定等测试手段,探求两种氯盐对碱激发矿渣强度的作用机理,可为实现碱激发矿渣在近海高氯盐环境下的应用提供可靠的理论依据.

1 试验材料与方法

1.1 试验材料

1) 首钢矿渣:首都钢铁集团磨细高炉矿渣,比表面积 400 m^2/kg,其化学组成如表 1 所示.

表 1 矿渣的化学组成

Table 1 Chemical composition of slag

化学组成	CaO	SiO_2	Al_2O_3	MgO	SO_3	TiO_2	Fe_2O_3	K_2O	MnO	Na_2O	P_2O_5	其他
质量分数/%	44.06	28.72	15.13	5.34	3.81	0.96	0.83	0.47	0.23	0.2	0.08	0.17

2) 分析纯试剂:广东西陇化工有限公司生产的氯化钠(NaCl),氯化钙($CaCl_2$),氢氧化钙($Ca(OH)_2$),硝酸银($AgNO_3$),铬酸钾(K_2CrO_4),盐酸(HCl,36%),硝酸(HNO_3,69%),硫氰酸钾(KSCN),水杨酸($C_7H_6O_3$),丙酮(C_3H_6O,99.5%),甲醇(CH_4O,99.9%)等试剂.

3) 拌合水:蒸馏水.

1.2 试验方案

净浆材料选用碱激发矿渣(以 $Ca(OH)_2$ 作为碱激发剂,矿渣与 $Ca(OH)_2$ 以质量比 9∶1 混合,记为 KC),依据海水主要氯盐种类与可能的浓度范围[17],分别掺入不同量的 NaCl 和 $CaCl_2$(以 Cl^- 掺入质量分数表示,且分别以 X 系列、Y 系列表示),以水灰比 1∶2 制作净浆试块,具体净浆试验方案见表 2、表 3. 标准养护至规定龄期,测定净浆试块的无侧限抗压强度. 同时对净浆的微观结构和水化产物进行 XRD、TG-DTG、液相离子浓度测试、水化程度测定等.

表 2 掺加 NaCl 的 KC 净浆 X 系列试验方案

Table 2 Test plan of X KC paste mixed with NaCl

X 系列	X0	X1	X2	X3	X4	X5	X6	X7	X8
Cl^- 质量分数/%	0	0.5	1.5	3	5	8	10	15	20

表 3 掺加 $CaCl_2$ 的 KC 净浆 Y 系列试验方案

Table 3 Test plan of Y KC paste mixed with $CaCl_2$

Y 系列	Y0	Y1	Y2	Y3	Y4
Cl^- 质量分数/%	0	0.5	1.5	3	5

1.3 试验方法

1) 净浆试块制备、养护及强度测定.

将净浆原料按配比分别称重放置于 NJ-160B 搅拌机的搅拌锅内,慢速搅拌 30 s;按水灰比掺入拌合水后慢速搅拌 90 s;将锅壁四周浆体刮入锅内后,再快速搅拌 60 s;将浆体分 3 层放入已涂抹脱模剂的20 mm × 20 mm × 20 mm 六联钢试模中,每层浆体由手工插捣密实,装满后再经振动台振捣密实;试样表面刮平后覆塑料薄膜,室温保湿养护 1d 后脱模,将脱模后净浆试块放入养护箱内养护(温度为 20 ± 2℃,湿度为≥95%),养护至规定龄期后使用加载速度为 1 mm/min 的路面材料强度试验仪测试 KC 净浆试块的无侧限抗压强度,每种配比试块的抗压强度取值为 3 个试块强度的均值.

2) XRD 测试.

测试采用岛津 LabX XRD-6000X 射线衍射仪进行测试,仪器参数:管压 40 kV,管流 40 mA,Cu Kα辐射,λ = 0.154 nm,扫描范围(2θ)0° ~ 60°,扫描速度 4(°)/min,步进0.02°. 试样制备:取测定了抗压强度后的净浆中心部分的碎块置于真空干燥箱内于 60℃烘干至恒重,研磨试样细度在200 ~ 400 目,干燥封存.

3) 净浆孔隙液液相离子浓度测试.

测试按照《水运工程混凝土试验规程》(JTJ270—1998)[18]和《水泥化学分析方法》(GB/T 176—2008)[19]规范测定. 试样采用测定了净浆抗压强度后的净浆中心碎块,将碎块粉碎研磨后过 0.63 mm 筛,然后置于真空干燥箱中 100℃烘 2 h,取出后放入干燥器冷却保存,以备使用.

① 总 Cl^- 含量(C)测定步骤:取适量磨细干燥后的净浆试样置于锥形瓶中,加入 100 mL 15% 的稀硝酸,塞紧瓶塞后浸泡 24 h,期间不定时摇动锥形瓶. 过滤后对浸出液进行离子浓度测定.

吸取体积为 V 的浸出液置于锥形瓶中,由滴定管加入过量体积的 V_2 硝酸银标准溶液(记 C_{AgNO_3})后,再用硫氰酸钾标准液(记 C_{KSCN})滴定,滴定至溶液出现红色且维持 10 s 不褪色为终点,记下硫氰酸钾标准液的体积消耗量 V_3. 按式(1)计算可得总 Cl^- 含量 C(mmol/L).

$$C = 100 \times (C_{AgNO_3}V_2 - C_{KSCN}V_3)/V \quad (1)$$

② 游离 Cl^- 含量(C_1)测定步骤:取适量磨细

干燥后的净浆试样置于锥形瓶中,加入 100 mL 蒸馏水,塞紧瓶塞后先振荡 1 ~ 2 min,之后连续浸泡 24 h. 对浸出液过滤后进行离子浓度测定.

吸取体积 V_1 浸出液置于锥形瓶中,加 0.5% 酚酞指示剂 1 ~ 2 滴,使溶液呈微红色,再用 5% 稀硫酸中和至无色后,加入浓度为 5% 铬酸钾指示剂,用硝酸银标准液(记 C_{AgNO_3})滴定至出现砖红色沉淀为止,记下硝酸银标准液体积消耗量 V_4. 按式(2)计算可得游离 Cl^- 含量 C_1(mmol/L).

$$C_1 = 100 \times C_{AgNO_3} V_4 / V_1 \tag{2}$$

③ 游离 OH^- 含量(P)测定步骤:浸出液制备方法与游离 Cl^- 相同,吸取体积 V_5 浸出液后置于锥形瓶中,加 1% 甲基橙指示剂 4 滴,用酸式滴定管滴入盐酸标准液(记 C_{HCl})至溶液由黄色突变到橙色为终点,记下盐酸标准液体积消耗量 V_6. 按式(3)计算可得游离 OH^- 含量 P(mmol/L).

$$P = 100 \times C_{HCl} V_6 / V_5 \tag{3}$$

4) 矿渣水化程度(K)测试.

采用 Kondo 和 Oshsawa 提出的水杨酸-甲醇-丙酮选择性溶解法[20]测试矿渣水化程度:取净浆中心部分碎块,粉碎研磨后过 0.08 mm 筛后置于 60℃ 真空干燥箱中烘至恒重,取 0.5 g 干燥后试样与 2.5 g 水杨酸,35 mL 丙酮及 15 mL 甲醇在锥形瓶中混合并密封保存 24 h,用滤纸过滤后,将滤纸同滤渣一并放入箱式电阻炉内 850℃ 灼烧 10 min,冷却后称量残余物质量 m,按式(4)计算得到矿渣水化程度 K.

$$K = (0.5 - m)/m \tag{4}$$

5) TG-DTG 测试.

测试采用耐驰 TG209F3 热重分析仪进行测试. 仪器参数:炉内采用高纯氮气作为保护气和吹扫气,采用液氮降温,仪器循环水温为 38℃,以 Al_2O_3 空坩埚作为参比样. 温度控制:本次热重实验温度范围为 50 ~ 200℃,升温速率控制在 10 ~ 40 K/min. 试样制备:将测定了抗压强度后的净浆中心部分碎块用过量无水乙醇于玛瑙研磨仪中湿磨成浆状,然后放入真空干燥箱 60℃ 真空干燥至恒重,干燥封存.

2　试验结果

2.1　净浆抗压强度试验结果

掺加 NaCl 和 $CaCl_2$ 的 X 与 Y 系列净浆试块在 7 d 和 28 d 龄期下的无侧限抗压强度随 Cl^- 掺入质量分数变化曲线如图 1 所示.

由图 1(a)可见,在 7 d 和 28 d 龄期下,掺加 NaCl 的 X 净浆强度随 Cl^- 掺量的增加而提高. 28 d龄期下 Cl^- 掺量为 0% ~ 5% 区间内强度增长速率较快,在 Cl^- 掺量大于 5% 后强度增长趋缓,Cl^- 掺量为 20% 的 X8 的抗压强度比无氯离子的 X0 的强度提高了约 50%. 由图 1(b)可见,在 7 d 和 28 d 龄期下,掺加 $CaCl_2$ 的 Y 系列净浆试块的抗压强度基本未随 Cl^- 掺量的增加而改变.

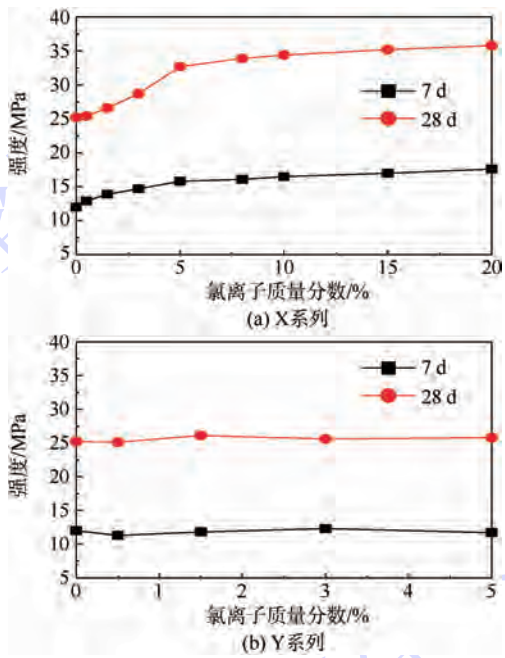


图 1　不同 Cl^- 掺入比下 X,Y 净浆抗压强度变化曲线

Fig. 1　Relationship between chloride ion content and compressive strength of X, Y paste

2.2　XRD 试验结果

KC 粉末,X 系列、Y 系列净浆在 7d 和 28d 龄期下的 XRD 图谱见图 2、图 3,图中 CH,Fs,CSH,NaCl 分别表示氢氧化钙($Ca(OH)_2$)、水化氯铝酸钙($3CaO \cdot Al_2O_3 \cdot CaCl_2 \cdot 10H_2O$)、水化硅酸钙凝胶(C-S-H)和氯化钠(NaCl)的主要衍射峰. X 中 Fs 的半定量分析见图 4.

由图 2(a)、图 3(a)可见,KC 水化后 $Ca(OH)_2$ 衍射峰消失.

由图 2、图 4 可见,掺加 NaCl 的 X 系列净浆中,出现了水化氯铝酸钙 Fs 的衍射峰,随着 NaCl 掺量升高,Fs 衍射峰高度提高(图 2(a)),衍射峰面积增强,生成量增加(图 4),同时 X 系列净浆中 Fs 衍射峰的增强伴随着 NaCl 衍射峰的显著减弱或消失(图 2(b)). 由图 3 可见,掺加 $CaCl_2$ 的 Y 系列净浆中亦出现了 Fs 的衍射峰,且随着 $CaCl_2$ 掺量的增加 Fs 衍射峰在增强(图 3(a)),同时 Fs 衍射峰的增强也伴随着 $CaCl_2$ 衍射峰的消失(图 3(b)).

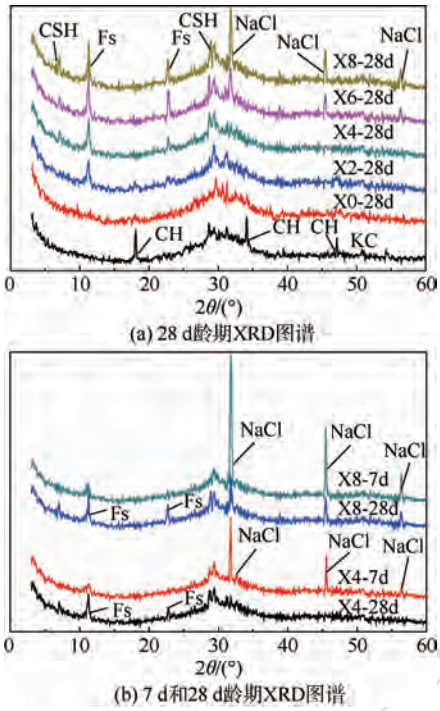


图 2 KC 与不同 Cl^- 掺量下 X 系列净浆的 XRD 图谱

Fig. 2 X-ray diffraction of KC and X paste with different chloride ion content

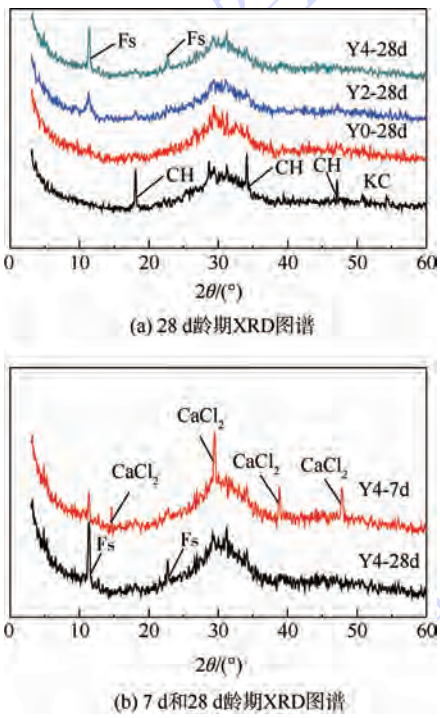


图 3 KC 与不同 Cl^- 掺量下 Y 系列净浆的 XRD 图谱

Fig. 3 X-ray diffraction of KC and Y paste with different chloride ion content

2.3 净浆孔隙液液相离子浓度测试结果

表 4 为测试得到的 28 d 龄期下 X 和 Y 系列净浆孔隙液液相中总 Cl^- 含量(C)、游离 Cl^- 量(C_1)、被结合 Cl^- 量(C_2)、游离 OH^- 量(P)及 pH 值,其中 C_1 为未被结合的游离态 Cl^- 含量,C_2 为被矿渣化学结合生成含氯水化物的化合态 Cl^- 含量,$C = C_1 + C_2$,pH 值 = 14 + lg[OH^-]. 净浆中 C_2 和 P 随 Cl^- 掺入比的关系变化曲线示于图 5、图6.

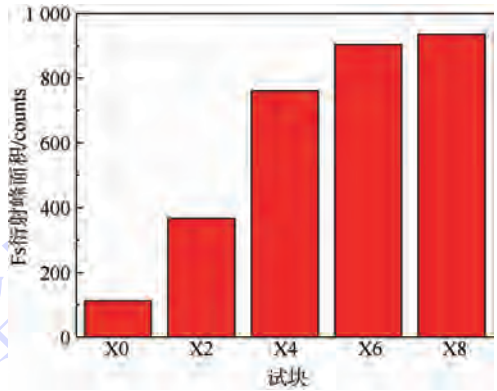


图 4 28 d 龄期下 X 系列净浆 XRD 图谱中 Fs 半定量分析图

Fig. 4 Semi-quantitative analysis of Fs in X-ray diffraction of X paste within curing time of 28 d

表 4 X,Y 系列净浆孔隙液液相中 Cl^-,OH^- 含量及 pH 值

Table 4 Cl^-, OH^- concentrations and pH value in the pore solutions obtained from X, Y paste

试块	C_1/(mmol·L^{-1})	C/(mmol·L^{-1})	C_2/(mmol·L^{-1})	P/(mmol·L^{-1})	pH 值
X0/Y0	—	—	—	87	12.94
X1	83	282	199	122	13.09
X2	412	845	433	191	13.28
X3	889	1 690	801	271	13.43
X4	1 613	2 817	1 204	395	13.59
X5	3 210	4 507	1 297	400	13.60
X6	4 327	5 634	1 307	405	13.61
X7	7 095	8 451	1 356	414	13.62
X8	9 896	11 268	1 372	458	13.66
Y1	53	271	218	84	12.94
Y2	329	880	551	92	12.96
Y3	660	1 626	966	81	12.90
Y4	1 503	2 804	1 301	82	12.91

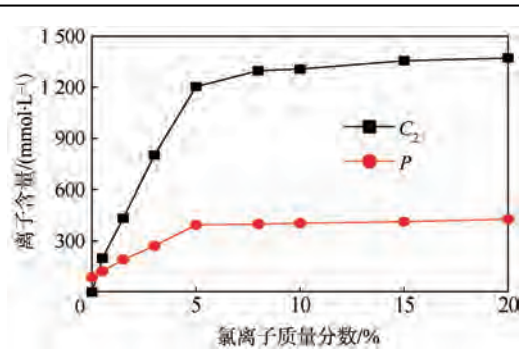


图 5 不同 Cl^- 掺量下 X 系列净浆中离子量变化曲线

Fig. 5 Relationship between chloride ion content and ion concentrations in X paste

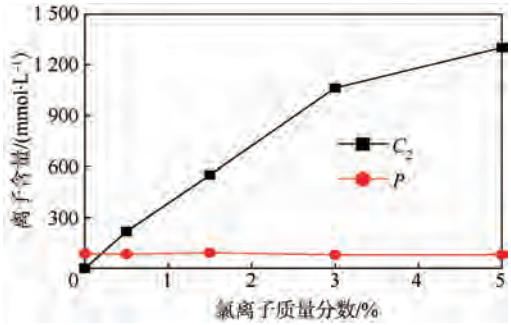


图 6　不同 Cl^- 掺量下 Y 系列净浆中离子量变化曲线

Fig. 6　Relationship between chloride ion content and ion concentrations in Y paste

由表 4 和图 5 可知,随 NaCl 掺入量增加,X 系列净浆孔隙液液相中 C_2 和 P 在增加,在 Cl^- 掺量为 0% ~5% 区间内 C_2 和 P 增长较快,在 Cl^- 掺量大于 5% 后 C_2 和 P 的增长明显趋缓,同时净浆孔隙液 pH 值升高;而由表 4 和图 6 可见,随 $CaCl_2$ 掺入量增加,Y 系列净浆孔隙液液相中 C_2 虽也增加,但 P 基本未变,孔隙液 pH 值也基本未变.

2.4　矿渣水化程度测试结果

对测定了 28 d 强度后的净浆 X0,X2,X4,X6,X8 试样进行矿渣水化程度测试,计算所得净浆的矿渣水化程度随 Cl^- 掺入比变化图如图 7 所示.

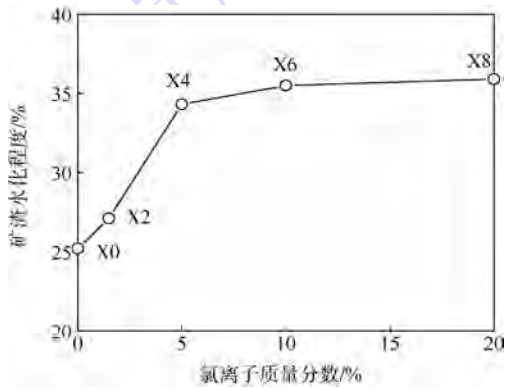


图 7　不同 Cl^- 掺量下 X 系列净浆的矿渣水化程度

Fig. 7　Hydration degree of X paste slag with different chloride ion content

由图 7 可见,随着 NaCl 掺量的增加,X 系列净浆中矿渣的水化程度在升高,无 Cl^- 的 X0 的矿渣水化度为 25.2%,而 Cl^- 掺量为 20% 的 X8 的矿渣水化度提高到 35.9%.

2.5　TG-DTG 测试结果

依据 DTG 测试曲线,通过仪器的 Proteus 分析软件可根据人工输入的温度范围、水化物的分子量以及水或气体在该水化物中所占的质量分数,自动计算出该水化物在测试样品中的质量分数[21-22].同时由文献[22-23]可知,CSH 结晶水失重基本在 50 ~200℃ 温度区间范围内完成,X 系列净浆中除 CSH 以外的其他水化物均不在此温度区间内.因此对 28 d 龄期下 X0,X2,X4,X6,X8 净浆进行 TG-DTG 测试,得到的在 50 ~200℃ 温度区间内 DTG 测试曲线见图 8 并计算得到各净浆的 CSH 所占质量分数.不同 Cl^- 掺量下 X 系列净浆的 CSH 含量变化曲线如图 9 所示.

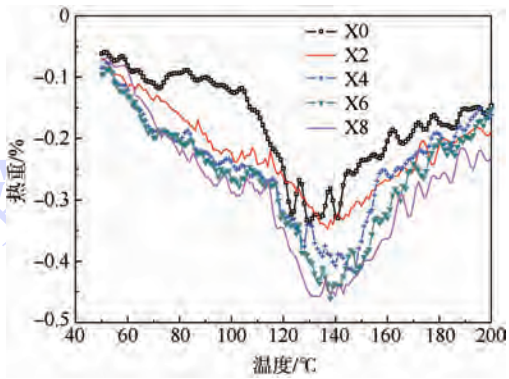


图 8　X 系列净浆 DTG 分析曲线

Fig. 8　DTG analysis curve of X paste

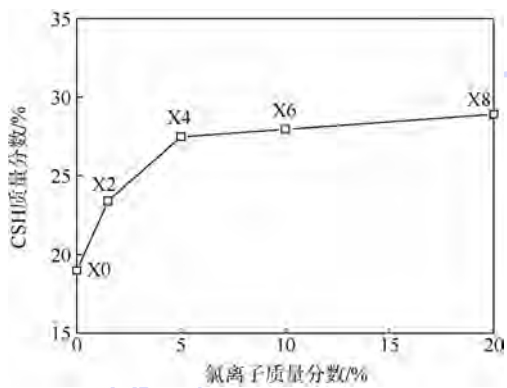


图 9　不同 Cl^- 掺量下 X 中 CSH 含量曲线

Fig. 9　CSH content of the X paste withdifferent chloride ion content

由图 9 可见,随着 NaCl 掺量的增加,X 系列净浆中 CSH 质量分数明显增加,在 Cl^- 掺量为 0% ~5% 区间内增长较快,在 Cl^- 掺量大于 5% 后增长趋缓,X8 中的 CSH 生成量比 X0 的 CSH 生成量增加了约 48%.

3　分析与讨论

由 XRD 测试结果可见,KC 中 $Ca(OH)_2$ 晶体作为碱激发剂参与矿渣水化而被消耗.掺有氯盐的 X 和 Y 系列净浆中均有水化产物 Fs 生成,Fs 生成量随 Cl^- 掺量增加而增加,且净浆中 Fs 生成的同时 NaCl 和 $CaCl_2$ 被大量消耗.矿渣中活性氧化铝($Al_2O_3^{r-}$)、氧化钙(CaO)在碱激发剂作用下

结合氯离子所生成的含氯水化物主要是 Fs,且可稳定存在[24]. 结合表 4 及图 5、图 6 可见,X 系列净浆随 NaCl 掺量的增加,被结合的 Cl^- 量与游离的 OH^- 量在增加;Y 系列净浆则随 $CaCl_2$ 掺量的增加,仅被结合的 Cl^- 量增加而游离的 OH^- 量基本未变. 由此提出 KC 净浆在 NaCl 和 $CaCl_2$ 环境下可能的化学反应方程式(5)和式(6):

$$Al_2O_3^{r-} + 4CaO + 2NaCl + 11H_2O = 3CaO \cdot Al_2O_3 \cdot CaCl_2 \cdot 10H_2O(Fs) + 2NaOH \quad (5)$$

$$Al_2O_3^{r-} + 3CaO + CaCl_2 + 10H_2O = 3CaO \cdot Al_2O_3 \cdot CaCl_2 \cdot 10H_2O(Fs) \quad (6)$$

由式(5)和式(6)可知,由于氯盐阳离子有别,NaCl 与 $Al_2O_3^{r-}$ 发生反应生成 Fs 和 NaOH,而 $CaCl_2$ 与 $Al_2O_3^{r-}$ 反应仅生成 Fs. X 系列净浆中 Cl^- 被化学结合生成 Fs 的同时生成有强碱 NaOH,其可电离出大量 OH^-;而 Y 系列净浆中 Cl^- 被化学结合仅生成 Fs 一种物质.

NaCl 的掺入对 X 净浆强度有增强作用,而 $CaCl_2$ 的掺入对 Y 净浆强度无明显增强作用,两种系列净浆抗压强度差异的产生是掺入氯盐阳离子差异的结果,X 系列净浆中不仅有 Fs 还有 NaOH 生成,而 Y 系列净浆中只有 Fs 生成,由此推知矿渣净浆中生成的 Fs 对净浆抗压强度无影响,而生成的 NaOH 可能是提高 X 净浆抗压强度的根本原因.

矿渣水化反应的主要水化产物是 CSH,其是矿渣强度产生的主要原因. X 系列净浆的 XDR 图谱(图 2(a))显示 X 系列净浆的水化产物除 Fs 以外还有 CSH,且 CSH 衍射峰随 Cl^- 掺量的增加而增加,说明净浆中胶凝性水化物 CSH 生成量在增加,浆体胶结作用得到增强,这可能是 X 系列净浆强度提高的主因. 作为强碱的 NaOH 有利矿渣玻璃体离解作用,可显著促进矿渣水化生成 CSH[25-26]. 结合反应式(5)和图 5、图 7、图 9,可见随着 X 净浆中 Cl^- 掺量的增加,净浆中 NaOH 生成量增加,NaOH 电离后净浆孔隙液中游离 OH^- 增加,液相碱度升高,矿渣水化程度提高,水化生成的胶凝性水化物 CSH 也在增加. 由于矿渣中 $Al_2O_3^{r-}$ 含量有限,直接影响反应式(5)的进行,因此当 $Al_2O_3^{r-}$ 被反应消耗后,反应式(5)会被限制,造成出现净浆孔隙液游离 OH^- 的变化曲线表现为 Cl^- 掺量为 0% ~5% 区间内曲线增长较快,在 Cl^- 掺量大于 5% 后增长明显趋缓,Cl^- 掺量为 5% 时出现的转折点可能是 $Al_2O_3^{r-}$ 的消耗点. 净浆孔隙液游离 OH^- 变化又直接影响了矿渣水化程度及水化生成的 CSH,造成矿渣水化程度及水化生成 CSH 随 Cl^- 掺量的变化趋势与游离 OH^- 变化趋势一致,且这些变化趋势与图 1(a)中 X 系列 28 d 强度随 Cl^- 掺量变化趋势完全一致,这也进一步说明了净浆中 NaOH 生成、水化物 CSH 增长及净浆强度提高有着重要的内在联系. 综合以上测试结果,可以得出:X 系列净浆中的 NaCl 与 $Al_2O_3^{r-}$ 反应生成了 NaOH,使净浆孔隙液中游离 OH^- 增加,孔隙液液相碱度升高,促进了矿渣水化,使其水化生成的 CSH 量明显增多,浆体胶结作用增强,最终宏观表现为净浆强度的提高.

4 结 论

通过强度测试和多种手段的微观分析,得出氯盐环境下矿渣净浆强度变化规律、矿渣作用机理与强度变化的内在联系,结论如下:

1) 掺加 NaCl 的 KC 净浆强度随着 Cl^- 掺量增加而增加;掺 $CaCl_2$ 的 KC 净浆强度随着 Cl^- 掺量增加基本保持不变.

2) 掺加 NaCl 和 $CaCl_2$ 的 KC 净浆中生成的水化产物 Fs 对 KC 净浆抗压强度的提高无明显作用.

3) 掺加 NaCl 的 KC 净浆中生成的 NaOH 是净浆抗压强度随 NaCl 掺量增加而显著提高的主要原因,这是因为 NaOH 生成提高了净浆液相碱度,有利于促进矿渣水化,使 CSH 生成量增加,浆体胶结作用增强,故有效提高了 KC 净浆的抗压强度.

参考文献(References)

[1] Mehta P K. Concrete durability-fifty years' progress[C]//Proceeding of 2th International Conference on Concrete Durability. Dsetroit: ACI SP126-1, 1991: 1-31.

[2] 戈雪良. 混凝土抗海水侵蚀试验研究及其抗蚀性能预测[D]. 武汉:武汉大学,2005.
Ge X L. Experimental study on sea water corrosive action of concrete and forecast the corrode resisting property of concrete[D]. Wuhan: Wuhan University, 2005(in Chinese).

[3] 范颖芳,张英姿,胡跃东,等. 氯化钠侵蚀混凝土力学性能的试验研究[J]. 大连海事大学学报,2008,34(1):125-128.
Fan Y F, Zhang Y Z, Hu Y D, et al. Experimental study on property of concrete corroded by NaCl[J]. Journal of Dalian Maritime University, 2008, 34(1): 125-128(in Chinese).

[4] Douglas E, Bilodeau A, Malhotra V M. Properties and durability of alkali-activated slag concrete[J]. Canada Cent for Mineral

and Energy Technology,1992,89(5):509-516.

[5] 黎鹏平. 胶凝材料组成与钢筋混凝土氯离子腐蚀研究[D]. 广州:华南理工大学,2010.
Li P P, Research on corrosion of binder composition and reinforced concrete to chloride ion [D]. Guangzhou: South China University of Technology,2010(in Chinese).

[6] 余红发. 盐湖地区高性能混凝土的耐久性、机理与使用寿命预测方法[D]. 南京:东南大学,2004.
Yu H F. Study on high performance concrete in salt lake: durability, mechanism and service life prediction [D]. Nanjing: Southeast University,2004(in Chinese).

[7] Higgins D. The effect of GGBS on the durability of concrete[J]. Concrete,1991(9-10):17-21.

[8] Bentur A, Ben-Bassat M. Durability of high performance concretes in highly concentrated magnesium solutions [C]//Proceedings of the International Conference on Durability of Building Materials and Components 6. Tokyo: University of Tokyo Press,1992:632-637.

[9] Deja J, Malolepszy J. Resistance of alkali-activated slag mortars to chloride solution [J]. ACI Special Publication, 1989, 114: 1485-1499.

[10] Shi C, Roy D, Krivenko P. Alkali-activated cement and concretes[M]. Boca Raton: CRC Press,2006:8-10.

[11] 罗睿,蔡跃波,王昌义. 磨细矿渣净浆和砂浆结合外渗氯离子的性能[J]. 建筑材料学报,2001,4(2):148-153.
Luo R, Cai Y B, Wang C Y. Binding capability of chloride ions mortar and paste with ground granulated blast furnace slag[J]. Journal of Building Materials, 2001, 4(2): 148-153 (in Chinese).

[12] 罗睿,蔡跃波,王昌义,等. 磨细矿渣抗氯离子侵蚀性能的机理研究[J]. 土木工程学报,2002,35(6):100-104.
Luo R, Cai Y B, Wang C Y, et al. The mechanism of ground granulated blast-furnace slag to resist chloride corrosion [J]. China Civil Engineering Journal, 2002, 35(6): 100-104 (in Chinese).

[13] 孙国文,管学茂,孙伟,等. 水泥-矿渣体系对 Cl^- 结合能力及机理研究[J]. 武汉理工大学学报,2010,32(7):39-43.
Sun G W, Guan X M, Sun W, et al. Research on the binding capacity and mechanism of chloride ion based on cement-GGBS system[J]. Journal of Wuhan University of Technology, 2010, 32(7):39-43(in Chinese).

[14] 余红发,翁智财,孙伟,等. 矿渣掺量对混凝土氯离子结合能力的影响[J]. 硅酸盐学报,2007,35(6):801-806.
Yu H F, Weng Z C, Sun W, et al. Influences of slag content on chlorine ion binding capacity of concrete [J]. Journal of the Chinese Ceramic Society, 2007, 35(6): 801-806 (in Chinese).

[15] Talero R, Trusilewicz L, Delgado A, et al. Comparative and semi-quantitative XRD analysis of Friedel's salt originating from pozzolan and Portland cement[J]. Construction and Building Materials, 2011, 25(5): 2370-2380.

[16] Zhu Q, Jiang L H, Chen Y, et al. Effect of chloride salt type on chloride binding behavior of concrete [J]. Construction and Building Materials, 2012(37): 512-517.

[17] 葛安亮. 钢筋混凝土在海水中的腐蚀性能研究[D]. 青岛:中国海洋大学,2005.
Ge A L. The study of reinforced concrete corrosion in seawater [D]. Qingdao: Ocean University of China, 2005 (in Chinese).

[18] JTJ 270-1998 水运工程混凝土试验规程[S]. 北京:人民交通出版社,1998:202-207.
JTJ 270-1998 Testing code of concrete for port and waterway engineering[S]. Beijing: China Communications Press, 1998: 202-207 (in Chinese).

[19] GB/T 176-2008 水泥化学分析方法[S]. 北京:中国标准出版社,2008:27-28.
GB/T 176-2008 Methods for chemical analysis of cement[S]. Beijing: Standards Press of China, 2008: 27-28 (in Chinese).

[20] Kondo R, Ohsawa S. Studies on a method to determine the amount of granulated blast furnace slag and the rate of hydration of slag in cement [C]//Proceedings of the 5th International Congress on the chemistry of Cement. Tokyo: Archiworld, 1972: 225-262.

[21] 宁建国. 软土固化剂优化设计原理与方法[D]. 北京:北京航空航天大学,2007.
Ning J G. Principle and method of optimization design for soft-soil stabilizer [D]. Beijing: Beijing University of Aeronautics and Astronautics, 2007 (in Chinese).

[22] 廉慧珍. 建筑材料物相研究基础[M]. 北京:清华大学出版社,1996.
Lian H Z. The phase research foundation of building materials [M]. Beijing: Tsinghua University Press, 1996 (in Chinese).

[23] 薛君玕,吴中伟. 膨胀和自应力水泥及其应用[M]. 北京:中国建筑工业出版社,1985:150-162.
Xue J G, Wu Z W. Application of expansion and self-stressing cement[M]. Beijing: China Architecture & Building Press, 1985:150-162(in Chinese).

[24] Taylor H F W. Cement chemistry[M]. London: Thomas Telford Ltd, 1997:80-88.

[25] Fernández-Jiménez A, Puertas F. Effect of activator mix on the hydration and strength behaviour of alkali-activated slag cements [J]. Advances in Cement Research, 2003, 15(3): 129-136.

[26] Wang S D, Scrivener K L. 29 Si and 27 Al NMR study of alkali-activated slag [J]. Cement and Concrete Research, 2003, 33(5):769-774.

北航学报 赠阅

Experiment of chloride effect on strength of alkali-activated slag paste

CHENG Yin, HUANG Xin*

(School of Transportation Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: Based on the results of unconfined compression tests of alkali-activated slag (KC) paste mixed with sodium chloride (NaCl) and calcium chloride ($CaCl_2$), the influence law of the KC paste strength changing with chloride was studied. X-ray diffraction (XRD), thermal gravity-differential thermal gravity analysis (TG-DTG), the test of ionic concentrations in the pore solution and the test of degree of hydration were used to analyze the mechanism of effect of two kinds of chlorideon KC paste strength. The research results show that the KC paste compressive strength increased with the increase in the concentration of NaCl; and KC paste compressive strength basically unchanged with the increase in the concentration of $CaCl_2$. The Friedel's salt (Fs) was generated in KC paste mixed with NaCl and $CaCl_2$, but it has no effect on the paste strength. The generation of sodium hydroxide (NaOH) in the KC paste mixed with NaCl was the main reason for its strength enhancement. NaOH existed in the KC paste improved the liquid alkalinity of paste, promoted hydration of slag and generated more calcium silicate hydrate (CSH) to improve the paste strength.

Key words: NaCl; $CaCl_2$; alkali-activated slag; mechanism; Friedel's salt; NaOH

北京航空航天大学学报

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0281

基于涡方法生成大涡模拟进口条件的数值计算

谢宝林[1,2], 邵亮*[2]
(1. 北京航空航天大学 中法工程师学院, 北京 100191; 2. 法国里昂流体与声学实验室, 里昂 69130)

摘 要: 探索改进涡方法来生成大涡模拟的非定常进口条件. 改进方法中,为了避免局部漩涡数量过多或者过少,采用密度分布方式放置漩涡场. 并用 Langevin 方程控制漩涡运动,模拟实验方法中的蜂窝器,使改进后的涡方法生成的脉动速度场更加符合湍流的特征. 在已知雷诺平均的流场结果下,利用涡方法产生漩涡场,进而生成能满足大涡模拟所需要的非定常进口流场. 为了检验改进后的涡方法在生成脉动速度过程中的效果,在槽道中进行了对比数值试验,并借助于直接数值模拟数据做对比,对比分析槽道进出口的平均速度、涡量以及雷诺应力统计,证明改进后的涡方法在生成大涡模拟进口条件下是非常有效的.

关 键 词: 涡方法; 槽道; 湍流; 大涡模拟; 雷诺平均
中图分类号: O357
文献标识码: A **文章编号**: 1001-5965(2015)04-0701-06

关于生成大涡模拟非定常进口条件的研究一直以来都是一个难题. 在很多计算流体力学的数值模拟中,例如使用大涡模拟对叶轮机进行的数值模拟,计算结果在很大程度上受进口条件影响[1-2]. 大涡模拟进口的流场需要符合湍流的统计特性,生成大涡模拟进口条件的方法要尽可能地容易操作,这样针对不同的进口情况能快速有效地生成满足要求的流场. 近些年来,已经有一些方法可以生成大涡模拟的进口条件. 其中一类是直接在平均流场上叠加一个在时间和空间上无相关关系的脉动信号,但这类方法得到的流场无法满足湍流的统计特性,与真实的湍流相差甚远. 另一类方法是在进出口采用循环条件,初始速度场是在平均速度场上添加一个微小的扰动,然后代入到 Navier-Stokes 方程中计算,在流向计算长度足够长的情况下,能够得到充分发展的湍流,但由于流向计算长度太长,这种方法的计算成本较高. 还有一类通过傅里叶函数合成方法,但这类方法受限于进口的几何形状,应用起来不够方便[3]. 另外一类是涡方法,通过在进口处安放漩涡场,利用大小不同而且能够移动的漩涡场来生成脉动速度场[4-6]. 由于涡方法容易操作并且在流向上不需要太长的长度就能生成比较符合要求的湍流,所以本文针对此类方法进行了研究. 采用由里昂流体与声学实验室的 Sergent 提出的涡方法[4]对槽道湍流进行探索性数值计算实验. 在改进涡方法中,主要通过 Langevin 方程来控制漩涡的运动以模仿实验方法中的蜂窝器,期望通过较短的流向长度来生成充分发展的湍流.

1 涡 方 法

1.1 涡方法的基本原理

非定常的脉动速度场需要至少满足以下条件:平均脉动速度为零;湍动能 $k=\frac{1}{2}\overline{u_j^2}$($u_j$ 为脉动速度);耗散率 $\varepsilon=2\nu\overline{s_{ij}s_{ij}}$($s_{ij}$为瞬时变形率张

收稿日期: 2014-05-19; 录用日期: 2014-09-16; 网络出版时间: 2014-10-22 20:29
网络出版地址: www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0281.html
作者简介: 谢宝林(1989—),男,湖北随州人,硕士生,buaa1214@gmail.com
*通讯作者: 邵亮(1963—),男,江苏太仓人,研究员,lshao@ec-lyon.fr,主要研究方向为流体力学.

引用格式: 谢宝林, 邵亮. 基于涡方法生成大涡模拟进口条件的数值计算[J]. 北京航空航天大学学报, 2015, 41(4): 701-706.
Xie B L, Shao L. Generation of unsteady inflow conditions for compressible LES with a modified vortex method[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(4): 701-706 (in Chinese).

量,ν 为运动黏度)与 RANS(Reynolds-Averaged Navier-Stokes)计算结果匹配.

涡方法的主要目的是通过漩涡场生成湍流脉动速度场,基本原理构图如图 1 所示.

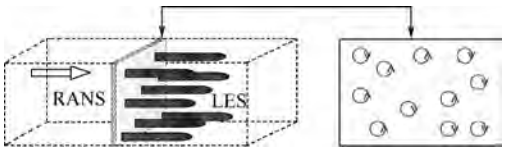


图 1 涡方法基本原理构图

Fig. 1 Configuration of vortex method

二维情况下,涡方法是基于涡量的动力学方程:

$$\frac{\partial \boldsymbol{\omega}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\omega = \nu \nabla^2 \boldsymbol{\omega} \tag{1}$$

$$\boldsymbol{u} = \nabla \times \boldsymbol{\psi} + \nabla \phi \tag{2}$$

式中,$\boldsymbol{\omega}$ 为涡量;$\boldsymbol{\Psi}$ 为二维情况下流体的流函数;ϕ 为速度势. 对式(2)求旋度,可以得到:

$$\boldsymbol{\omega} = -\nabla^2 \boldsymbol{\psi} \tag{3}$$

理想中的漩涡在空间中的分布方式为狄拉克分布,那么式(3)的解析解可以通过对涡量进行卷积运算得到:

$$\boldsymbol{\psi}(\boldsymbol{x}) = -\frac{1}{2\pi}\iint_{R^2} \ln|\boldsymbol{x} - \boldsymbol{x}'|\boldsymbol{\omega}(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' \tag{4}$$

将式(4)代入到式(2),就可以得到速度场和漩涡场的关系表达式[4](毕奥-萨伐尔定律):

$$\boldsymbol{u}(\boldsymbol{x}) = -\frac{1}{2\pi}\iint_{R^2} \frac{(\boldsymbol{x} - \boldsymbol{x}') \times \boldsymbol{\omega}(\boldsymbol{x}') \cdot z}{|\boldsymbol{x} - \boldsymbol{x}'|^2}\mathrm{d}\boldsymbol{x}' \tag{5}$$

式(5)为卷积计算,可以简写为

$$\boldsymbol{u} = \boldsymbol{K} * \boldsymbol{\omega} \tag{6}$$

式中,* 表示卷积计算;$\boldsymbol{K}$ 为卷积计算的核,在简单的二维涡方法中,$\boldsymbol{K} = \frac{z}{|z|^2}$.

在实际的三维情况下,漩涡在空间中的分布不可能像理想中的狄拉克分布那样,漩涡在空间中的分布由形状方程 $\boldsymbol{\xi}_\sigma$ 和环量值 Γ 来定义,其中 σ 是漩涡的作用半径. 在任意一个位置 $\boldsymbol{x}$ 由漩涡 i 得到的旋量:

$$\boldsymbol{\omega}_i = \Gamma_i(\boldsymbol{x}_i)\boldsymbol{\xi}_\sigma(\boldsymbol{x} - \boldsymbol{x}_i) \tag{7}$$

由式(6),得到相应的脉动速度场为

$$\boldsymbol{u}_i = -\frac{1}{2\pi}\Gamma_i(\boldsymbol{x}_i)\boldsymbol{K}_\sigma(\boldsymbol{x} - \boldsymbol{x}_i) \tag{8}$$

式中 $\boldsymbol{K}_\sigma = \boldsymbol{K} * \boldsymbol{\xi}_\sigma$

全场的速度和漩涡可以用以下的关系式给出:

$$\boldsymbol{u}(\boldsymbol{x}) = -\frac{1}{2\pi}\iiint_{R^3} N(\boldsymbol{x}')\Gamma(\boldsymbol{x}')\boldsymbol{K}_\sigma(\boldsymbol{x} - \boldsymbol{x}')\mathrm{d}\boldsymbol{x}' \tag{9}$$

$$\boldsymbol{\omega}(\boldsymbol{x}) = \iiint_{R^3} N(\boldsymbol{x}')\Gamma(\boldsymbol{x}')\boldsymbol{\xi}_\sigma(\boldsymbol{x} - \boldsymbol{x}')\mathrm{d}\boldsymbol{x}' \tag{10}$$

式中,$N(\boldsymbol{x}')$ 为在任意一个位置 $\boldsymbol{x}'$ 的漩涡场的密度.

1.2 运用涡方法的基本流程

这个方法最早是由 Sergent 提出的,并使用商用计算流体力学软件 Fluent 进行了验证[7]. 涡方法的使用流程具体如下.

在进口界面上随机分布一个漩涡场.

针对所添加的每一个漩涡,确定其环量值以及作用半径(σ 和 Γ).

在数值模拟计算的每一个网格上,计算出由漩涡场生成的脉动速度场,然后将其叠加到平均速度的剖面上.

控制漩涡的运动. 例如,在进口界面上平移或者使其绕着某一轴旋转.

每隔一段时间(例如 k/ε),可以调转漩涡的方向.

通过漩涡场计算出相应的脉动速度场.

2 改进后的涡方法的参数设定

2.1 漩涡的形状函数

漩涡的形状函数采用式(11)给出的形式,这种形状函数是改进的高斯函数型,与原始的涡方法中对形状函数的设定保持一致.

$$\boldsymbol{\xi}_\sigma(\boldsymbol{x} - \boldsymbol{x}') = \frac{1}{2\pi\sigma^2}\left(2\mathrm{e}^{-\frac{|\boldsymbol{x}-\boldsymbol{x}'|^2}{2\sigma^2}} - 1\right)\mathrm{e}^{-\frac{|\boldsymbol{x}-\boldsymbol{x}'|^2}{2\sigma^2}}z \tag{11}$$

2.2 漩涡的作用半径

原始的涡方法中,Sergent 用式(12)来给出漩涡的作用半径 σ,这个表达式基于 k 和 ε:

$$\sigma = \frac{C_\mu^{3/4}}{2}\frac{k^{3/2}}{\varepsilon} \tag{12}$$

式中,k 为当地的湍动能;C_μ 为湍流模型常数.

在改进后的涡方法中,对漩涡半径的参数设定是基于均匀各向同性湍流的假设. 通过匹配由速度场表达的湍动能以及由漩涡场来表达的耗散率,可以得到近似于泰勒尺度的漩涡半径:

$$\sigma = \sqrt{\frac{4\nu k}{3(2\ln 3 - 3\ln 2)\varepsilon}} \tag{13}$$

2.3 漩涡的环量和密度分布

基于均匀各向同性湍流的假设所得到的漩涡的环量与密度的关系表达式为

$$\Gamma = 4\sqrt{\frac{\pi S k}{3N(2\ln 3 - 3\ln 2)}} \tag{14}$$

式中,N 为漩涡的数量. 原始的涡方法中,漩涡的

分布方式是随机的，N 个漩涡随机的分布在进口处. 这种随机分布漩涡的方式有一定缺陷，可能导致局部的漩涡数量远大于或者远小于实际需要的漩涡数量.

为了更为精确地确定在不同位置添加漩涡的数量，当地的漩涡数量应该和当地的湍动能和耗散率相关联. 由此提出漩涡场的密度分布方式：

$$N = 4\pi\sigma^2\varepsilon/\Gamma^2 \tag{15}$$

2.4　对漩涡场的运动的控制

原始的涡方法中，漩涡的移动是完全随机的. 这样产生的一个副作用是，它会打乱脉动速度场在时间和空间的相关性. 因此改进后的涡方法通过 Langevin 方程[8-9]来控制漩涡的运动：

$$X(t+\Delta t) = X(t) - X(t)\frac{\Delta t}{T_L} + \sqrt{\frac{2k\Delta t^3}{T_L}}\Xi(t) \tag{16}$$

式中，$X(t)$ 为漩涡中心在 t 时刻的位置；Δt 为时间步长；T_L 为该过程的时间尺度，也叫作释放时间，$T_L = \dfrac{4k}{3C_\mu\varepsilon}$，$C_\mu = 0.09$；$\Xi(t)$ 为一个白噪信号.

实质上，这个过程是模仿实验方法中的蜂窝器来生成湍流[10]. 蜂窝器通过网格来扰动平均速度场，进而生成脉动速度，这与添加的漩涡场的作用是一样的. 蜂窝器的网格所具备的分形结构与通过控制方程来控制漩涡场的运动具有相似性. 以 Langevin 方程控制漩涡的运动，来重构流场在时间和空间的相关关系，以便尽快地生成湍流，这与蜂窝器的网格的分形结构所起的作用是一样的.

3　数值模拟实验

通过一个雷诺数为 395 的槽道湍流算例来检验改进后的涡方法. 数值模拟使用的程序为法国里昂流体与声学实验室（LMFA）开发的 Turb' Flow. 槽道的大小为 $2\pi h \times 2h \times \pi h$，$h$ 是槽道半高（$h = 0.01$ m）. 3 个方向的网格数为 $49 \times 89 \times 41$. 流向和展向网格均匀分布，y 方向的网格分布如下：

$$y = h \times \frac{\tanh(2.65\eta)}{\tanh 2.65}$$

式中，$\eta = 2\dfrac{j - j_{\min}}{j_{\max} - j_{\min}} - 1$，$j$ 为沿 y 方向的网格编号值. 具体网格分布如图 2 所示.

为了加速计算，整个槽道并行分配为 8 个网格块. 槽道 y 方向的两个面是无滑移壁面，x 方向是流向，进口处的边界条件通过涡方法添加，出口处为压力边界. z 方向采用周期性边界条件.

数值模拟计算所需要的平均场由 RANS 计算得到，所采用的 RANS 模型是 Wilcox k-ω 模型[11]. 大涡模拟中使用的亚格子模型为 WALE 模型[12-13].

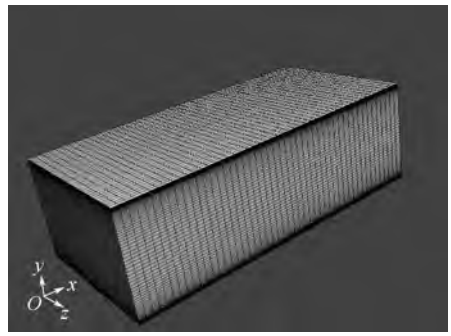


图 2　槽道的网格分布示意图

Fig. 2　Meshes of channel

为了和原始的涡方法进行对比，第 1 种模拟计算采用了在进口界面随机分布的 800 个漩涡来生成漩涡场，所有漩涡在进口界面的平移也是随机的，按照式（11）、式（13）和式（14）来设定漩涡的形状函数、作用半径和环量值，这与原始的涡方法是一致的. 第 2 种计算采用了改进后的涡方法，按照式（11）、式（13）和式（14）设定漩涡的形状函数、作用半径和环量值，但漩涡场采用式（15）的密度分布方式，总的漩涡数目依然为 800 个，并且运用 Langevin 方程（16）来控制漩涡场的运动. 试验结果的参考数据来自于 DNS（Direct Numerical Simulation）的计算结果[14-16].

4　计算结果及分析

4.1　平均速度场

实验所得到的平均速度场的结果由图 3 和图 4 给出. 在进口处，由于平均速度场是根据 RANS 的计算结果给定，所以无论是在平均速度的线性区还是对数区，它都比较符合直接数值模

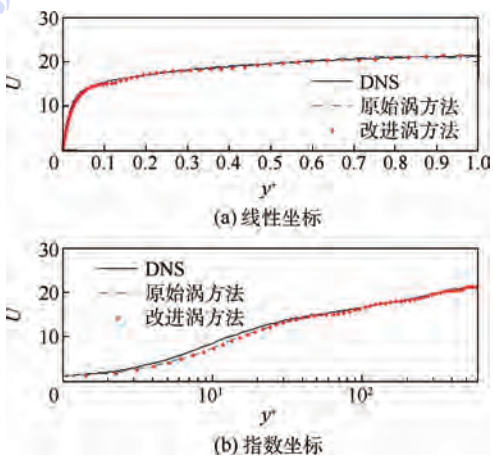


图 3　进口平均速度场的统计结果

Fig. 3　Mean velocity of inlet

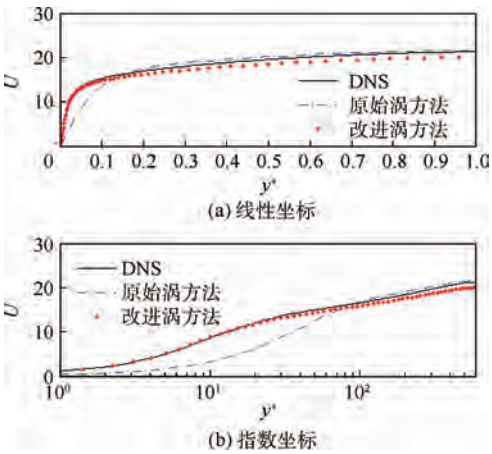


图 4 出口平均速度场的统计结果

Fig. 4 Mean velocity of outlet

拟的结果. 在出口处,平均速度也能和直接数值模拟的数据相吻合. 因为湍流沿着流向不断发展过程中,脉动速度越来越符合湍流的脉动统计特性,对平均速度场统计的影响越来越小.

4.2 涡 量

脉动涡量 ω'的结果由图 5 和图 6 给出. 很容易观察到,在进口处,传统的涡方法不能够得到准确的统计结果. 而采用密度分布方式并且由 Langevin方程来控制漩涡运动的改进涡方法能够生成与 DNS 统计数据相吻合的脉动涡量场,这是由于采用密度分布漩涡的方式能准确地根据不同

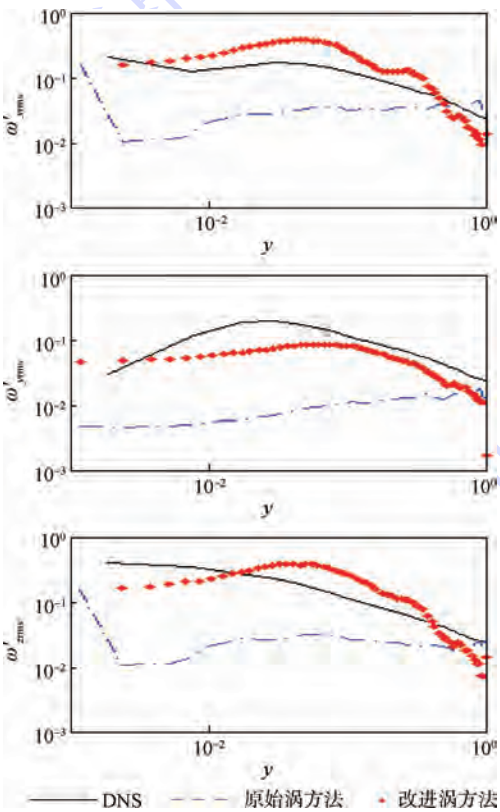


图 5 进口处的涡量统计结果

Fig. 5 Vorticity of inlet

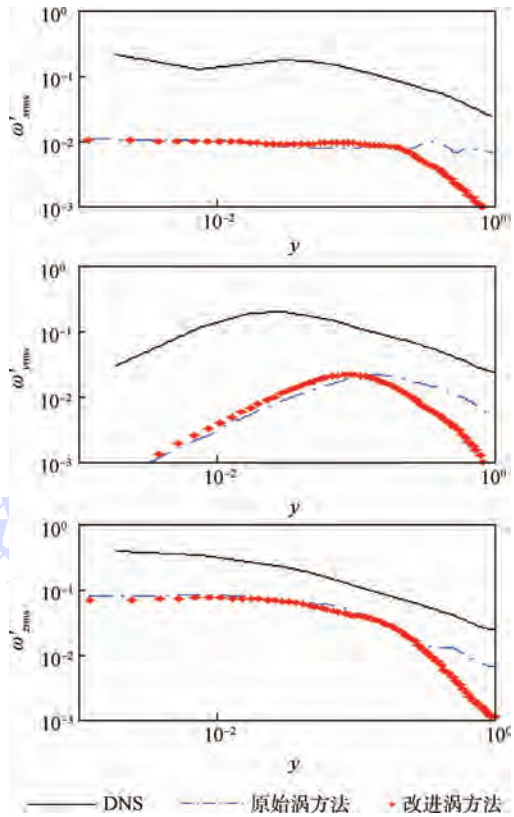


图 6 出口处的涡量统计结果

Fig. 6 Vorticity of outlet

位置的需求量来分布漩涡的数目,从而得到更接近真实湍流的脉动速度场.

流体向下游流动的过程中,湍动能在不同尺度的脉动间传递,直至传递到 Kolmogorov 尺度并耗散掉,这个过程中能量会衰减. 并且由于进口处添加的漩涡场产生的脉动速度场在统计上不能完全符合湍流的统计特性,这种初始场在通过 Navier-Stokes 方程求解的过程中,会摧毁掉初始场所包含的流场信息. 但从曲线的走势来看,出口处,曲线更加平滑,更加趋近于 DNS 的统计曲线. 对比DNS 数据,在大小上约有一个量级的误差.

4.3 雷诺应力

图 7 和图 8 给出了进出口处的雷诺应力的统计结果. 进口处,改进后的涡方法中,漩涡场的分布跟当地的湍动能和耗散率相关,所以得到的脉动速度的统计更接近 DNS 结果. 在出口处,由于脉动场得到了一定程度的发展,所以雷诺应力的统计曲线更为平滑,改进后的涡方法得到的$\langle uu\rangle$和$\langle uv\rangle$的统计曲线更为接近 DNS 结果. 同时从统计结果可以看出,流体在向下游流动的过程中,湍动能有所衰减,初始场所包含的信息在通过 Navier-Stokes 方程求解的过程中在慢慢丢失,流体更加符合湍流真实的统计特性.

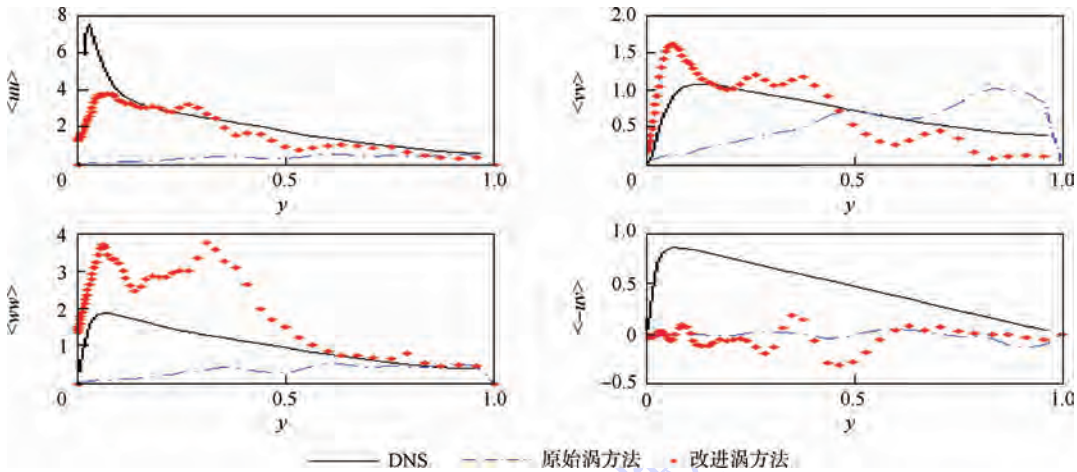


图 7　进口处的雷诺应力统计结果

Fig. 7　Reynolds stress of inlet

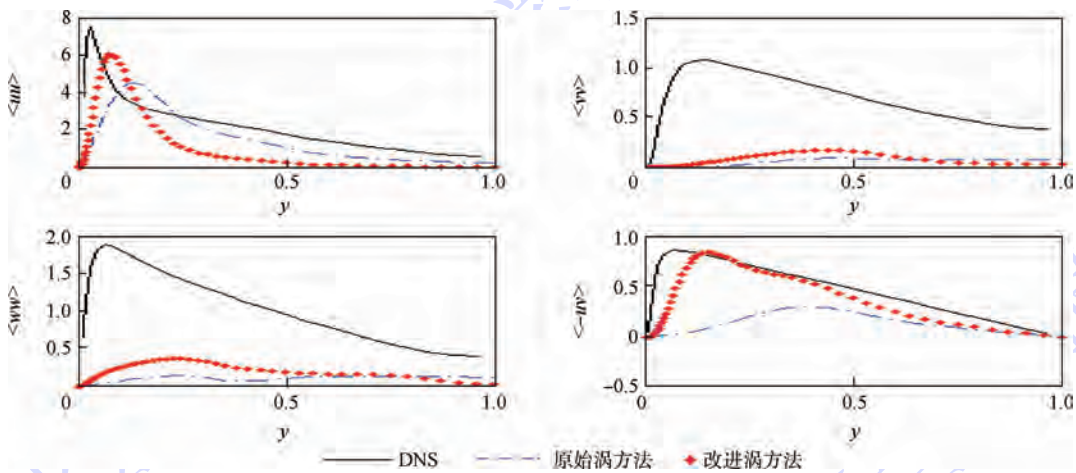


图 8　出口处的雷诺应力统计结果

Fig. 8　Reynolds stress of outlet

5　结　论

实验证明，涡方法能够通过 RANS 的结果来生成脉动场. 改进后的涡方法采用密度分布方式来分布进口漩涡场并通过 Langevin 方程控制漩涡的运动.

1）相比于原始的涡方法，对比涡量和雷诺应力的统计结果，不论是在槽道的进口处，还是出口处，改进后的涡方法生成的脉动速度场更加接近真实的湍流.

2）在出口处，湍流得到一定程度的发展，从统计结果来看，更加符合真实湍流的统计特性，这个过程中，湍流所需要的流向发展长度约为 10 h，对比传统的使用循环条件[2]来生成大涡模拟进口条件的方法，涡方法显得更加有效.

3）另一方面，从进出口处雷诺应力的统计数据来看（$\langle w'w' \rangle$ 和 $\langle v'v' \rangle$），改进后的涡方法所生成的脉动速度场还不能完全和 DNS 数据契合，这方面还需要进一步的探索和改进.

致谢　感谢北京航空航天大学的方乐老师以及法国里昂流体与声学实验室的高峰博士对我的帮助和指导.

参考文献（References）

[1] Nilsen K M, Kong B, Fox R O, et al. Effect of inlet conditions on the accuracy of large eddy Simulations of a turbulent rectangular wake[J]. Chemical Engineering Journal, 2014, 250: 175-189.

[2] Payri R, Gimeno J, Marti-Aldaravi P, et al. Study of the influence of the inlet boundary conditions in a LES simulation of internal flow in a diesel injector[J]. Mathematical and Computer Modelling, 2013, 57(7): 1709-1715.

[3] Tabor G R, Baba-Ahmadi M H. Inlet conditions for large eddy simulation: a review[J]. Computers & Fluids, 2010, 39(4): 553-567.

[4] Sergent E. Vers une methodologie de couplage entre la simulation des grandes echelles et les modeles statistiques[D]. Lyon: Ecole Centrale de Lyon, 2002.

[5] Yokota R, Barba L A. FMM-based vortex method for simulation of isotropic turbulence on GPUs, compared with a spectral method[J]. Computers & Fluids, 2013, 80(1): 17-27.

[6] Leonard A. Vortex methods for flow simulation[J]. Journal of Computational Physics, 1980, 37(3): 289-335.

[7] Mathey F, Cokljat D, Bertoglio J P, et al. Assessment of the vortex method for large eddy simulation inlet conditions[J]. Progressin Computational Fluid Dynamics, 2006, 6(1-3): 58-67.

[8] Chorin A J, Hald O H. Generalized Langevin equations[C]// Stochastic Tools in Mathematics and Science. New York: Springer, 2013: 171-198.

[9] Kursawe J, Schulz J, Metzler R. Transient aging in fractional Brownian and Langevin-equation motion[J]. Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, 2013, 88(6): 062124.

[10] Djenidi L, Tardu S F, Antonia R A. Relationship between temporal and spatial averages in grid turbulence[J]. Journal of Fluid Mechanics, 2013, 730: 593-606.

[11] Parent B, Sislian J P. Validation of the Wilcox k-ω model for flows characteristic to hypersonic airbreathing propulsion[J]. AIAA Journal, 2004, 42(2): 261-270.

[12] Ducros F, Nicoud F, Poinsot T. Wall-adapting local eddy-viscosity models for simulations in complex geometries[C]//Proceedings of 6th ICFD Conference on Numerical Methods for Fluid Dynamics. Oxford: Oxford University Computing Laboratory 1998: 293-299.

[13] Schumann U. Subgrid scale model for finite difference simulations of turbulent flows in plane channel and annuli[J]. Journal of Computational Physics, 1975, 18(4): 3762404.

[14] del Alamo J C, Jimenez J. Direct numerical simulation of the very large anisotropic scales in a turbulent channel[J]. In CTR Annual Research Briefs, 2001: 329-341.

[15] Moser R D, Kim J, Mansour N N. Direct numerical simulation of turbulent channel flow up to $Re = 590$[J]. Physics of Fluids, 1999, 11(4): 943-945.

[16] Grotzbach G. Direct numerical and large eddy simulation of turbulent channel flows[J]. Encyclopedia of Fluid Mechanics, 1987, 6: 1337-1391.

Generation of unsteady inflow conditions for compressible LES with a modified vortex method

XIE Baolin[1,2], SHAO Liang[*2]

(1. Sino-French Engineering School, Beijing University of Aeronautics and Astronautics, Beijing 100191, China;
2. Laboratory of Fluid Mechanics and Acoustics, Lyon 69130, France)

Abstract: This article is to explore and modify vortex method to generate large eddy simulation (LES) inlet condition. In the modification, a density distribution of vortex which can avoid unreasonable vortex quantity in local place was used. The movement of vortex was controlled by Langevin equation, which mimics the active grid turbulence so as to generate an unsteady flow field which is more like the real turbulence. Knowing the results of Reynolds-averaged Navier-Stokes equations (RANS) and reconstructing the turbulence by vortex method (VM), the method was tested in the experimental channel. The statistical results of both inlet and outlet of the channel were given, which include the mean velocity, the vorticity and the Reynolds stress. By analyzing the generated field with the reference of direct numerical simulation (DNS) results and comparing with the results of original method, during the development of turbulence in the channel, it shows that the modified vortex method is viable and cost-effective for generating unsteady inflow condition for LES.

Key words: vortex method; channel flow; turbulence; large eddy simulation (LES); Reynolds-averaged Navier-Stokes (RANS)

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0297

偏置弹簧对 SMA 驱动器输出性能影响规律

杨宝锋，张小勇*，闫晓军

（北京航空航天大学 能源与动力工程学院，北京 100191）

摘 要：偏置弹簧参数对形状记忆合金(SMA,Shape Memory Alloy)驱动器的输出性能有着非常大的影响.为获得其影响规律,通过试验研究了不同偏置弹簧刚度以及不同弹簧预压缩载荷下,SMA 驱动器的作动位移、响应速度的变化规律以及循环次数对驱动器性能的影响规律.试验结果表明,驱动器最大作动位移及响应速度随弹簧刚度或预载的增大而降低;驱动器性能在前几次循环中衰减较大,随着循环进行性能逐渐趋于稳定;弹簧预载的增大使驱动器性能稳定所需的循环次数减少,而弹簧刚度对驱动器输出性能的循环变化规律影响不大.

关 键 词：形状记忆合金；驱动器；循环性能；锁紧机构；偏置弹簧
中图分类号：V44；V45
文献标识码：A **文章编号**：1001-5965(2015)04-0707-06

形状记忆合金(SMA,Shape Memory Alloy)是20世纪60年代初发现并逐步发展起来的一种新型功能材料[1],目前已广泛应用于航空航天、能源、医学等各个领域[2-3].由于SMA材料内部特殊的热弹性马氏体相变,使其具有形状记忆效应和超弹性效应.其中,形状记忆效应是指材料在应力作用下发生塑性变形,应力卸除经加热后,材料发生奥氏体相变,塑性变形完全消失,材料能恢复到原来形状的现象[4].形状记忆合金驱动器(简称SMA驱动器)就是利用SMA材料的形状记忆效应工作的驱动器.

SMA驱动器,是以形状记忆合金作为驱动元件的一种新型的智能驱动器,具有功重比高、结构简单、体积小、反应快并且可靠性高等诸多优点[5-6],容易实现驱动器的自动化与微型化,有着传统驱动器无法比拟的性能优点,已成为航天解锁装置以及仿生机器人等的理想驱动装置.

近年来,国内外学者已成功设计了多种类型的SMA驱动器,并对驱动器力学性能进行了相应的试验研究[7].然而,对SMA驱动器输出性能的试验研究主要集中在驱动器的加热方式[8-10]、加热电流[11]、冷却方式[12-13]等与驱动元件(SMA丝或SMA弹簧)响应直接相关的参数对驱动器输出性能的影响方面.而对于SMA驱动器其他部件参数,如选用的偏置弹簧的刚度、装配时弹簧的预压缩载荷(简称预载)大小等对驱动器性能影响方面的研究还比较匮乏.尤其是针对偏置弹簧刚度以及弹簧预载对SMA驱动器单次输出性能及多次循环输出性能衰减影响方面的研究,还需要进一步开展.而实际上,选用不同弹簧刚度或不同初始预载大小对驱动器输出性能如作动位移、响应速度等参数有非常大的影响.因此,对不同参数配置下的SMA驱动器输出性能进行深入研究,掌握其影响规律,对SMA驱动器的输出性能优化设计有着重要的意义.

收稿日期：2014-05-23；**录用日期**：2014-08-20；**网络出版时间**：2014-09-22 14:40
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0297.html
基金项目：国家自然科学基金资助项目(11272025)；中央高校基本科研业务费专项资金资助项目(YWF-14-RSC-014)
作者简介：杨宝锋(1989—),男,陕西延安人,硕士生,ybf1000@qq.com
*通讯作者：张小勇(1983—),男,贵州安顺人,讲师,zhangxy@buaa.edu.cn,主要研究方向为形状记忆合金智能结构、航天机构、固体力学.

引用格式：*杨宝锋，张小勇，闫晓军. 偏置弹簧对SMA驱动器输出性能影响规律[J]. 北京航空航天大学学报，2015，41(4)：707-712. Yang B F，Zhang X Y，Yan X J. Relationship between bias spring parameters and output performances of SMA actuators[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：707-712 (in Chinese).*

北航学报 赠阅

基于此,本文通过试验分别研究了选用不同偏置弹簧刚度以及在不同弹簧预载下 SMA 驱动器的力学循环特性.获得了驱动器作动位移、响应速度随弹簧刚度、预紧力的变化规律,并获得驱动器循环次数对其性能衰减的影响规律.

1 试验系统及方法

1.1 试验系统

为了更接近工程中实际使用的驱动器,本文直接采用了 SMA 空间磁悬浮飞轮锁紧机构[14-15]中的解锁驱动器作为试验研究对象,如图 1 所示(点划线线框中为 SMA 驱动器部分).图 2 为相应的 SMA 驱动器结构原理图.

图 1 SMA 空间锁紧机构

Fig. 1 SMA locking device

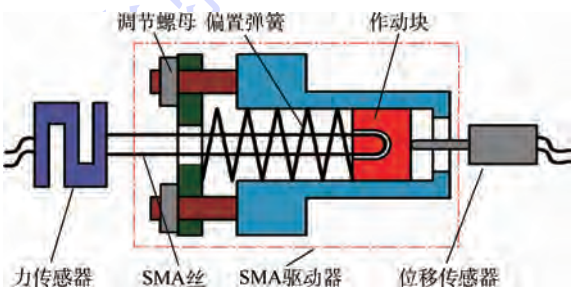


图 2 SMA 驱动器结构原理图

Fig. 2 Schematic diagram of the SMA actuator

正常情况下,驱动器在偏置弹簧弹回复力的作用下处于一个确定状态,即图 2 中作动块处于最顶端位置,对应于空间解锁机构的设计锁紧状态.当对驱动器进行通电加热时,SMA 丝在很短的时间内收缩,产生巨大的回复力,驱动作动块发生作动,同时压缩偏置弹簧,使解锁机构完成解锁功能.断电后,SMA 丝逐渐冷却,发生马氏体相变,输出力降低,在偏置弹簧回复力的作用下,SMA 丝被拉长,驱动器完成复位.

驱动器作动过程中,需测量不同偏置弹簧刚度以及不同的弹簧预紧力下 SMA 驱动器的输出力 F、作动位移 u 以及响应时间 t.如图 2 所示,采用位移传感器测量作动块的位移以获得驱动器的作动位移 u;通过力传感器与 SMA 丝直接相连来获取驱动器输出力 F;驱动器的响应时间 t 为驱动器刚开始通电到驱动器作动位移达到最大时所需时间,由响应时间 t 以及最大作动位移 u_{max} 即可获得响应速度.偏置弹簧预紧力大小可通过图 2 中调节螺母进行调整.整个试验系统如图 3 所示.点划线框中为试验核心部分,其组成如图 4 所示.

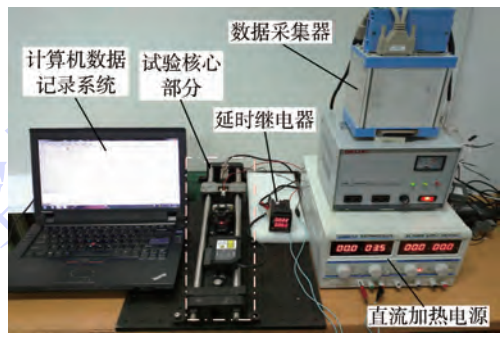


图 3 试验系统

Fig. 3 Testing system

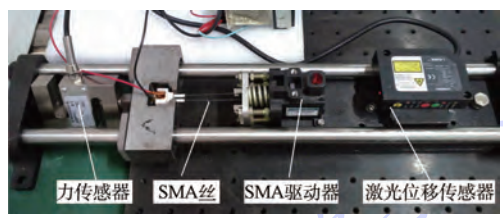


图 4 试验系统核心组成部分

Fig. 4 Core part of the testing system

1.2 试验方法

SMA 驱动器中的 SMA 丝选用沈阳天贺新材料有限公司提供直径 0.7 mm 的 NiTi 形状记忆合金丝(原子百分比为 Ni:50.8%,Ti:49.2%),奥氏体相变温度为 A_s = 50℃.试验前,将 SMA 丝在 450℃退火 5 min,以消除加工硬化影响.此后,对每根丝进行 5% 的预拉伸应变.

试验过程中,通过延时继电器控制稳压直流电源对驱动器的间歇通电和断电,通电电流为 3A,加热时间为 60 s,冷却时间为 90 s.

SMA 驱动器的输出性能影响参数试验分为 3 类:

1)不同偏置弹簧预载对输出性能影响.试验选择刚度为 8.9 N/mm 的弹簧,调节偏置弹簧预载分别为 100,130,160 N 下进行试验,研究驱动器的输出力、作动位移以及响应速度的变化规律.

2)不同偏置弹簧刚度对输出性能影响.选取 3 根刚度为 5.6,8.9,12.4 N/mm 的偏置弹簧,在预载为 100 N 下进行试验,研究驱动器的输出力、作动位移以及响应速度的变化规律.

3）循环次数对性能衰减的影响. 分别对第 1）、第 2）类试验进行 100 次循环驱动，研究驱动器输出性能随循环次数的变化规律.

2　试验结果及分析

2.1　SMA 驱动器输出性能

2.1.1　不同弹簧预载下的输出性能

不同偏置弹簧预载下，驱动器在第 1 次循环过程中的输出力与作动位移之间关系如图 5 所示. 图 6 给出了驱动器最大位移随弹簧预载的变化规律曲线.

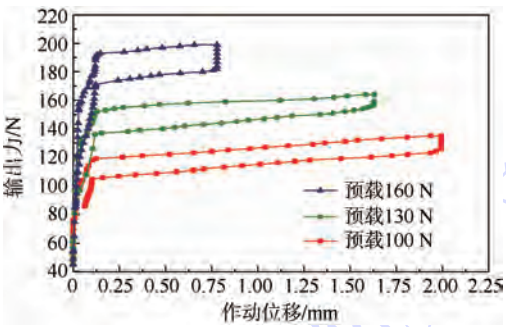


图 5　不同弹簧预载下输出力-位移曲线

Fig. 5　Force-displacement curve under different preloads

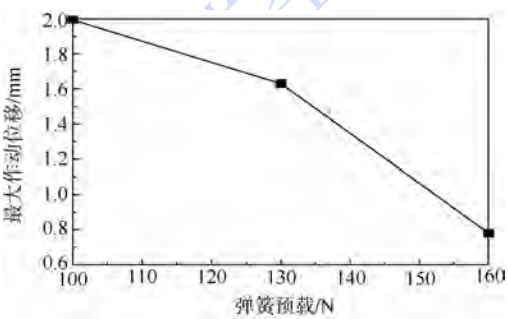


图 6　驱动器最大作动位移随预载变化曲线

Fig. 6　Stroke of actuator under different preloads

由图 5、图 6 可以看出，随着弹簧预载的增加，驱动器的最大作动位移逐渐减小，而驱动器输出力却逐渐增加. 这主要是由于驱动器作动过程中，必须克服偏置弹簧的回复力，弹簧预载越大，作动相同位移所需的力也就越大，因此，输出力也相应增大. 可以推断，当弹簧预载超过一定值后，驱动器将失去作动能力.

驱动器第 1 次循环的响应速度随弹簧预载的变化规律如图 7 所示. 可以看出，随着弹簧预载的增加，驱动器的响应速度呈线性降低的趋势. 这是因为 SMA 丝中存在应力会导致其相变温度升高，而且基本呈线性比例增大. 弹簧预载越大，SMA 丝中的应力就越大，在同样加热电流作用下，其响应时间也就越长，响应速度也相应降低.

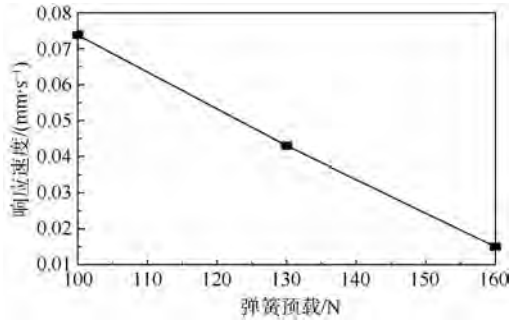


图 7　驱动器响应速度随预载变化曲线

Fig. 7　Response rate of actuator under different preloads

因此，在实际工程应用中，为增加驱动器作动位移、提高驱动器响应速度，可适当降低驱动器装配时偏置弹簧的预载.

2.1.2　不同弹簧刚度下的输出性能

不同偏置弹簧刚度下，驱动器在首次循环过程中，输出力与作动位移之间关系如图 8 所示. 最大作动位移随弹簧刚度变化规律如图 9 所示.

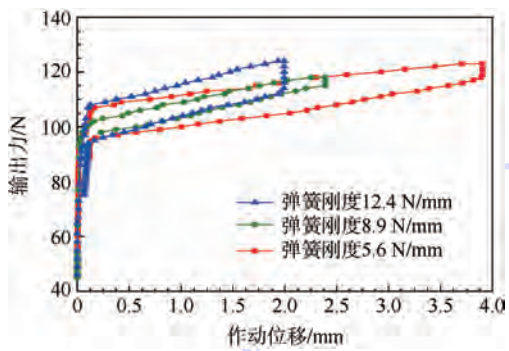


图 8　不同弹簧刚度下输出力-位移曲线

Fig. 8　Force-displacement curve under different stiffness

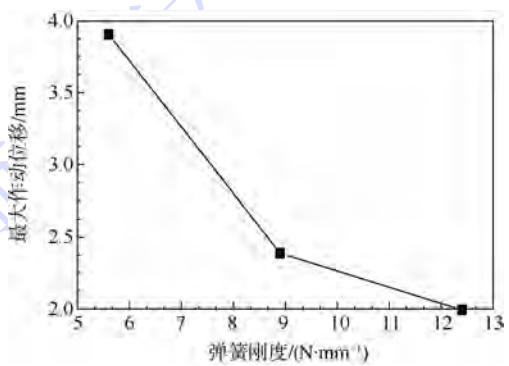


图 9　驱动器最大作动位移随弹簧刚度变化曲线

Fig. 9　Stroke of actuator under different stiffness

由图 8、图 9 可以看出，随着弹簧刚度的增加驱动器最大作动位移逐渐减小. 此外，从驱动器作动位移与输出力的关系可以看出不同弹簧刚度下，驱动器循环曲线斜率有所改变，随着弹簧刚度的增加曲线斜率也逐渐增大. 对比图 5 中，同一弹簧在不同预载下，曲线斜率基本相同，则可以看出，驱动器作动过程中，作动位移与输出力之间关

系主要由偏置弹簧弹性刚度决定.

驱动器稳定后的响应速度随弹簧刚度的变化规律如图 10 所示. 可以看出,随着弹簧刚度增加,驱动器响应速度呈线性减小趋势.

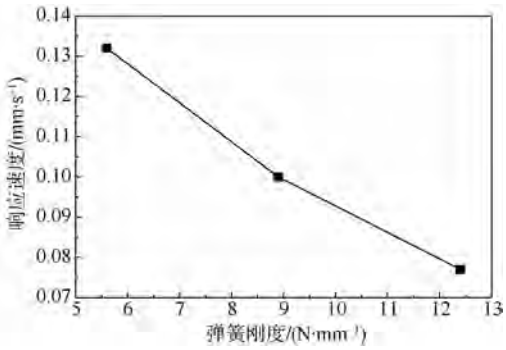


图 10 驱动器响应速度随弹簧刚度变化曲线

Fig. 10 Response rate of actuator under different stiffness

因此,工程应用中,可通过选用刚度较小的弹簧来提高驱动器响应速度和作动位移,以提升驱动器作动能力.

2.2 SMA 驱动器循环性能

2.2.1 循环过程中输出力/位移关系

图 11 所示偏置弹簧刚度为 8.9 N/mm、预紧力为 100 N 时,SMA 驱动器前 100 次循环试验中所得到的输出力-位移曲线.

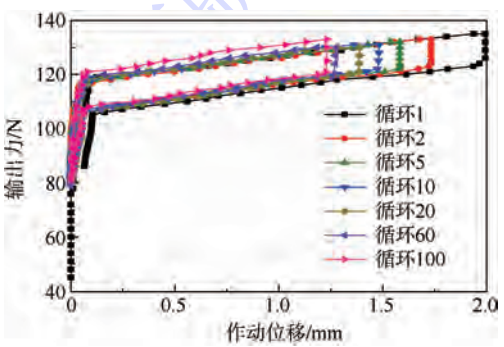


图 11 循环过程中的输出力-位移曲线

Fig. 11 Force-displacement curve in different cycles

由图 11 可以看出,在每次循环过程中,SMA 丝通电加热,驱动器输出力逐渐增大,当输出力达到可以克服系统阻力大小时,对应图中 120 N 左右,驱动器才开始作动,作动过程中,输出力随作动位移增加而增大,最终达到稳定状态(最大位移状态);由 2.1.2 节分析可知,作动过程中,输出力增大幅度主要由偏置弹簧刚度大小决定. 当驱动器停止加热时,输出力降低,但作动位移并没有随之立即减小,而是表现出迟滞现象,等输出力降低一定值后,作动位移才开始减小,最后达到复位. 驱动器作动过程同样也表现出迟滞现象.

循环过程中出现的迟滞现象,主要是由于驱动器作动过程中,作动块与驱动器内壁之间摩擦引起的. 驱动器在通电加热时,SMA 丝的输出拉力 F_{SMA} 迅速增大,与此同时,作动块与内壁之间的静摩擦力 f 以及弹簧回复力 F_s 也逐渐增大,并且与 F_{SMA} 方向相反(如图 12(a)所示),因此驱动器没有立即作动,表现出迟滞现象;当 SMA 丝拉力大于弹簧回复力与最大静摩擦力时,驱动器开始作动.

驱动器达到最大作动位移并稳定后,弹簧回复力保持不变,摩擦力 f 减小为 0;在驱动器复位过程中,断电一瞬间,F_{SMA} 开始减小,而作动块与内壁之间的摩擦力 f 却逐渐增大,并且方向与 F_{SMA} 相同(如图 12(b)所示),两者之和保持不变,与弹簧的回复力 F_s 相等,因此位移保持不变,表现出迟滞现象. 当 F_{SMA} 减小幅度大于最大静摩擦力 f_{max} 的时候,即两者之和小于弹簧回复力 F_s 时,作动位移开始减小,驱动器逐渐复位. 因此,为减小驱动器迟滞现象,应尽量保证驱动器内壁以及作动块光滑,减小两者之间摩擦力.

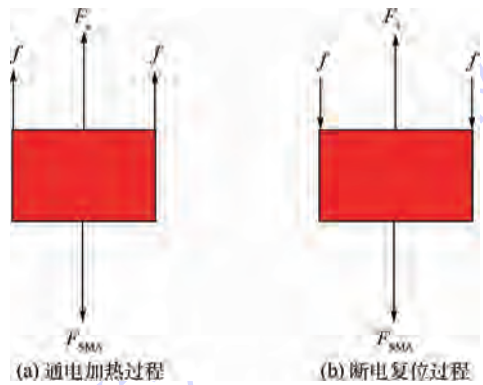


图 12 SMA 驱动器作动块受力图

Fig. 12 Force diagram of the SMA actuator

2.2.2 不同弹簧预载下的循环性能

不同弹簧预载下,在 100 次循环中,驱动器最大作动位移变化趋势如图 13 所示.

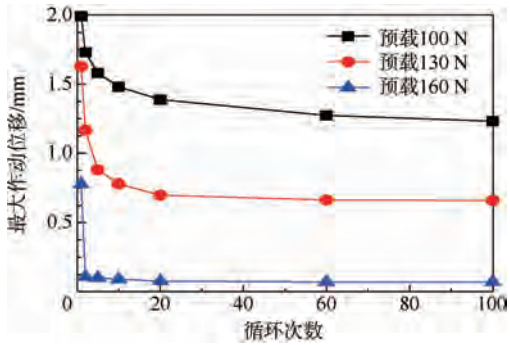


图 13 不同预载下最大作动位移循环变化曲线

Fig. 13 Stroke-cycle curve under different preloads

由图 13 可以看出,在驱动器的前几次循环

中，作动位移衰减幅度较大，随着循环进行，衰减现象逐渐减小，最后趋于稳定状态. 这种规律在其他各组试验中，均可发现. 此外，随着弹簧预载的增加，驱动器前几次循环衰减幅度增大，最大作动位移达到稳定时所需的循环次数逐渐减小. 当预载为 160 N 时，由于循环衰减严重，驱动器在完成 1 次循环之后，已几乎失去作动能力，输出位移接近 0.

图 14 所示的是不同弹簧预载下驱动器的响应速度随循环次数的变化情况. 可以看出，驱动器响应速度大小在前几期循环中波动较大，但随着循环的进行，驱动器稳定后，响应时间也逐渐稳定下来. 此外，弹簧预载仅影响响应速度大小，对响应速度循环变化规律的影响较小.

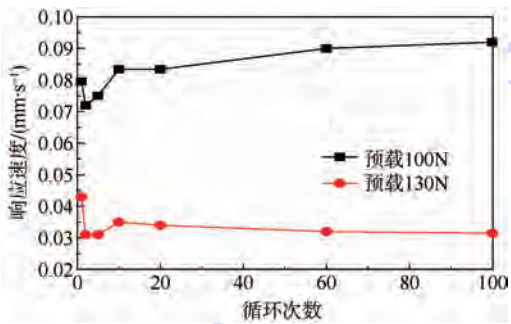


图 14　不同弹簧预载下响应速度循环变化曲线

Fig. 14　Response rate-cycle curve under different preloads

工程应用中，要保证驱动器运行的可靠性，在驱动器使用之前，应先对驱动器进行多次预循环，使驱动器达到性能稳定状态，以防止由于衰减效应而导致驱动器在实际使用中失效. 此外，为使驱动器在更少的循环次数下稳定，在满足所需作动位移下，可适当增大弹簧预紧力.

2.2.3　不同弹簧刚度下的循环性能

不同弹簧刚度下，驱动器在前 100 次循环中的最大作动位移变化趋势如图 15 所示.

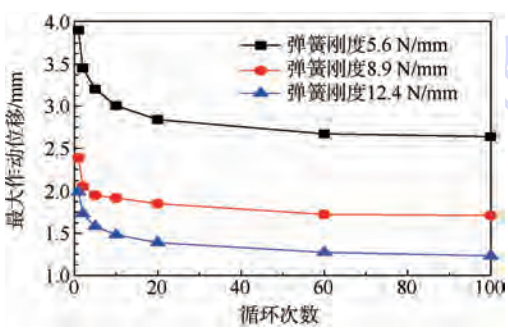


图 15　不同弹簧刚度下最大作动位移循环变化曲线

Fig. 15　Stroke-cycle curve under different stiffness

由图 15 可以看出，随着弹簧刚度的增加，驱动器作动位移逐渐减小，但驱动器最大作动位移随循环次数变化趋势基本相同，驱动器最大位移达到稳定所需循环次数也基本一致，由此可见弹簧刚度对驱动器的循环作动能力影响很小.

图 16 所示的是不同弹簧刚度下驱动器的响应速度随循环次数的变化曲线. 可以看出，与图 13类似，驱动器在前几次循环中，响应速度大小波动较大，随着循环进行逐渐稳定下来. 总的来说，弹簧刚度的增大，会使驱动器响应速度降低，但对驱动器响应速度随循环的整体变化规律影响不大.

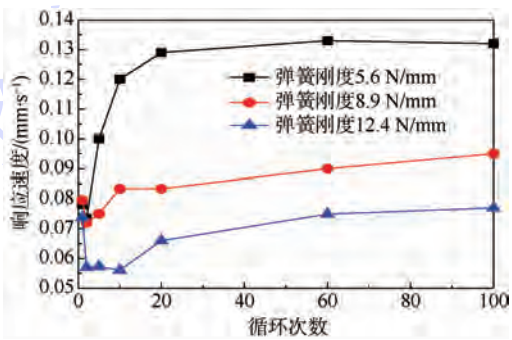


图 16　不同弹簧刚度下响应速度循环变化曲线

Fig. 16　Response rate-cycle curve under different stiffness

3　结　论

1）SMA 驱动器最大作动位移随偏置弹簧刚度、预载的增大逐渐减小；驱动器响应速度也随弹簧刚度、预载的增加而逐渐降低.

2）SMA 驱动器在前几次循环中性能衰减很大，随着循环次数增加性能逐渐趋于稳定. 在实际使用中，应对驱动器进行多次预循环使其性能稳定，防止由于性能衰减效应而造成的驱动器工作失效.

3）弹簧预载的增大，会使驱动器性能稳定所需的预循环次数减少. 而弹簧刚度对驱动器输出性能循环变化规律影响不大.

4）驱动器循环过程中，会出现迟滞现象. 为降低迟滞效应，应使驱动器作动块与驱动器内壁尽可能光滑或采取其他措施，以减小两者之间摩擦力.

参考文献（References）

[1] 高滨. 形状记忆合金在航天器分离机构上的应用[J]. 航天返回与遥感，2005，26(1)：48-52.
Gao B. Shape memory alloys and their application to aerospace separation mechanisms[J]. Spacecraft Recovery & Remote Sensing，2005，26(1)：48-52(in Chinese).

[2] 赵连城，郑玉峰. 形状记忆与超弹性镍钛合金的发展和应用[J]. 中国有色金属学报，2004，14(1)：323-326.
Zhao L C，Zheng Y F. Development and applications of nickel-ti-

tanium alloys with shape memory effect and superelasticity[J]. The Chinese Journal of Nonferrous Metals, 2004, 14(1): 323-326(in Chinese).

[3] Zhang X Y, Yan X J, Yang Q L. Design and experimental validation of compact, quick-response shape memory alloy separation device[J]. Journal of Mechanical Design, 2014, 136(1): 011009.

[4] 贺玉珊. 形状记忆效应[J]. 教学仪器与实验, 1990, 6(2): 34-36.
He Y S. Shape memory effect[J]. Educational Equipment and Experiment, 1990, 6(2): 34-36(in Chinese).

[5] 张义辽. SMA 直线驱动器结构原理及实验研究[D]. 合肥: 中国科学技术大学, 2010.
Zhang Y L. The theory of structure and experimental research on SMA linear actuator[D]. Hefei: University of Science and Technology of China, 2010(in Chinese).

[6] 王宏. 用于仿生推进的 SMA 驱动器研究[D]. 哈尔滨: 哈尔滨工程大学, 2007.
Wang H. Research on SMA actuator applied to biomimetic propel[D]. Harbin: Harbin Engineering University, 2007(in Chinese).

[7] Zhang X Y, Yan X J. Continuous rotary motor actuated by multiple segments of shape memory alloy wires[J]. Journal of Materials Engineering and Performance, 2012, 21(12): 2643-2649.

[8] 杨汝清, 桑子刚. 机器人形状记忆合金(SMA)驱动器及其控制探讨[J]. 机器人, 1990, 12(6): 50-53.
Yang R Q, Sang Z G. The explore of the shape memory alloy (SMA) actuator and its control on robot[J]. Robot, 1990, 12(6): 50-53(in Chinese).

[9] 熊克, 沈文罡. SMA 丝缠绕角对扭力驱动器驱动性能影响的试验研究[J]. 机械工程学报, 2004, 39(12): 123-128.
Xiong K, Shen W G. Experimental research on the relationship between SMA wire wound angle and SMA torsion actuator primary performances[J]. Chinese Journal of Mechanical Engineering, 2004, 39(12): 123-128(in Chinese).

[10] Lee H J, Lee J J. Evaluation of the characteristics of a shape memory alloy spring actuator[J]. Smart Materials and Structures, 2000, 9(6): 817-823.

[11] 唐森, 董二宝, 许雯, 等. 新型 SMA 扭转驱动器设计及实验研究[J]. 功能材料与器件学报, 2011, 17(4): 407-411.
Tang S, Dong E B, Xu W, et al. Design and experimental study of new SMA rotary actuator[J]. Journal of Functional Materials and Devices, 2011, 17(4): 407-411(in Chinese).

[12] 王社良, 薛凌云, 苏三庆. 形状记忆合金驱动器的力学性能[J]. 西安建筑科技大学学报, 1999, 31(3): 233-236.
Wang S L, Xue L Y, Su S Q. Mechanical behavior of shape memory alloy actuator[J]. Journal of Xi'an University of Architecture & Technology, 1999, 31(3): 233-236(in Chinese).

[13] 李宝成, 董二宝, 李永新, 等. 新型 SMA 丝致动器的结构设计与实验研究[J]. 机械与电子, 2010(1): 7-8.
Li B C, Dong E B, Li Y X, et al. Structrue design and experiment research of a novel actuator based on SMA wire[J]. Machinery & Electronics, 2010(1): 7-8(in Chinese).

[14] 闫晓军, 张小勇, 聂景旭, 等. 采用 SMA 驱动的小型空间磁悬浮飞轮锁紧机构[J]. 北京航空航天大学学报, 2011, 37(2): 127-131.
Yan X J, Zhang X Y, Nie J X, et al. Prototype SMA actuated locking device for small space magnetic bearing flywheels[J]. Journal of Beijing University of Aeronautics and Astronautics, 2011, 37(2): 127-131(in Chinese).

[15] Zhang X Y, Yan X J, Zhang S W, et al. Development of a novel shape memory alloy-actuated resettable locking device for magnetic bearing reaction wheel[J]. Review of Scientific Instruments, 2014, 85(1): 015006.

Relationship between bias spring parameters and output performances of SMA actuators

YANG Baofeng, ZHANG Xiaoyong*, YAN Xiaojun

(School of Energy and Power Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: The parameters of the bias spring have great impacts on the output performances of the shape memory alloy (SMA) actuator. To investigate the rules of influences, the output stroke and response rate of a SMA actuator with different bias spring stiffness and precompression load were studied. The output degradation due to cyclic actuation was also investigated. Experimental results show that the strokes and response rates decrease with the increase of preloads or stiffness. The performance of the actuator degrades significantly in the first few cycles and stabilizes eventually with the increase of the number of cycles. The test results also indicate that less cycling is needed to stabilize the actuator when the preload is increased, and the stiffness seems to affect the cyclic performance of the actuator slightly.

Key words: shape memory alloys; actuator; cyclic performance; locking device; bias spring

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0472

基于时空关联图模型的视频监控目标跟踪

张诚，马华东*，傅慧源

（北京邮电大学 智能通信软件与多媒体北京市重点实验室，北京 100876）

摘 要：多摄像机监控环境下的无重叠视域目标跟踪问题十分具有挑战性，其原因在于跟踪目标在网络中的转移与运动规律往往具有不确定性. 目标跟踪的关键问题在于摄像机之间的目标关联以及如何依据网络拓扑结构来找到目标之间的对应关系. 提出了一种图模型来对摄像机网络中的时空关联关系进行表达. 图模型中的节点表示目标在摄像机视域中的出现区域和消失区域，边由时间与空间关系进行约束. 提出了一种将目标外观模型与图模型相融合的跟踪方法，其中外观模型通过协方差描述子进行特征融合，同时，结合二部图匹配策略来解决多摄像头目标跟踪中的识别与匹配问题. 在真实监控视频上的实验验证了该方法的有效性.

关 键 词：目标跟踪；多摄像机网络；时空关联图模型；二部图匹配；数据关联
中图分类号：TP391
文献标识码：A **文章编号**：1001-5965(2015)04-0713-08

与单摄像机下的监控不同，多摄像机监控网络的视野范围往往较大，存在复杂的网络拓扑，区域之间没有视域重叠，特定目标的跟踪作为监控事件分析、目标轨迹预测、物理空间检索等监控任务的基础环节之一，已逐渐成为视频监控领域的研究热点.

多摄像机监控网络中的目标跟踪即实现目标的连续跟踪，其核心问题是判断不同位置摄像机在不同时刻观测到的对象是否相同，这主要涉及以下 3 个子问题：①单摄像目标跟踪. 这是多摄像头下目标跟踪的基础[1]，不同之处在于跨头跟踪没有起始帧的初始化步骤，需要首先进行运动目标的检测与分割[2-3]；②目标识别与匹配. 目标匹配是把不同时刻出现在不同摄像机前的目标进行匹配，从而识别出相同的目标，实现不同摄像机中同一目标的对应. 传统的目标匹配理论往往注重建立目标的表现模型(AM，Appearance Model)，这些特征主要包括：颜色特征、几何特征、区域特征等，已有的研究工作往往选取这些特征中的一种或者几种进行融合来建立表现模型参与匹配[4]；③跨摄像头目标跟踪. 即如何协同多个摄像头来可靠地跟踪目标，这里需要解决的是目标的关联问题. 需要特别指出的是，非重叠区域的跨摄像头目标跟踪并不同于传统的目标重识别问题，由于不同摄像机前的颜色、光照、观测距离和角度等方面的差异，仅仅依赖于表现模型并不足以完成精确的目标匹配. 在特定环境的视频监控中，目标运动和转移的规律依赖于摄像机网络的拓扑结构，一旦建立多摄像机之间的时间、空间、因果等拓扑关系，非重叠区域的跨摄像头目标跟踪算法就能

收稿日期：2014-04-28；**录用日期**：2014-08-01；**网络出版时间**：2014-12-03 14:50
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0472.html
基金项目：国家 863 计划资助项目(2014AA015101)；国家自然科学基金资助项目(61402048)；国家工信部物联网发展专项资金；北京市教育委员会共建项目
作者简介：张诚(1990—)，男，湖北黄冈人，硕士生，zhangcheng@bupt.edu.cn
*通讯作者：马华东(1964—)，男，河南南阳人，教授，mhd@bupt.edu.cn，主要研究方向为物联网与传感网、多媒体系统与网络.

引用格式：*张诚，马华东，傅慧源. 基于时空关联图模型的视频监控目标跟踪[J]. 北京航空航天大学学报，2015，41(4)：713-720.*
Zhang C，Ma H D，Fu H Y. Object tracking in surveillance videos using spatial-temporal correlation graph model[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：713-720 (in Chinese).

获得时空关系的约束,从而使得算法的效率得到大幅提高.综上所述,多摄像机监控下的目标跟踪问题,需要把摄像机的拓扑关系与目标的表现模型匹配算法进行结合.

目前研究者已经提出了很多将一系列数据点以及每对数据点之间的相似性或者距离转化为图结构的方法.为了将目标在摄像机网络中的运动和转移关系进行表示,本文利用有向带权图来建立时空关联关系:图中节点表示摄像机视域(FOV,Field of Views)中目标的出现区域(source region)和消失区域(sink region),节点之间的边表示目标在相邻两个节点之间在空间物理位置直接相邻,边的权重定义为目标在不同节点之间移动的时间概率分布.当目标在网络中移动时,目标关联算法采用了基于时空约束与特征模型融合的二部图最大匹配策略,对不同摄像机观测到的图像帧序列建立权重二部图,通过求取最大权重匹配推断目标下一步移动的节点位置,从而实现对目标的关联跟踪.

1 研究现状

1.1 目标跟踪

单摄像头目标跟踪是跨头跟踪的基础,它主要涉及如下 3 个模块[5]:首先是运动目标的有效表达.目标的表示是跟踪的主要模块,常用的目标表示方法包括使用一些全局特征、局部特征、模板、灰度直方图、生成模型、判别模型等为目标进行建模;其次是搜索机制.在目标跟踪的过程中,需要依据当前帧中目标的位置寻找下一帧中目标的最佳匹配位置,这就依赖于一定的搜索算法对目标的运动状态进行估计和假设.常用的检索策略包括:粒子滤波算法、马尔科夫链蒙特卡洛方法、局部最优化搜索、密集采样等.最后是跟踪模型的更新.为了适应在跟踪过程中目标外观的剧烈变化,表示模型需要进行实时更新以此来避免跟踪框的偏移情况.TLD 算法[6]是近年来提出的一种新颖高效的视频跟踪算法,该算法能够满足实际监控视频中的长时鲁棒跟踪任务,本文将使用 TLD 算法进行单摄像头下的目标跟踪.

1.2 目标识别与关联

很多研究者将不同摄像机之间的目标识别与目标关联问题称之为目标重识别[4].给定输入目标,为了在待匹配目标集合中寻找最匹配的对象,通常需要以下两个步骤:首先,建立输入目标与待匹配对象的特征表示方法;然后,定义合理的相似性度量方法进行度量,从而确定待匹配数据中是否含有相同的输入目标.然而,目标重识别的问题解决框架并不能很好地完成跨摄像头的目标跟踪任务,这是由于:①在多摄像头监控中,外部环境具有不可控性,比如监控视角、距离的改变、环境光、背景的复杂性等,大部分重识别方法依赖于目标外观特征所建立的表现模型变得不再可靠.②单纯的外观特征匹配并没有考虑多摄像机网络具有的空间拓扑关系与时间关联约束,由于目标的运动规律往往依赖于摄像头网络的时空拓扑,如果将时空约束纳入到目标跟踪的框架,将大幅促进识别的效果.

近年来已经有很多研究者开始关注多摄像机监控网络的拓扑结构[7-13],并利用其来构建跨摄像机跟踪任务中的目标关联关系,通常的方法是将摄像机的视域划分为不同的语义节点,通过定义不同节点之间的边来进行摄像机监控网络中的目标跟踪、活动分析、轨迹预测等工作.Makris 等[7]和 Stauffer[8]最开始研究多摄像头监控中的网络结构并以此来分析视频监控活动,该网络中节点由目标出现区域和消失区域构成,边由目标运动路径构成(摄像机内部或者不同摄像机之间).Tieu 等[9]假设目标在摄像机节点之间的转移时间满足一定的概率分布,并依据非参数贝叶斯框架来估计相机之间的拓扑关系.Loy 等[10]在每一个摄像机的可视范围内划定不同的语义区域,然后通过分析目标在不同区域内的活动来进行不同摄像机之间的目标关联分析.Wang 等[11]利用摄像机节点的网络图模型来进行目标运动轨迹分析,该模型不需要训练数据,并且能够自适应地进行调整.刘少华等[12]提出了一种基于最小费用流模型的关联算法,能够估算并修正效用函数的偏差.万九卿等[13]提出了高阶时空观测模型并在此基础上建立了数据关联问题的动态贝叶斯网络描述,该算法能够降低运算量,并且能够应用于目标外观模型未知的情况.

1.3 本文的贡献

本文的主要贡献有 3 点:①与其他只研究摄像机网络拓扑的方法不同,本文将特征融合与时空关联相结合来进行多摄像机监控环境下的目标跟踪,这样就可以充分利用现有的各种优秀的特征提取方法并将其融入提出的算法框架;②本文采用的是区域协方差算子(RC,Region Covariance descriptor)对目标外观特征进行融合,实验表明,该方法在噪声和尺度上具有较好的鲁棒性;③在目标关联阶段,采用了针对多个跟踪序列的

二部图匹配策略来解决多摄像头目标跟踪中的重识别与数据关联问题,相比于传统的匹配方法,可以降低匹配的错误率,修正误匹配.

2　时空关联图模型的建立

本文利用有向带权图模型来表示摄像机网络的空间拓扑结构和目标运动的时间关联性,并利用其进行多摄像头监控下的目标跟踪. 摄像机网络的拓扑结构表示目标在摄像机间的转移关系,如图 1 所示,将每个摄像机视域 P 中目标出现区域和消失区域定义为节点集合 V,节点之间的边集合 E 表示目标在相邻两个节点之间在空间物理位置一步可达,于是图模型可定义为:$G=\langle V,E\rangle$,$v_i\in V$,$e(i,j)\in E$,$1\leqslant i,j\leqslant K$,$v_i$ 表示图中节点,$e(i,j)$ 表示节点 v_i 与节点 v_j 的边,同一摄像机视域内的节点由内部邻边相连(intra-camera),不同摄像机视域间的节点由外部邻边相连(inter-camera),K 为图中的节点个数.

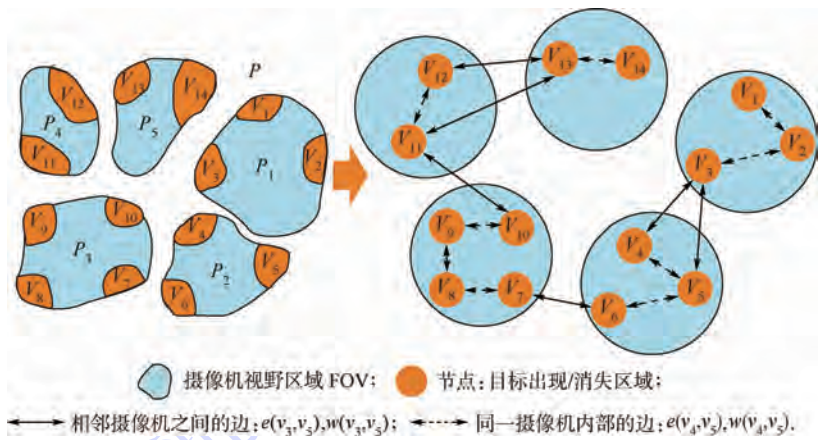


图 1　摄像机网络与图模型建立示意图

Fig. 1　Example of camera network and its corresponding graph model

2.1　基于网络拓扑的空间关联

空间关系定义了节点之间是否建立边,以及边的方向. 图模型的建立过程中,如果两个节点之间在物理空间位置上一步可达,即不经过其他任何节点,则为它们之间建立一条边,边的方向定义为目标移动的方向. 图 1 中有 5 个摄像机,其视域(FOV)分别为 $P_1\sim P_5$,$v_i\in V$ 表示目标在 FOV 中的出现与消失区域,同一摄像机内部的目标运动轨迹用虚线箭头表示,不同摄像机之间的目标移动轨迹用实线箭头表示. 这样的拓扑关系建立,能够满足跨摄像机下目标跟踪过程中的两个约束:

约束 1　如果跟踪目标在当前摄像机区域中的节点处能够监测到,则使用单摄像头跟踪算法对其进行持续跟踪,这时节点之间不需要进行目标关联计算,可将跟踪匹配时的不确定性降到最低.

约束 2　如果跟踪目标在某一节点处消失,则在图模型中使用广度优先策略(BFS, Breadth First Search)在其邻域(实线箭头)一步可达的节点处进行目标检测,对检测到的目标提取特征融合,再进行目标的匹配与关联.

2.2　基于统计学习的时间关联

图模型中边的权重由目标在不同节点之间转移的概率决定,在实际应用系统中,采用统计学习方法为目标的运动建立时间约束,从而定义节点之间合理的权值 $w(v_i,v_j)$,$v_i,v_j\in V$. 在实验过程中,首先对目标通过相邻 v_i 和 v_j 之间的时间进行统计,可以得到一组观测时间序列 $T_{ij}=(t_1,t_2,\cdots,t_n)$,该观测序列的分布可以由高斯分布进行描述,也就是说节点 v_i 和 v_j 之间的边 $e(i,j)$ 可以根据高斯分布 $N(e_{ij}^t,\sigma_{ij}^2)$ 进行建模. 通过训练数据,就可以由该分布的期望与方差来获得这条边的权重.

实际情况下,想要获取一组观测序列的统计规律往往比较困难,这是由许多因素决定的:包括目标的运动规律、摄像机的地理位置、每天的监控环境变化等. 本文中,利用事先获得的训练数据(如各个节点之间的通过时间),来近似地估计每条边的权重. 在计算某一条边权时,对所有观测时间进行聚类并计算每个类内的方差. 具体的操作过程中,σ_{ij} 是由当前边属于哪一个类而决定的. 图 2 是两个节点之间边权的直方图实例,从中可以看到,观测数据可以被划分为 3 个类别,每个类别中可以得到相应的方差值. 本文使用的聚类方法为 K-means,该算法可以自适应地决定聚类类别的数量.

文献[14]在求取图模型中最优边权时采用了一个期望权重相似度方程(EWS, Expected Weighted Similarity),本文受其启发,定义图模型

的边权重为

$$w(e) \triangleq \exp(e) - 1 \tag{1}$$

该权重方程能够为相应的边定义相似度最大的边权得分，如果 $p(e) \sim N(e', \sigma)$，定义该边的能量为

$$E[w(e)] = \exp\left(e' + \frac{\sigma^2}{2}\right) - 1 \tag{2}$$

使用 EWS 方程定义的边权能量值既考虑了观测数据的均值也考虑了数据的方差.

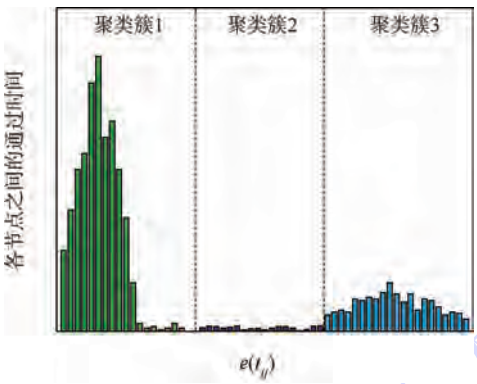


图 2 通过训练数据学习而来的节点 v_i 和 v_j 之间边权统计直方图

Fig. 2 Histogram of edge weights between v_i and v_j learned during training phase

3 数据关联与相似性度量

3.1 特征提取与相似性度量

本节具体阐述了不同节点处观测到的目标相似性度量方法. O_n^i 表示第 i 个节点处观测到的目标 n. 对于两个不同节点 v_i 和 v_j 处的两个目标关联性大小，本文使用下式进行度量：

$$\mathrm{Cor}(O_n^i, O_m^j) = \alpha \cdot \mathrm{Sim}(O_n^i, O_m^j) + (1 - \alpha)E[w(e)] \tag{3}$$

式中，$\mathrm{Sim}(O_n^i, O_m^j)$ 为目标表现特征项（AM），本文由两个目标区域的协方特征矩阵的相似性来度量，具体定义见下文. $w(v_i, v_j)$ 为目标时空关联项（GM，Graph Model），定义为两个节点之间的边权. 对于算法中的参数，将设定一个最优化的经验范围，具体见本文实验部分.

Tuzel 等[15]在 2006 年将区域协方差描述子引入到目标特征匹配中去，这种基于图像二阶统计特性的协方差矩阵描述子（RC，Region Covariance）在目标检测与跟踪领域都得到了应用并且取得了良好的效果. 为了提高跟踪的可靠性，本文使用 RC 来实现目标多特征的融合，对目标的外观表现模型进行表示. 跟踪目标的每个像素点 k 生成一个 d 维的特征向量 $\boldsymbol{z}_k$，该特征向量可以是像素的颜色、灰度、亮度、梯度等信息. 本文中特征向量采用的是 16 位 RGB 颜色空间下均匀量化的直方图[16]与 SURF 特征[17].

对于一个特定的大小为 $W \times H$ 的矩形区域 R，使用 $\{\boldsymbol{z}_k\}_{k=1,2,\cdots,s,s=W\times H}$ 来表示 R 的 d 维的特征向量，并且使用以下的协方差矩阵来表示目标区域 R：

$$\boldsymbol{C}_R = \frac{1}{s-1}\sum_{k=1}^{s}(\boldsymbol{z}_k - \boldsymbol{\mu})(\boldsymbol{z}_k - \boldsymbol{\mu})^{\mathrm{T}} \tag{4}$$

式中，$\boldsymbol{\mu} = \frac{1}{s-1}\sum_{k=1}^{s}\boldsymbol{z}_k$，表示所有的点的均值化.

式(4)中，正定的实对称矩阵 $\boldsymbol{C}_R$ 由于在计算中本身就已经进行了去均值化，因此它对目标区域的亮度均值变化具有不变性. 在目标识别的应用中，尺度不变性、旋转不变性和光照不变性具有非常重要的意义. 在对运动目标进行提取与协方差矩阵表示之后，需要采用一定的相似性度量算法进行特征匹配，从而判定是否为同一目标. 文献[18]给出了一种协方差矩阵之间的距离度量方法，对于两个协方差矩阵 $\boldsymbol{C}_1$ 和 $\boldsymbol{C}_2$，其相关性可表示为

$$\mathrm{Sim}(\boldsymbol{C}_1, \boldsymbol{C}_2) = \sqrt{\sum_{i=1}^{s}\ln^2\lambda_i(\boldsymbol{C}_1, \boldsymbol{C}_2)} \tag{5}$$

式中，$\{\lambda_i(\boldsymbol{C}_1, \boldsymbol{C}_2)\}_{i=1,2,\cdots,s}$ 是 $\det(\lambda\boldsymbol{C}_1 - \boldsymbol{C}_2) = 0$ 中 λ 的解.

$$\lambda\boldsymbol{C}_1 - \boldsymbol{C}_2 = \sqrt{\boldsymbol{C}_1}\left(\lambda\boldsymbol{E} - \sqrt{\boldsymbol{C}_1^{-1}}\boldsymbol{C}_2\sqrt{\boldsymbol{C}_1^{-1}}\right)\sqrt{\boldsymbol{C}_1} \tag{6}$$

式中，正定的实对称阵 $\sqrt{\boldsymbol{C}_1^{-1}}\boldsymbol{C}_2\sqrt{\boldsymbol{C}_1^{-1}}$ 的特征值是正实数，这样 $\{\lambda_i(\boldsymbol{C}_1, \boldsymbol{C}_2)\}_{i=1,2,\cdots,s}$ 的计算即可以转化为对矩阵 $\sqrt{\boldsymbol{C}_1^{-1}}\boldsymbol{C}_2\sqrt{\boldsymbol{C}_1^{-1}}$ 的特征值的求解；$\boldsymbol{E}$ 为单位矩阵.

3.2 基于二部图匹配策略的目标关联

当目标在摄像机之间移动时，需要判断其最有可能出现的下一个节点，这是一个图像匹配的问题. 传统的基于单幅图像匹配的重识别框架并不适用于跨摄像头跟踪任务，这是由于在某个摄像头下进行目标检测时有可能出现第一帧就检测错误的情况，而且由于监控环境的复杂性，跟踪目标经常会由于遮挡而出现跟踪框漂移的情况，为了避免上述两种情况可能导致的跟踪失败，本文采用了基于两个观测序列帧匹配的策略，这样可以降低匹配的错误率，修正误匹配. 具体方法为：需要将目标消失节点 v_i 处的观测序列 O^i：$\{O_1^i, O_2^i, \cdots, O_n^i\}$ 与目标可能出现的节点 v_j 处的观测序列 O^j：$\{O_1^j, O_2^j, \cdots, O_m^j\}$ 建立关联关系，具体如图 3 所示. 注意，依据之前建立的时空关联图模型，此处的 v_j 应为 v_i 一步可达的所有节点. 为了求解一

个最优的 $O^i \mapsto O^j$ 关联匹配,目标关联性问题可以转化为一个二部图匹配问题.

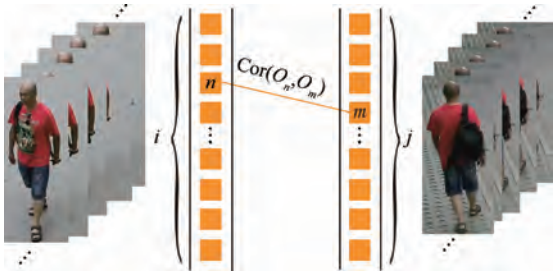


图 3　视频帧序列的二部图匹配

Fig. 3　Bipartite frame sequence matching

为了将问题形式化,本文定义两个矩阵:$\{\mathrm{Cor}(O_n^i, O_m^j)\}_{N\times M}$ 为得分矩阵,矩阵元素 $\mathrm{Cor}(O_n^i, O_m^j)$ 表示节点 v_i 中的检测目标 O_n^i 与节点 v_j 中的检测目标 O_m^j 相关联时产生的得分值,其中 $N = |O^i|, M = |O^j|$. $\{\boldsymbol{X}(O_n^i, O_m^j)\}_{N\times M}$ 为 0-1 二值矩阵,如果目标 O_n^i 与目标 O_m^j 相匹配,则 $\boldsymbol{X}(O_n^i, O_m^j) = 1$. 特别地,这里需要定义一个约束:$\sum_{m=1}^{M} \boldsymbol{X}(O_n^i, O_m^j) = 1, \forall n \in N$,其作用是保证二值矩阵 $\{\boldsymbol{X}(O_n^i, O_m^j)\}_{N\times M}$ 中每一行有且仅有一个 1 值,其余均为 0 值,即保证了 O^i 中所有观测对象在 O^j 中有且仅有唯一的对应. 于是问题的目标方程即可表示为求解一个最大权重的完备匹配:

$$\hat{M} = \mathrm{argmax} \sum_{n=1}^{N} \sum_{m=1}^{M} \mathrm{Cor}(O_n^i, O_m^j) \cdot \boldsymbol{X}(O_n^i, O_m^j) \tag{7}$$

在解决二部图最大匹配问题的众多策略之中,匈牙利算法[19]是最为经典的算法之一,它可以在多项式时间内解决问题,该算法的复杂度为 $O(N^3)$,其中 N 为观测帧序列 O^i 中的目标个数. 本文使用匈牙利算法对式(7)进行求解,即可得到一个 O^i 到 O^j 的最大权完备匹配,并得到两个节点包含的观测序列的总体目标关联度 $M(v_i, v_j)$,这里需要设定一个阈值为 0.5(见实验设计),当关联度大于该阈值时,认为是一个可信的匹配,否则认为该目标已经离开所有监控摄像机的可视范围. 通过对节点 v_i 所有一阶邻域的节点的关联度 M 进行排序,即可得到目标最有可能的出现节点 v_j:

$$v_j = \mathrm{argmax}_{j\in\{j\,|\,\exists e(v_i,v_j),\,\mathrm{s.t.}\,v_i\to\mathrm{next}=v_j\}} M(v_i, v_j) \tag{8}$$

4　实　验

4.1　实验设计

对多摄像机目标跟踪方法的研究一直都缺少相关的实验数据以及实验系统. 本实验使用的数据来源于 PKU-SVD-B 数据集[20],本文建立如图 4 所示的摄像机监控网络,用以模拟复杂监控网络中的若干单元,4 个摄像机监控视域无重叠,采集视频在时间上保持同步,在每个摄像机独立监控的范围内首先需要对移动目标进行检测和分割[3],这里采用帧差法获取目标区域,然后采用 TLD[6]算法在当前视域内进行跟踪.

图 4　摄像机视域及网络拓扑

Fig. 4　Topology of camera views

如图 4 所示,所建立的图模型包含 4 个摄像机区域,9 个节点,对于每个节点在时间轴上设置 F_i ($i = 1, 2, \cdots, N$)个观测帧序列,每个观测帧序列在时间轴上具有一定的范围(例如前后10 帧). 如果指定对象的运动轨迹依据图模型转移到该节点,则认为这条轨迹包含了对应的节点. 对于某一个指定跟踪目标,在某个包含的节点范围内,当所有匹配帧和 Ground Truth 有超过 50% 的帧能够正确匹配上,则认为该跟踪对象在跨摄像头跟踪时得到了一个可信的匹配节点. 对某一跟踪目标的评价指标为

$$T_{\mathrm{rec}} = \frac{\sum_{i=1}^{P} C_i}{\sum_{i=1}^{P} G_i} \tag{9}$$

$$T_{\mathrm{pre}} = \frac{\sum_{i=1}^{P} C_i}{\sum_{i=1}^{P} D_i} \tag{10}$$

$$F_1 = \frac{2 \cdot T_{\mathrm{pre}} \cdot T_{\mathrm{rec}}}{T_{\mathrm{pre}} + T_{\mathrm{rec}}} \tag{11}$$

式中,T_{rec} 和 T_{pre} 分别表示召回率和准确率;P 为总的节点个数;G_i ($G_i < F_i$) 为跟踪对象的 Ground Truth 轨迹序列在第 i 个节点处的数目;C_i 为跟踪对象在节点 i 处正确匹配到的帧的数目;D_i 为跟

北航学报 赠阅

踪对象在节点 i 处包含的所有匹配帧数.

4.2 结果与分析

实验中,采集 5 个行人在摄像机网络中的运动轨迹,图 5 为他们在摄像机中的跟踪实例,在这些实例中,由于摄像机的拍摄角度,目标出现的运动轨迹经常是从远到近或是由近及远,因此单摄像头目标跟踪算法应该对目标的形变有较好的鲁棒性.

图 5 5 个目标跟踪实例

Fig. 5 Five examples of image of people in network

为了更好地评估算法性能,同时保证衡量标准的公正性,实验确保每种方法均使用相同的参数在全部 5 个跟踪实例上进行实验. 对于式(3)中的参数 α,根据 Sim(·)项和 $E[w(e)]$项的权重,首先,通过手工调整参数过程中确定的算法最佳性能的范围为 0.05 ~ 0.15. 然后,在对每个对象进行跟踪时,在经验范围内随机选择 10 组参数进行实验,图 6 为每个跟踪实例的 10 组随机参数的结果得到的箱型图,即每个箱型柱由 10 组数据的分布得到. 从中可以展示出每个实例的平均 F_1 得分、上边缘值和下边缘值等,而且能更为细致地体现不同参数情况下算法性能的差异,大部分数据集中在平均值附近,说明在 α 的经验范围内算法性能比较稳定.

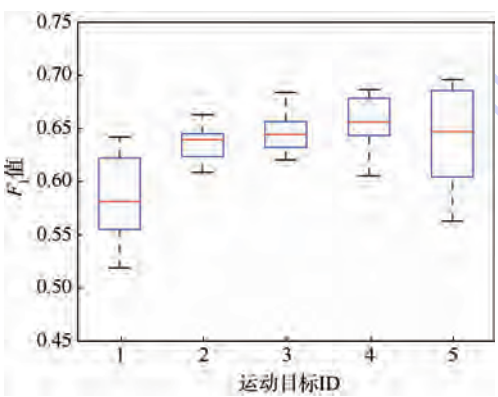


图 6 在不同跟踪目标上的随机参数实验

Fig. 6 Experiment of random parameters in different tracking cases

对比实验中,算法使用的最优参数为 5 组实例的平均 F_1 得分最大情况下的参数. 在表 1 中列出了 5 组实例中目标最终轨迹的准确率、召回率和 F_1 得分. 本文结合目标表现模型与图模型时空关联的方法相比只使用表现模型的基线方法在跟踪的精度上能够得到提升(约 10% 的提升). 传统的基于特征的识别方法没有加入时空关联约束,大部分目标由于在外观上难以分辨,会产生较多误匹配,因此准确率往往较低.

表 1 目标跟踪精度及对比

Table 1 Tracking accuracy and comparison

目标 ID	方法对比	准确率	召回率	F_1 值
1	AM + GM	0.595 6	0.600 6	0.598 0
	AM	0.492 3	0.495 0	0.493 6
2	AM + GM	0.655 9	0.626 4	0.640 8
	AM	0.501 9	0.599 6	0.546 4
3	AM + GM	0.651 9	0.659 2	0.655 5
	AM	0.561 1	0.604 0	0.581 7
4	AM + GM	0.653 7	0.659 0	0.656 3
	AM	0.558 0	0.597 6	0.577 1
5	AM + GM	0.602 1	0.687 8	0.642 1
	AM	0.505 6	0.559 0	0.530 9

在目标识别、跟踪的应用中,好的算法应该对环境的变化有较强的鲁棒性. 图 7 中比较了只依赖 AM 跟踪与 AM + GM 方法在噪声与目标尺度变化方面的鲁棒性结果. 使用高斯噪声对检测的

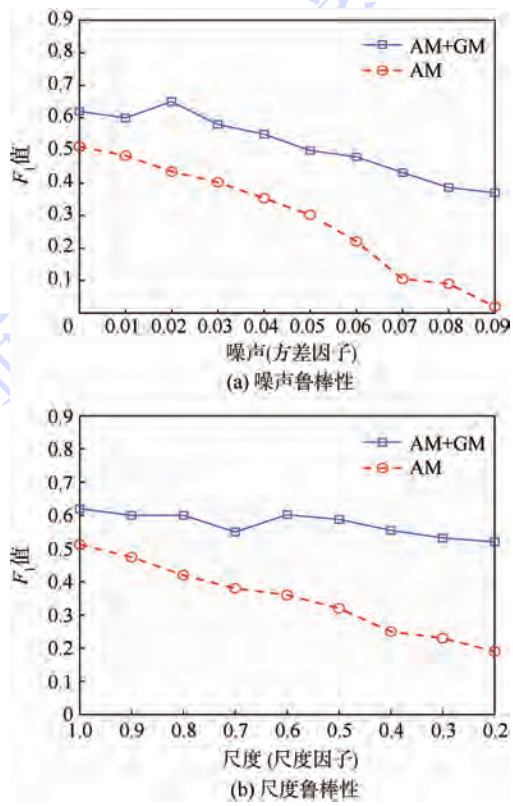


图 7 跨摄像头跟踪时不同方法对噪声和尺度的鲁棒性

Fig. 7 Robustness of different approaches to noise and scale variances in object tracking

北航学报 赠阅

帧序列进行模糊处理,可以清晰地发现随着噪声的逐渐增大,AM 方法会逐渐失效,而加入了时空关联图模型辅助的方法则对噪声有一定的鲁棒性.与此同时,监控视频中的运动目标由于远离或者靠近摄像机,往往会产生较大幅度的尺度改变,实验表明,本文提出的方法对于跟踪目标的尺度大小也具有很好的鲁棒性.

5　结　论

1)本文提出了一种将特征融合与时空关联相结合来进行多摄像机监控环境下的目标跟踪方法.时空关联图模型能够为目标跟踪过程中的目标关联提供可靠的依据.

2)实验表明,本文的方法在实际应用中有着较好的适用性,相比于传统的基于特征模型跟踪的方法在跟踪精度方面有较大提高.

3)由于多摄像机网络的复杂性,目标移动过程中的时间延迟等因素,如何将算法进一步地扩大到更大范围的应用场景以及更多元的目标种类将是下一步的研究工作.

参考文献(References)

[1] Wang X. Intelligent multi-camera video surveillance: a review [C]//Pattern Recognition Letters. Netherlands: Academic Press, Elsevier, 2013, 3(4): 3-19.

[2] Fu H, Ma H, Ming A. EGMM: an enhanced Gaussian mixture model for detecting moving objects with intermittent stops[C]// IEEE International Conference on Multimedia and Expo. Piscataway, NJ: IEEE, 2011, 7041(2): 1-6.

[3] Fu H, Ma H, Liu L. Robust human detection with low energy consumption in visual sensor network[C]//IEEE International Conference on Mobile Ad-hoc and Sensor Networks. Piscataway, NJ: IEEE, 2011: 91-97.

[4] Zheng W S, Gong S, Xiang T. Reidentification by relative distance comparison[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(3): 653-668.

[5] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark [C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2013, 9(4): 2411-2418.

[6] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.

[7] Makris D, Ellis T, Black J. Bridging the gaps between cameras [C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2004, 2: II205-II210.

[8] Stauffer C. Estimating tracking sources and sinks[C]//IEEE Conference on Computer Vision and Pattern Recognition Workshop. Piscataway, NJ: IEEE, 2003: 35.

[9] Tieu K, Dalley G, Grimson W E L. Inference of non-overlapping camera network topology by measuring statistical dependence [C]//IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2005, 2: 1842-1849.

[10] Loy C C, Xiang T, Gong S. Multi-camera activity correlation analysis[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2009: 1988-1995.

[11] Wang X, Tieu K, Grimson W E L. Correspondence-free activity analysis and scene modeling in multiple camera views[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(1): 56-71.

[12] 刘少华,赖世铭,张茂军.基于最小费用流模型的无重叠视域多摄像机目标关联算法[J].自动化学报,2010,36(10):1484-1489.
Liu S H, Lai S M, Zhang M J. A min-cost flow based algorithm for objects association of multiple non-overlapping cameras[J]. ACTA Automatica Sinica, 2010, 36(10): 1484-1489 (in Chinese).

[13] 万九卿,刘青云.基于高阶时空模型的视觉传感网络数据关联方法[J].自动化学报,2012,38(2):236-247.
Wan J Q, Liu Q Y. Data association in visual sensor networks based on high-order spatio-temporal model[J]. Acta Automatica Sinica, 2012, 38(2): 236-247 (in Chinese).

[14] Song B, Roy-Chowdhury A K. Stochastic adaptive tracking in a camera network[C]//IEEE 11th International Conference on Computer Vision. Piscataway, NJ: IEEE, 2007: 1-8.

[15] Tuzel O, Porikli F, Meer P. Region covariance: a fast descriptor for detection and classification[C]//European Conference on Computer Vision. Heidelberg: Springer, 2006: 589-600.

[16] Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]//European Conference on Computer Vision. Heidelberg: Springer, 2008: 262-275.

[17] Bay H, Tuytelaars T, Van Gool L. Surf: speeded up robust features[C]//European Conference on Computer Vision. Heidelberg: Springer, 2006: 404-417.

[18] Förstner W, Moonen B. A metric for covariance matrices[C]// Geodesy-the Challenge of the 3rd Millennium. Heidelberg: Springer, 2003: 299-309.

[19] Kuhn H W. Variants of the Hungarian method for assignment problems[J]. Naval Research Logistics Quarterly, 1956, 3(4): 253-258.

[20] Large-scale object tracking over a multiple-camera network [EB/OL]. The First IEEE International Conference on Multimedia Big Data. [2014-04-20]. http://www.bigmm2015.org/Challenge.asp.

北航学报 赠阅

Object tracking in surveillance videos using spatial-temporal correlation graph model

ZHANG Cheng, MA Huadong*, FU Huiyuan

(Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Object tracking in non-overlapping multi-camera surveillance is a challenging problem since the transition time between cameras varies greatly from individual to individual with uncertainty. The key problem of object tracking in wide areas is data association and how to find correspondences between objects via camera topology. A novel graph model was proposed to capture the spatial-temporal correlation among objects, which are moving in the camera network. Source/sink regions are graph nodes, and the graph edges are constructed by the spatial and temporal constrains. Specifically, a tracking method combining appearance model and graph model was proposed to solve the problem of object re-identification and data association via bipartite matching in multi-camera object tracking. In addition, region covariance descriptor was utilized to fuse the appearance feature. Experiments with real videos validate the proposed approach.

Key words: object tracking; multi-camera network; spatial-temporal graph model; bipartite matching; data association

北京航空航天大学学报

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0478

抗代间污染攻击的网络编码签名方案

彭天丽，尚涛*，刘建伟

（北京航空航天大学 电子信息工程学院，北京 100191）

摘 要：为适应实时应用的需求，网络编码中引入了代的概念. 针对网络编码易受代间污染攻击导致消息发生串扰的问题，提出了基于代标识符的网络编码签名方案. 首先，方案中设置了依据代标识符生成的两级私钥，包括代私钥以及在此基础扩展成的消息私钥，以便节点判断消息的代属性来决定是否对它继续进行编码；其次，利用双线性对构造了具有同态性质的签名算法，通过分离两级私钥进行批验证，节点可同时验证同一代的所有消息. 最后，通过随机预言模型分析，证明了该方案在适应性选择消息攻击下是安全的. 通过开销分析表明方案能有效减少验证开销，提高系统性能.

关 键 词：网络编码；代间污染攻击；同态签名；批验证；双线性对
中图分类号：TN918.1
文献标识码：A **文章编号**：1001-5965(2015)04-0721-06

网络编码[1]可使组播网络中的信息传输达到最大流最小割这一理论上界，也可提高网络鲁棒性和安全性，所以在实际网络中得到了广泛的应用. 但是网络编码在带来收益的同时存在许多的安全传输问题[2]. 其中一类主要安全问题就是网络编码容易受到恶意节点的污染攻击[3]，导致目的节点无法正确解码. 对于网络中的攻击者篡改网络中传输的数据或向网络中输入随机数据来干扰网络通信的污染攻击，Yu等[4]提出了基于RSA的同态签名方案，裴恒利等[5]在基于RSA同态签名方案的基础上设计了一种融合时间戳和同态签名的网络编码签名方案，它可同时抵御污染攻击与重放攻击，随后Shang等[6]在格上通过计算消息与其签名的距离提出了一种基于距离的安全网络编码算法，它的安全性可规约为格上固定长度向量问题(FLVP)，同时格的特性使得它比基于RSA的签名方案更加安全. Chou等[7]提出了应用于实际无线网络中的分代网络编码方法，它能更好地满足信息实时传输的需求，然而这样的分代网络编码不仅面临代内污染攻击问题也存在着代间污染攻击问题. 上述已有方案可以有效地解决代内污染攻击的问题，然而对于不同消息子空间的消息发生串扰引起的代间污染攻击问题并未加以讨论. 针对这个问题，Zhao等[8]与Kehdi等[9]提出了验证每个消息向量是否属于源消息向量张成的线性子空间的方法，方法主要思想是通过生成源消息张成线性子空间的零空间，由于零空间的任意向量与传输的消息向量正交，因而给中间节点分发零空间的消息向量即可验证消息的完整性. 然而，此方法一方面需预先知道整个文件信息，一方面公钥信息需

收稿日期：2014-07-30；**录用日期**：2014-10-11；**网络出版时间**：2014-12-22 10:41
网络出版地址：www.cnki.net/kcms/detail/11.2625.V.20141222.1041.001.html
基金项目：国家973计划资助项目(2012CB315905)；国家自然科学基金资助项目(61272501)；Research Promotion Grant-in-Aid for KUT Graduates of Special Scholarship Program；中央高校基本科研业务费资助项目(YWF14DZXY012)
作者简介：彭天丽(1991—)，女，重庆人，硕士生，ptianli@aliyun.com
*通讯作者：尚涛(1976—)，男，辽宁营口人，副教授，shangtao@buaa.edu.cn，研究方向为网络安全、网络编码.

引用格式：*彭天丽，尚涛，刘建伟. 抗代间污染攻击的网络编码签名方案[J]. 北京航空航天大学学报，2015，41(4)：721-726.*
Peng T L，Shang T，Liu J W. Signature scheme for network coding against inter-generation pollution attacks[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：721-726 (in Chinese).

北航学报 赠阅

频繁更新. 接着 Boneh 等[10]提出了一种利用双线性对构造的只需常数级规模公钥的网络编码签名方案,它通过对线性子空间进行签名,可抵御代内和代间污染并在随机预言模型中被证明可满足 CDH 假设.

针对代间污染攻击问题并思及 Boneh 方案在验证过程中需对每个接收的消息执行双线性对操作带来了很大的网络开销,本文受到 Tseng 等[11]的启发,将批验证[12]与网络编码相结合,提出了一种结合批验证的抗代间污染攻击的同态网络编码签名方案. 该方案一方面能够抵御代间污染攻击,另一方面与 Boneh 等的方案相比提高了节点验证效率,减少了系统开销.

1 相关研究

1.1 随机线性网络编码

随机线性网络编码的基本思想是,编码时将所有消息进行线性组合,所采用的系数均为随机选取;解码时,采用高斯消元法求解线性方程组从而恢复出原始消息[13].

随机线性网络编码的具体步骤如下:

1) 网络模型如图 1,一个源节点 S 将要传输的消息分为 l 代,每代拥有 m 条消息,将这 m 条消息发送到 t 个目的节点 $d_1, d_2, \cdots, d_t$ 时,源节点首先将每条原始消息 $\boldsymbol{M}_i (i = 1, 2, \cdots, m)$ 设定为选自有限域 $\mathbf{Z}_q$ 的长度为 n 的向量,则原始消息 $\boldsymbol{M}_i$ 可表示为 $\boldsymbol{M}_i = (v_{i1}, v_{i2}, \cdots, v_{in}), v_{ij} \in \mathbf{Z}_q$.

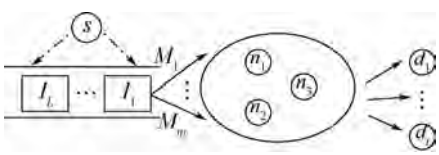


图 1 网络模型图

Fig. 1 Network model

然后将每条原始消息向量用长度为 m 的单位向量进行扩展生成扩展向量[14] $\boldsymbol{M}'_i$:

$$\boldsymbol{M}'_i = (\underbrace{0, \cdots, 0}_{i-1}, 1, \underbrace{0, \cdots, 0}_{m-i}, v_{i1}, v_{i2}, \cdots, v_{in})$$

对 m 个扩展向量进行随机网络编码,编码后向量为 $\boldsymbol{w} = \sum_{i=1}^{m} a_i \boldsymbol{M}'_i, \boldsymbol{w} \in \mathbf{Z}_q^N, N = m + n$,其中 a_i 是从有限域中随机选取的. 将代标识符 I 随着消息一起传输,则编码后消息格式如图 2 所示.

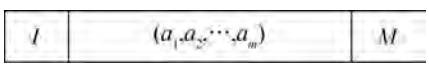


图 2 编码后的消息格式

Fig. 2 Message format after network coding

2) 中继节点收到 k 个消息向量 $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_k$,将它们进行线性组合,得新的编码向量 $\boldsymbol{W} = \sum_{i=1}^{k} a_i \boldsymbol{w}_i \in \mathbf{Z}_q^N$.

3) 目的节点收到 m 个线性无关的消息向量后,可得到一个 m 行 $m + n$ 列的矩阵,将其前 m 列构成的矩阵记为 $\boldsymbol{C}$,后 n 列构成的矩阵记为 $\boldsymbol{P}$,则可由式(1)恢复得到原消息.

$$(\boldsymbol{M}_1, \boldsymbol{M}_2, \cdots, \boldsymbol{M}_m) = \boldsymbol{C}^{-1}\boldsymbol{P} \tag{1}$$

1.2 网络编码代间污染攻击

代标识符的主要作用就是区分不同的编码向量所张成的消息子空间,如果不同消息子空间的消息发生了串扰就产生了代间污染. 造成代间污染有两种攻击方式[15],具体代间污染攻击模型如图 3 所示. 在模型中,n_1 和 n_2 是普通节点,n_3 和 n_4 是恶意节点,n_5 是污染节点.

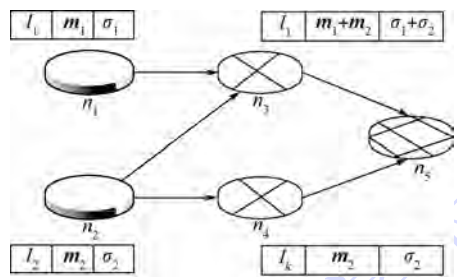


图 3 代间污染攻击模型

Fig. 3 Model of inter-generation pollution attack

一种攻击方式是对消息的攻击. 节点 n_1 输出代 I_1 的消息,节点 n_2 输出代 I_2 的消息. 当恶意节点 n_3 收到来自代 I_1 的消息 $\boldsymbol{m}_1$ 和代 I_2 的消息 $\boldsymbol{m}_2$ 时,将其编码得到消息组合 $\boldsymbol{m}' = \boldsymbol{m}_1 + \boldsymbol{m}_2$,通过同态签名,可得到消息 $\boldsymbol{m}'$的合法签名 $\sigma' = \sigma_1 + \sigma_2$. 但是由于新产生的消息组合 $\boldsymbol{m}'$既不属于代 I_1 也不属于代 I_2,所以该消息属于污染消息,当其参与目的节点解码时将导致解码错误.

另一种攻击方式是对代标识符的攻击. 恶意节点 n_4 转发代 I_2 消息 $\boldsymbol{m}_2$ 到节点 n_5 的过程中,随意伪造或篡改 I 信息,消息将指向错误的消息子空间,则节点 n_5 也会受到污染.

上述两种攻击方式均会导致线性子空间的消息发生串扰,使目的节点无法正确解码.

1.3 双线性对

相对于有限域乘法群上的离散对数问题,椭圆曲线加法群上离散对数问题的求解更难,它可用更小的密钥保证更高级别的安全性. 所以在 Boneh 等[16]借助椭圆曲线上的双线性对给出了基于身份的加密体制后,双线性对被用来构造各种密码协议.

北航学报 赠阅

令 G_1 是阶为大素数 q 的加法循环群，生成元为 P. 令 G_2 是阶为 q 的乘法循环群. 假设 G_1，G_2 上的离散对数问题都是困难的. 若映射 $e: G_1 \times G_1 \rightarrow G_2$ 满足如下性质则为双线性对.

1）双线性性：对于任意的 $P, Q \in G_1$，$a, b \in \boldsymbol{Z}_q^*$，都有 $e(aP, bQ) = e(P, Q)^{ab}$.

2）非退化性：存在 $P, Q \in G_1$ 使得 $e(P, Q) \neq 1$.

3）可计算性：对于任意的 $P, Q \in G_1$，存在多项式时间算法能够计算 $e(P, Q)$.

对双线性映射 e 的性质 1）进行推导可得如式(2)的等式：

$$e(aP, bQ) = e(P, Q)^{ab} = e(P, abQ) = e(abP, Q) \tag{2}$$

2　签名方案

2.1　定　义

网络编码同态签名方案由下列四元组概率多项式时间算法构成.

1）系统参数建立(Setup). 输入安全参数 1^τ，输出素数 q、源节点公钥以及私钥.

2）签名(Sign). 输入私钥，一个随机选取的代标识符 $I \in \{0,1\}^*$，一个消息向量 $\boldsymbol{w}$，输出消息的签名 σ.

3）消息组合(Combine). 输入一个代标识符，k 个消息向量 $\boldsymbol{w}_i$，k 个随机编码系数 $a_i \in \boldsymbol{Z}_q$，输出消息组合 $\boldsymbol{W}$ 及其签名 σ.

$$\boldsymbol{W} = \sum_{i=1}^{k} a_i \boldsymbol{w}_i \in \boldsymbol{Z}_q^N \tag{3}$$

4）批验证(Batch Verify). 输入一组待验证的消息向量 $\boldsymbol{w}_i$，公钥和当前传输的代标识符 $I \in \{0,1\}^*$，经验证算法输出有效或者无效.

2.2　具体构造

对于四元组算法的具体构造如下.

1）系统参数建立(Setup)：给定系统参数 1^τ，系统选取参数 G_1, G_2, P, q, e, H_G，其中 H_G 是一个单向抗碰撞 hash 函数 $H_G: \{0,1\}^* \rightarrow G_1$. 然后系统选取随机数 $\alpha \in \boldsymbol{Z}_q^*$ 和 $r_i \in \boldsymbol{Z}_q^*$ $(i = 1, 2, \cdots, N)$ 作为系统私钥. 计算 $D_i = \alpha r_i \cdot P$ $(i = 1, 2, \cdots, N)$ 作为系统公钥. 假设源节点要发送 l 个代的消息，那么给每个代随机选取一个代标识符 I，计算 $G_{pk} = H_G(I)$. 源节点公开公共参数 $\{G_1, G_2, e, q, P, \{D_i\}_{i=1}^N, G_{pk}, H_G\}$.

2）签名(Sign)：给定代 I 的一组扩展后的消息向量 $\boldsymbol{M}' = (\boldsymbol{M}_1', \boldsymbol{M}_2', \cdots, \boldsymbol{M}_m')$，其中 $\boldsymbol{M}_i' = (v_1, v_2, \cdots, v_N)$. 计算 $G_{sk} = \alpha \cdot G_{pk}$.

计算签名为

$$\sigma = \sum_{i=1}^{N} G_{sk} \cdot r_i v_i \bmod q \tag{4}$$

式中，G_{sk} 可看成代对应的私钥，$G_{sk} \cdot r_i$ 可看成消息对应的私钥. 所以同一代的消息对应着相同的代私钥和消息私钥.

3）消息组合(Combine)：当接收到 k 个属于代 I 的消息向量后，节点对其进行随机线性网络编码操作得到新的消息向量为 $\boldsymbol{W} = \sum_{i=1}^{k} a_i \boldsymbol{w}_i$，并利用签名的同态性质，由 k 条消息组合中的签名直接生成新消息对应的签名 $\mathrm{sign}(\boldsymbol{W})$.

性质 1　方案具有同态性质.

证明　设 $\boldsymbol{w}_i = (v_{i1}, v_{i2}, \cdots, v_{iN})$ $(i = 1, 2, \cdots, k)$ 是长度为 N 的向量，则由式(4)可得：

$$\mathrm{sign}(\boldsymbol{w}_i) = \sum_{i=1}^{N} G_{sk} \cdot r_i v_i \bmod q$$

那么中间节点收到 k 个消息向量后，进行随机线性网络编码得到向量 $\boldsymbol{W} = \sum_{i=1}^{k} a_i \boldsymbol{w}_i$. 对其进行签名计算有

$$\begin{aligned} \mathrm{sign}(\boldsymbol{W}) &= \mathrm{sign}\left(\sum_{i=1}^{k} a_i \boldsymbol{w}_i\right) = \\ &\sum_{j=1}^{N} G_{sk} \cdot r_j \left(\sum_{i=1}^{k} a_i v_{ij} \bmod q\right) = \\ &\sum_{i=1}^{k} a_i \left(\sum_{j=1}^{N} G_{sk} \cdot r_j v_{ij} \bmod q\right) = \\ &\sum_{i=1}^{k} a_i \mathrm{sign}(\boldsymbol{w}_i) \end{aligned} \tag{5}$$

因此证明了方案具有同态性质. 从而得到新消息向量的签名计算式为

$$\mathrm{sign}(\boldsymbol{W}) = \sum_{i=1}^{k} a_i \mathrm{sign}(\boldsymbol{w}_i) \tag{6}$$

4）批验证(Batch Verify)：在对代 I 的消息进行批验证时，借鉴文献[7]中对分代网络编码的缓冲模式，假设节点对一个代消息的缓冲时间为 t，则节点在缓冲时间结束前对所有代标识符为 I 的消息进行验证. 设节点在代 I 缓冲时间结束前收到了 k 个代 I 的数据包，其形式分别为 $(I, \boldsymbol{w}_1, \sigma_1), (I, \boldsymbol{w}_2, \sigma_2), \cdots, (I, \boldsymbol{w}_k, \sigma_k)$，其中，$\boldsymbol{w}_i = (v_1', v_2', \cdots, v_N')$. 那么首先计算 $H_G(I)$，验证 $H_G(I) = G_{pk}$ 是否成立，然后验证式(7)是否成立：

$$e\left(P, \sum_{i=1}^{k} \sigma_i\right) = e\left(\sum_{i=1}^{k} \left(\sum_{j=1}^{N} D_j \cdot v_{ij}'\right), G_{pk}\right) \tag{7}$$

若式(7)成立则所有消息都属于代 I 的消息，不存在代间污染攻击. 否则再对每一个消息向量

进行单独验证确定哪一个是污染消息向量,验证方式与式(7)类似:

$$e(P,\sigma)=e\left(\sum_{i=1}^{N}D_i\cdot v_i',G_{\mathrm{pk}}\right) \tag{8}$$

3 方案分析

3.1 正确性分析

对方案中的两个验证公式进行正确性的分析.

1) 节点验证单个签名的情况.

$$\begin{aligned}e(P,\sigma)&=e\left(P,\sum_{i=1}^{N}G_{\mathrm{sk}}\cdot r_iw_i\right)=\\&e\left(P,\sum_{i=1}^{N}\alpha\cdot G_{\mathrm{pk}}\cdot r_iw_i\right)=\\&e\left(P\cdot\sum_{i=1}^{N}\alpha\cdot r_iw_i,G_{\mathrm{pk}}\right)=\\&e\left(\sum_{i=1}^{N}D_iw_i,G_{\mathrm{pk}}\right)\end{aligned} \tag{9}$$

2) 节点批验证的情况.

$$\begin{aligned}e\left(P,\sum_{i=1}^{k}\sigma_i\right)&=e\left(P,\sum_{i=1}^{k}\left(\sum_{j=1}^{N}G_{\mathrm{sk}}\cdot r_jw_{ij}\right)\right)=\\&e\left(P,\sum_{i=1}^{k}\left(\sum_{j=1}^{N}\alpha\cdot G_{\mathrm{pk}}\cdot r_jw_{ij}\right)\right)=\\&e\left(P\cdot\sum_{i=1}^{k}\left(\sum_{j=1}^{N}\alpha\cdot r_jw_{ij}\right),G_{\mathrm{pk}}\right)=\\&e\left(\sum_{i=1}^{k}\left(\sum_{j=1}^{N}D_jw_{ij}\right),G_{\mathrm{pk}}\right)\end{aligned} \tag{10}$$

由以上两个推导过程可证明方案的正确性.

3.2 安全性分析

1) 安全定义.

定义 1 在随机预言模型下,如果一个概率多项式时间的攻击者在以下的安全游戏中的优势是可忽略的,则网络编码同态签名方案是安全的.

① 初始化:挑战者运行 Setup 算法生成公开参数和系统私钥,并把系统公开参数发送给攻击者,保存私钥.

② 询问:攻击者指定一系列的消息子空间 $V_i\in\boldsymbol{Z}_q^N$. 对于每一个 i,挑战者从 $\{0,1\}^*$ 任意选择一个代标识符 I_i 与之对应,通过签名算法 Sign 生成代 I_i 中每个消息向量的签名 σ_i,并将签名 σ_i 和代标识符 I_i 发送给攻击者.

③ 伪造:攻击者输出 $I^*\in\{0,1\}^*$,非零消息向量 $\boldsymbol{m}^*$($\boldsymbol{m}^*$ 不属于代 I_i),以及签名 σ^*. 若 $I^*=I_i$,或者 $I^*\neq I_i$ 时验证 Verify($I^*,\boldsymbol{m}^*,\sigma^*$)合法,则说攻击者赢得了本次游戏.

2) 安全性证明.

① 在 $I^*=I_i$ 但 $\boldsymbol{m}^*$ 不属于代 I_i 的消息的情况下证明方式与文献[15]类似.

在询问阶段,攻击者选择由向量($\boldsymbol{M}_1,\boldsymbol{M}_2,\cdots,\boldsymbol{M}_m$)张成的消息子空间 $\boldsymbol{V}$ 并向挑战者询问该消息子空间中各消息向量的签名. 挑战者收到询问请求后通过随机预言机进行应答. 挑战者首先选取 I 作为 $\boldsymbol{V}$ 的代标识符,然后通过签名算法生成各向量 $\boldsymbol{M}_i$ 的签名 σ_i,并将签名 σ_i 和代标识符 I 发送给攻击者. 由此攻击者可获得式(11)的方程组:

$$\begin{bmatrix}v_{11}&v_{12}&\cdots&v_{1N}\\v_{21}&v_{22}&\cdots&v_{2N}\\\vdots&\vdots&\cdots&\vdots\\v_{m1}&v_{m2}&\cdots&v_{mN}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_N\end{bmatrix}=\begin{bmatrix}\sigma_1\\\sigma_2\\\vdots\\\sigma_m\end{bmatrix} \tag{11}$$

式中 $x_i=G_{\mathrm{sk}}\cdot r_i$

攻击者输出伪造消息向量($I,\boldsymbol{m}^*,\sigma^*$),$\boldsymbol{m}^*\notin\boldsymbol{V}$. 如果此伪造消息量能够通过验证,那么以下方程组成立:

$$\begin{bmatrix}v_{11}&v_{12}&\cdots&v_{1N}\\\vdots&\vdots&\cdots&\vdots\\v_{m1}&v_{m2}&\cdots&v_{mN}\\v_1^*&v_2^*&\cdots&v_N^*\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_N\end{bmatrix}=\begin{bmatrix}\sigma_1\\\vdots\\\sigma_m\\\sigma_*\end{bmatrix} \tag{12}$$

假设式(11)的系数矩阵与增广矩阵的秩为 λ,则其解的个数为 $q^{n-\lambda}$,式(12)的秩为 $\lambda+1$,解为 $q^{n-\lambda-1}$,同时也可知式(11)中将有 q 个解满足式(12). 当 q 足够大时,攻击者赢得游戏的概率 $\frac{1}{q}$可忽略不计.

② $I^*\neq I_i$ 时,在随机预言模型下,哈希函数 H_{G} 被看成是一个随机预言机,而攻击者没有能力在多项式时间内找到 $I^*\neq I_i$ 使得 $H_{\mathrm{G}}(I^*)=H_{\mathrm{G}}(I_i)$,所以 $G_{\mathrm{pk}}^*\neq G_{\mathrm{pk}}$,验证时不能通过.

3.3 开销分析

将本方案与 Boneh 方案的验证开销进行比较分析. 两个方案开销的比较分析分为对验证计算开销的比较以及验证时对数据包等待时间的比较.

1) 验证计算开销的比较.

在比较验证计算开销时,不考虑数据包等待带来的开销.

Boneh 方案中,其签名形式为

$$\sigma=\left(\prod_{i=1}^{N}H(I,i)^{v_i}\right)^{\alpha}$$

验证形式为

$$e(\sigma,h)=e\left(\prod_{i=1}^{N}H(I,i)^{w_i},u\right)$$

式中，α 为系统私钥；$u = h^{\alpha}$ 为系统公钥.

本方案和 Boneh 方案中均采用了双线性对来构造网络编码签名方案. 由于在两个方案验证过程中的模加、模乘以及哈希操作的开销远小于双线性对操作的开销，所以以下的验证计算开销分析中只考虑双线性对操作的开销比较.

对两个方案的验证计算开销进行比较之前先假设中间节点收到同代的消息条数平均为 λ. 令 C_e 为执行双线性对操作所需时间.

① 不存在代间污染攻击的情况下，Boneh 方案对每条消息进行验证时需进行两次双线性对操作，其所需时间为 $(2 \times \lambda) C_e$. 由于不存在代间污染攻击，所以本方案下只需要进行一次批验证即可，由验证式(7)可知本方案进行批验证只需两次双线性对操作，所需时间为 $2 \times C_e$.

② 当存在代间污染攻击的情况下，Boneh 方案所需时间与不存在代间污染攻击下所需时间一样为 $(2 \times \lambda) C_e$. 使用本方案先进行批验证，确定存在代间污染攻击后，还需对每条消息进行逐一验证确定哪一个是污染消息，所以需要时间为批验证时间加上所有单个验证的时间，其总耗时为 $(2 \times \lambda + 2) C_e$. 此种情况下两方案所需时间之比为

$$\frac{2\lambda}{2\lambda + 2} = \frac{\lambda}{\lambda + 1}$$

综合两种情况考虑，假设节点受到污染攻击的概率为 p. 则采用本方案验证所需的平均时间为 $p(2\lambda + 2) C_e + (1 - p) 2C_e = 2C_e(p\lambda + 1)$，此时两方案验证时间之比为 $\xi = \dfrac{2\lambda C_e}{2C_e(p\lambda + 1)} = \dfrac{\lambda}{p\lambda + 1}$. 当 $\lambda \geqslant 2$，即至少有两条同代的消息到达验证节点，且节点受到代间污染攻击概率 $p < \dfrac{1}{2}$ 时，ξ 始终大于 1，即本方案的验证开销更小.

由以上分析可知，本方案和 Boneh 方案相比，节点在有多条同代消息到达且发生代间污染概率 $p < \dfrac{1}{2}$ 情况下的计算开销更小.

2）数据包等待时间比较.

在比较数据包等待时间带来的开销时，不考虑验证计算开销. 图 4 和图 5 分别为单验证模型和批验证模型. 单验证时，节点收到数据包立刻开始验证，数据包不需要等待；批验证时，数据包等待直到缓冲时间结束前，节点进行批验证. 由图 4 和图 5 可见，在相同的时间内，因为数据包到达速率一样，所以二者缓冲数据包数量相同，数据包等待时间不影响批验证开销.

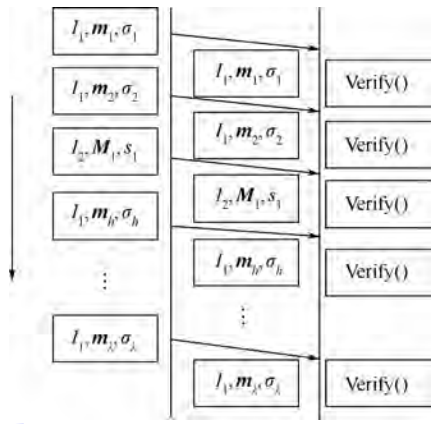


图 4　单验证模型

Fig. 4　Individual verification model

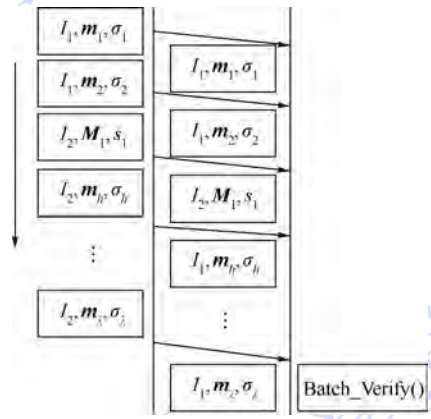


图 5　批验证模型

Fig. 5　Batch verification model

综上考虑，结合批验证的方案与普通的单验证方案相比能够有效地节省验证开销，提高系统效率.

4　结　论

1）提出一种具有同态性质的网络编码签名方案，解决网络编码中存在的因代间污染攻击造成的消息串扰问题.

2）在节点有多条同代消息到达且发生代间污染攻击概率 $p < \dfrac{1}{2}$ 情况下，本方案与 Boneh 方案相比，验证效率更高.

参考文献（References）

[1] Ahlswede R, Cai N, Li S. Network information flow[J]. IEEE Transactions on Information Theory, 2000, 46(4): 1204-1216.

[2] He M, Chen L, Wang H, et al. Survey on secure transmission of network coding in wireless networks[C]//International Conference on Computer Science and Service System. Washington, D. C.: IEEE, 2012: 1216-1219.

[3] 曹张华，唐元生. 安全网络编码综述[J]. 计算机应用，2010，30(2)：499-505.

北航学报 赠阅

Cao Z H, Tang Y S. Survey on secure network coding[J]. Journal of Computer Applications, 2010, 30(2): 499-505(in Chinese).

[4] Yu Z, Wei Y, Ramkumar B, et al. An efficient signature-based scheme for securing network coding against pollution attacks [C]//Proceedings of International Conference on Computer Communications(INFOCOM). Washington, D. C.: IEEE, 2008: 1409-1417.

[5] 裴恒利,尚涛,刘建伟. 融合时间戳和同态签名的安全网络编码方法[J]. 通信学报,2013,34(4):28-35.
Pei H L, Shang T, Liu J W. Secure network coding method merged with timestamp and homomorphic signature[J]. Journal on Communication, 2013, 34(4): 28-35(in Chinese).

[6] Shang T, Pei H L, Liu J W. Secure network coding based on lattice signature[J]. China Communication, 2014(1): 138-151.

[7] Chou P A, Wu Y, Jain K. Practical network coding[C]//Proceedings of the Annual Allerton Conference on Communication Control and Computing, 2003.

[8] Zhao F, Kaller T, Medard M, et al. Signatures for content distribution with network coding[C]//IEEE International Symposium on Information Theory. Washington, D. C.: IEEE, 2007: 556-560.

[9] Kehdi E, Li B. Null keys: limiting malicious attacks via null space properties of network coding[C]//Proceedings of International Conference on Computer Communication (INFOCOM). Washington, D. C.: IEEE, 2009: 1224-1232.

[10] Boneh D, Freeman D, Katz J, et al. Signing a linear subspace: signature schemes for network coding[C]//12th International Conference on Practice and Theory in Public Key Cryptography. Berlin: Springer, 2009: 68-87.

[11] Tseng Y M, Wu T Y, Wu J D. Towards efficient ID-based signature schemes with batch verifications from bilinear pairing [C]//International Conference on Availability, Reliability and Security. Washington, D. C.: IEEE, 2009: 935-940.

[12] Camenisch J, Hohenberger S, Pedersent M Q. Batch verification of short signatures[C]//Advances in Cryptology-Eurocrypt 2007. Berlin: Springer, 2007: 246-263.

[13] 黄佳庆,李宗鹏. 网络编码原理[M]. 北京:国防工业出版社,2012:31.
Hang J Q, Li Z P. Network coding principles[M]. Beijing: National Defense Industry Press, 2012: 31(in Chinese).

[14] Lee S H, Gerla M, Krawczyk H, et al. Performance evaluation of secure network coding using homomorphic signature[C]//International Symposium on Network Coding (NetCod). Piscataway, NJ: IEEE, 2011: 1-6.

[15] Liu G J, Wang B. Secure network coding against intra/inter-generation pollution attacks[J]. Communications, China, 2013, 10(8): 100-110.

[16] Boneh D, Franklin M. Identity-based encryption from the Weil pairing[C]//Advances in Cryptology-CRYPTO 2001. Berlin: Springer, 2001: 213-229.

Signature scheme for network coding against inter-generation pollution attacks

PENG Tianli, SHANG Tao*, LIU Jianwei

(School of Electronic and Information Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: The concept of generation is introduced into network coding so as to adapt to the needs of real-time application. Considering that network coding is vulnerable to inter-generation pollution attacks and causes message crosstalk, a generation-identifier based signature scheme for network coding was proposed. Firstly, the two-level private keys, including the generation private key and the message private key, both derived by the generation identifier were set. This procedure enabled nodes to judge the generation attribute of messages and decided whether to continue coding. Secondly, a signature algorithm was constructed with homomorphic property by using the bilinear pairing. Meanwhile the two-level private keys were separated to verify signatures by batch, which made nodes verify all the massages of the same generation simultaneously. Finally, through the analysis of random oracle model, the scheme was proved secure against adaptively chosen message attack. The result indicates that the scheme can reduce computation cost and improve performance of the system.

Key words: network coding; inter-generation pollution attacks; homomorphic signature; batch verification; bilinear pairing

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0266

矩的显式积分算法研究与应用

付子豪，龚光红*
（北京航空航天大学 自动化科学与电气工程学院，北京 100191）

摘 要：矩的求解通常被用于求解有限元、体积、惯性矩等问题中. 基于矩的叠加性，首先给出了在三维空间中计算域的离散方式，并推导了矩的显式积分公式，随后将其推广到 n 维空间中，该表达式易于在计算机上实现；设计了矩的并行计算算法，并通过 Fortran 和 Python 混编的方式，实现了矩的并行计算；对多重精度下的样例数据给出了一个算例，实现了零阶矩和二阶矩的计算，并和串行算法、逐次降维算法作出比较，进行了效率分析和误差分析. 结果显示，矩的显式积分并行计算算法易于程序实现，并且在效率上高于串行算法，能够很容易推广到高维空间，该算法具有高度可并行性，误差主要来自计算域离散.

关 键 词：矩；显式积分公式；并行计算；逐次降维；计算域离散
中图分类号：TP301.6
文献标识码：A **文章编号**：1001-5965(2015)04-0727-05

在采用有限元方法[1]进行刚度矩阵计算、求解物体体积、求解物体的惯性矩[2-4]等各种力学参数时，常常需要涉及矩的计算. Soerjadi 等[5]给出了在二维平面情况下的计算公式. 该公式基于格林公式的一个特例得到，令格林公式中，向量场函数的任意一项恒为 0，导出将面积分转化为线积分的通式，然后将线积分用空间已知点的值表示出来，继而得出面积分的值. 然而，该方法要求向量场函数面积分项中只有一项，且该项需含有对空间所有维度的微分项，因此在三维或更高维度的空间中，无法继续用该方法得到结果.

Johnson 和 Mirtich 等[1,6]提出了在三维空间进行体积矩积分的方法. 该方法首先对于给定的标量场函数构造了一个向量场函数的散度形式，然后通过高斯公式将该向量场函数的散度转化为向量场在计算域表面的积分，随后将该积分拆写成标量场积分的形式，实现了将三维的体积矩转化为二维的面矩，随后再通过斯托克斯公式将面矩的积分转化为线积分和低阶面矩的和的形式，通过递归，得出计算域的矩. 虽然该方法解决了三维空间中体积矩的计算问题，但是仍存在几点不足. 首先，在通过斯托克斯公式将面矩转化为线积分与低阶面矩和的形式的时候，推导非常复杂；其次，过多地使用了递归计算，在程序实现上会增大运行时的开销；最后该方法只适用于三维空间中的体积矩，对于高维空间中的矩无法计算得到.

Sheynin 等[7-9]给出了求解计算域任意阶矩的公式. 并将其推广到了高维空间，该方法未使用降维或递归的手段，而是直接给出了矩的显式表达式，降低了问题的复杂度. 然而该表达式不便于计算机实现.

本文在 Sheynin 等基础之上推导了 n 维空间中的便于计算机程序实现的矩的显示公式. 并基于在三维空间的公式，使用 Python 与 Fortran 混编的技术，实现了三维空间下矩的计算，并以 Boeing777-200LR为例，给出了一个算例，评估了

收稿日期：2014-05-13；**录用日期**：2014-08-25；**网络出版时间**：2014-10-15 10:49
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0266.html
作者简介：付子豪(1990—)，男，四川德阳人，硕士生，fuzihaofzh@gmail.com
* **通讯作者**：龚光红(1968—)，女，四川石柱人，教授，ggh@buaa.edu.cn，主要研究方向为分布仿真与虚拟技术.

引用格式：付子豪，龚光红. 矩的显式积分算法研究与应用[J]. 北京航空航天大学学报，2015，41(4)：727-731. Fu Z H，Gong G H. Explicit moment integration algorithm and its application[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：727-731 (in Chinese).

并行计算的效能,并和逐次降维法进行了对比.

1 矩的定义与公式推导

1.1 矩的定义及物理意义

首先给出矩的定义,如式(1)所示,以下就该公式的计算进行讨论.

$$m_{k_1k_2\cdots k_n}=\int_{\Omega}x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n \tag{1}$$

式中,$x_1,x_2,\cdots,x_n$ 为 n 维空间中的坐标;$m_{k_1k_2\cdots k_n}$ 为在 $x_1,x_2,\cdots,x_n$ 上的 $k_1,k_2,\cdots,k_n$ 阶矩;Ω 为积分计算域. 当 $k_1=k_2=\cdots=k_n=0$ 时,$m_{k_1k_2\cdots k_n}$ 具有体积的物理意义,公式如下:

$$V=\int_{\Omega}1\cdot\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n \tag{2}$$

而当 $\sum_{i=1}^{n}k_i=2$ 时,$m_{k_1k_2\cdots k_n}$ 具有惯性矩的物理意义,简单起见,考虑三维空间中刚体的惯性矩张量 $\boldsymbol{I}$,有

$$\boldsymbol{I}=\begin{pmatrix}I_{xx}&I_{xy}&I_{xz}\\I_{yx}&I_{yy}&I_{yz}\\I_{zx}&I_{zy}&I_{zz}\end{pmatrix} \tag{3}$$

式中,对角线上的 I_{xx},I_{yy},I_{zz} 分别为在 x,y,z 轴上的转动惯量,其定义如下所示:

$$\begin{cases}I_{xx}=\int_{\Omega}(y^2+z^2)\,\mathrm{d}m\\I_{yy}=\int_{\Omega}(x^2+z^2)\,\mathrm{d}m\\I_{zz}=\int_{\Omega}(x^2+y^2)\,\mathrm{d}m\end{cases} \tag{4}$$

而非对角元素则是惯量积,定义如下所示:

$$\begin{cases}I_{xy}=I_{yx}=-\int_{\Omega}xy\mathrm{d}m\\I_{xz}=I_{zx}=-\int_{\Omega}xz\mathrm{d}m\\I_{yz}=I_{zy}=-\int_{\Omega}yz\mathrm{d}m\end{cases} \tag{5}$$

由上可见,体积和惯性矩都可以通过矩积分的方式来获取. 下面先给出三维空间中矩的公式以及推导过程,随后将其推广到 n 维空间中.

1.2 计算域的分割

对于不规则计算域的通式推导是困难的. 因此本文首先将任意形状的计算域进行分割,将整个计算域(Ω)转化为若干个三角形包围的多面体,再在这若干个四面体中进行计算,最后将结果叠加. 在进行四面体矩的计算时,通过坐标变换的方法在顶点为原点、3 条棱与坐标轴重合且长度为 1 的四面体(称作标准四面体)上进行计算,从而得到任意四面体的矩.

对于空间中的任意计算域,首先将其几何外形用三角形面片进行近似离散,对于每一个三角形,将其 3 个顶点与原点相连,便能够得到若干以三角形为底,以原点为顶点的四面体,如图 1 所示. 本文定义四面体和原点相接的 3 条棱的另一个端点的次序和三角形表面法向量满足右手定则,然后分别在每一个四面体上进行积分,得到每一个四面体的矩,然后将这些四面体的矩相加,由于有方向的定义,所有集合体外部的矩都会被抵消,因此所有四面体矩的和即是计算域的矩. 令 T 表示用三角化离散后的计算域,在 T 中,有 r 个顶点在原点的四面体,记作 T_i,有

$$m_T=\sum m_{T_i} \tag{6}$$

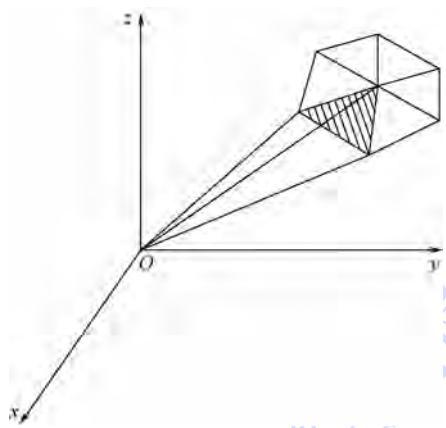


图 1 空间四面体的定义

Fig. 1 Definition of tetrahedron

1.3 n 维空间下标准四面体公式

首先推导在 n 维空间下的标准四面体的计算公式. n 维空间标准四面体积分定义为

$$M=\int_{\Omega}x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n \tag{7}$$

式中,$x_1,x_2,\cdots,x_n\geqslant 0,x_1+x_2\cdots+x_n\leqslant 1$.

使用 Dirichlet 积分公式有

$$M=\int_{\Omega}x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n}\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n=\frac{\Gamma(k_1+1)\Gamma(k_2+1)\cdots\Gamma(k_n+1)}{\Gamma\left(\sum_{i=1}^{n}k_i+n\right)}\cdot\frac{1}{\sum_{i=1}^{n}k_i+n}=\frac{k_1!k_2!\cdots k_n!}{\left(\sum_{i=1}^{n}k_i+n-1\right)!}\cdot\frac{1}{\sum_{i=1}^{n}k_i+n}=\frac{k_1!k_2!\cdots k_n!}{\left(\sum_{i=1}^{n}k_i+n\right)!} \tag{8}$$

式中,$\Gamma(n)$ 表示伽马函数,定义为 $\Gamma(n)=(n-1)!$.

1.4 三维空间任意四面体的矩

对于三维空间中任意四面体的矩,可以采用直接积分的方法来得到. 首先,选取四面体与原点

相接的 3 条棱为新的坐标系,将四面体用它的 3 条边表示,记作 $T(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})$,其中 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ 都是列向量,且 $\boldsymbol{a}=(a_1,a_2,a_3)^{\mathrm{T}}$,$\boldsymbol{b}=(b_1,b_2,b_3)^{\mathrm{T}}$,$\boldsymbol{c}=(c_1,c_2,c_3)^{\mathrm{T}}$,从而得到从旧坐标到新坐标的变换矩阵 $(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})^{-1}$,记作 $\boldsymbol{A}$,其中的元素为 $\{a_{ij}\}$,在旧坐标系下变量用 x_i 表示,而新坐标系下的变量用 y_i 表示,则坐标转换关系为 $\boldsymbol{x}=\boldsymbol{A}\boldsymbol{y}$.

记体积矩为 $m_{k_1k_2k_3}(T)$,$\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}$ 为直角坐标系的基,将向量 $\boldsymbol{x}$ 用 $\boldsymbol{A}\boldsymbol{y}$ 替换,即 $x_i=\sum_{j=1}^{3}a_{ij}y_j$,有

$$m_{k_1k_2k_3}(T)=\int_{T(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})}x_1^{k_1}x_2^{k_2}x_3^{k_3}\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3=$$
$$|\boldsymbol{A}|\int_{T(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k})}((\boldsymbol{A}\boldsymbol{y})_{(1)})^{k_1}((\boldsymbol{A}\boldsymbol{y})_{(2)})^{k_2}\cdot$$
$$((\boldsymbol{A}\boldsymbol{y})_{(3)})^{k_3}\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}y_3 \tag{9}$$

式中,$(\boldsymbol{A}\boldsymbol{y})_{(i)}$ 代表向量 $\boldsymbol{A}\boldsymbol{y}$ 的第 i 个分量.

$$m_{k_1k_2k_3}(T)=\int_{T(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k})}(a_{11}y_1+a_{12}y_2+a_{13}y_3)^{k_1}\cdot$$
$$(a_{21}y_1+a_{22}y_2+a_{23}y_3)^{k_2}\cdot$$
$$(a_{31}y_1+a_{32}y_2+a_{33}y_3)^{k_3}\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}y_3=$$
$$\int_{T(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k})}\sum_{k_{11}+k_{12}+k_{13}=k_1}\sum_{k_{21}+k_{22}+k_{23}=k_2}\sum_{k_{31}+k_{32}+k_{33}=k_3}$$
$$\frac{k_1!k_2!k_3!}{k_{11}!k_{12}!k_{13}!k_{21}!k_{22}!k_{23}!k_{31}!k_{32}!k_{33}!}\cdot$$
$$\prod_{p=1,q=1}^{3}a_{pq}^{k_{pq}}y_1^{k_{11}+k_{21}+k_{31}}y_2^{k_{12}+k_{22}+k_{32}}y_3^{k_{13}+k_{23}+k_{33}}\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}y_3 \tag{10}$$

这里,使用到了多项式展开定理. 式中,$\sum_{k_{i1}+k_{i2}+k_{i3}=k_i}$ 表示对所有满足 $k_{i1}+k_{i2}+k_{i3}=k_i$ 的组合进行求和[10]. 再代入标准四面体矩公式得:

$$m_{k_1k_2k_3}T(\boldsymbol{a},\boldsymbol{b},\boldsymbol{c})=$$
$$|\boldsymbol{A}|\sum_{k_{11}+k_{12}+k_{13}=k_1}\sum_{k_{21}+k_{22}+k_{23}=k_2}\sum_{k_{31}+k_{32}+k_{33}=k_3}\left(\prod_{p=1,q=1}^{3}a_{pq}^{k_{pq}}\right)\cdot$$
$$\frac{(k_{11}+k_{21}+k_{31})!(k_{12}+k_{22}+k_{32})!(k_{13}+k_{23}+k_{33})!}{k_{11}!k_{12}!k_{13}!k_{21}!k_{22}!k_{23}!k_{31}!k_{32}!k_{33}!}\cdot$$
$$\frac{k_1!k_2!k_3!}{(k_1+k_2+k_3+3)!} \tag{11}$$

1.5　矩公式在高维空间的推广

将三维空间的通式向高维空间推广,容易得到高维空间中的矩公式:

$$m_{k_1k_2\cdots k_n}T(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_n)=$$
$$|\boldsymbol{A}|\int_{T(i_1,i_2,\cdots,i_n)}\prod_{i=1}^{n}\left(\sum_{i=1}^{n}a_{ij}y_j\right)^{k_i}\mathrm{d}y_1\mathrm{d}y_2\cdots\mathrm{d}y_n=$$
$$|\boldsymbol{A}|\int_{T(i_1,i_2,\cdots,i_n)}\prod_{i=1}^{n}\left(\sum_{k_{i1}+k_{i2}+\cdots+k_{in}=k_i}\frac{k_i!}{k_{i1}!k_{i2}!\cdots k_{in}!}\cdot\right.$$
$$\left.\prod_{j=1}^{n}(a_{ij}y_j)^{k_{ij}}\right)\mathrm{d}y_1\mathrm{d}y_2\cdots\mathrm{d}y_n=$$
$$|\boldsymbol{A}|\int_{T(i_1,i_2,\cdots,i_n)}\left(\prod_{i=1}^{n}\times\sum_{k_{i1}+k_{i2}+\cdots+k_{in}=k_i}\prod_{p=1}^{n}\frac{k_p!}{k_{p1}!k_{p2}!\cdots k_{pn}!}\cdot\right.$$
$$\left.\prod_{q=1,r=1}^{n}(a_{qr})^{k_{qr}}\prod_{s=1}^{n}(y_s)^{k_{1s}+k_{2s}+\cdots+k_{ns}}\right)\mathrm{d}y_1\mathrm{d}y_2\cdots\mathrm{d}y_n=$$
$$|\boldsymbol{A}|\left(\prod_{i=1}^{n}\times\sum_{k_{i1}+k_{i2}+\cdots+k_{in}=k_i}\prod_{p=1}^{n}\frac{k_p!}{k_{p1}!k_{p2}!\cdots k_{pn}!}\cdot\right.$$
$$\left.\prod_{q=1,r=1}^{n}(a_{qr})^{k_{qr}}\cdot\frac{\prod_{s=1}^{n}\left(\sum_{t=1}^{n}k_{ts}\right)!}{\left(\sum_{u=1}^{n}k_u+n\right)!}\right) \tag{12}$$

式中,$\prod_{i=1}^{n}\times\sum_{k_{i1}+k_{i2}+\cdots+k_{in}=k_i}$ 表示 n 个求和号累加的形式. 由于多重求和号可以通过多重循环实现,因此该形式易于计算机程序设计.

2　矩的并行计算算法

在第 1 节中,推导了三维空间下任意矩积分的公式. 在本节,基于该公式设计出矩的并行计算算法,并对并行、串行显式积分公式算法和 Mirtich[6] 提出的逐次降维法进行比较.

2.1　矩的并行显式积分公式算法

由计算域的分解可知,矩的计算需要求出 n 个四面体的矩,然后将其相加,而这部分计算是相互间没有关联的,因此,可以通过并行计算的方法来实现,实现原理如图 2 所示.

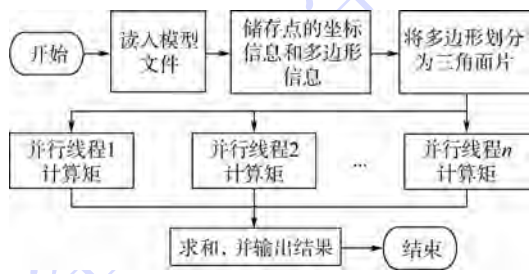


图 2　并行算法流程图

Fig. 2　Flowchart of parallel algorithm

本次计算采用 Boeing777-200LR 的三角化表面网格(图 3)作为测试几何体. 通过设置不同的三角剖分精度,得到了不同层级的网格. 通过对单位密度几何体的零阶矩 $\int_\Omega 1\mathrm{d}m$ 和二阶矩 $\int_\Omega y^2\mathrm{d}m$ 的计算,比较了 3 种方法的效率.

为了简化几何处理的复杂度,该程序使用 Python[11-12] 做外部接口;为了提升计算速度,实现并行计算,采用 Fortran 做底层计算,并行的部分也由 Fortran 来实现. 并行的部分通过在 Fortran 代码中加入 OpenMP[13-14],将循环自动地分配到各个 CPU 上,提高计算效率. 计算环境使用 Intel® Core™ i7-3770 CPU主频 3.4 GHz 4 核 8 线

北航学报 赠阅

程 8 GB RAM.

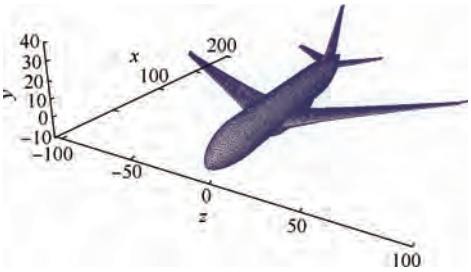


图 3 几何体表面三角网格

Fig. 3 Surface triangular mesh of geometry

2.2 结果分析

通过对不同精度的 Boeing777-200LR 网格[15]进行计算,结果如图 4 和表 1 所示.

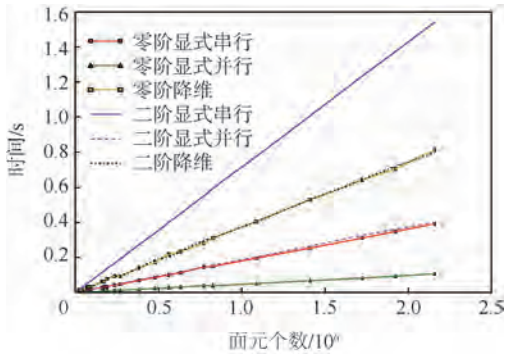


图 4 算例结果分析

Fig. 4 Example results analysis

表 1 计算结果

Table 1 Calculation results

单元数	零阶矩					二阶矩				
	显式积分公式法			逐次降维法		显式积分公式法			逐次降维法	
	矩/kg	串行时间/ms	并行时间/ms	矩/kg	时间/ms	矩/(kg·m²)	串行时间/ms	并行时间/ms	矩/(kg·m²)	时间/ms
20524	66373	3	2	66373	16	2004112	16	0	2004112	0
269948	66689	47	12	66689	93	2024921	109	47	2024926	94
478990	66700	85	23	66700	172	2025507	218	63	2025511	187
631228	66704	113	30	66704	234	2025687	281	93	2025692	234
769506	66705	145	39	66705	280	2025790	343	109	2025795	297
1087336	66708	196	51	66708	405	2025976	499	156	2025982	405
1721682	66710	311	82	66710	640	2026108	764	234	2026113	640
2158232	66711	392	106	66711	811	2026154	967	250	2026159	795

在该算例中,分别通过串行和并行的显式积分公式算法,计算了几何体的零阶和二阶矩,并和逐次降维法做了对比. 可以看出,随网格数的增大,计算时间线性增加;在同等情况下,并行效率大致为串行效率的 3.7 倍;在计算零阶矩时,显式积分公式法效率高于逐次降维法,而当矩的阶数增大时,逐次降维法的效率较高,但仍低于并行的显式积分公式法;显式积分公式法能很快推广到高维空间,而逐次降维法仅限于三维空间中.

在程序并行计算时,由于 CPU 4 核 8 线程一个核上的两线程存在干扰,不便于比较效率,因此在这里关闭 CPU 的超线程功能进行测试,相当于 4 核并行,效率达到了 3.7 倍左右,由于 OpenMP 运行的资源消耗,不能达到理想中 4 倍的运行效率.

对比矩的计算结果误差可知,矩的积分误差主要发生在几何体的离散过程中. 随着网格数增加,矩的值同时增大,而最后趋于收敛,对比第 1 组计算结果和最后一组计算结果,有 4.5% 的误差,这是由于离散网格随网格数的增加而逐渐贴近于真实边界. 其次,对比显式积分公式算法和逐次降维法在相同离散网格下的结果可知,由离散求和造成的矩计算的误差小,其数量级约在 10^{-6} 左右. 其原因是离散几何体矩的计算直接用公式表出,因此其计算误差只表现有限数值精度导致的舍入误差上. 综上,矩积分计算的误差主要由几何体离散导致,在实际计算时,需要采用合适的离散网格数进行离散.

3 结 论

通过分析和计算表明:

1) 将计算域用三角面片离散的方式,能有效地将任意计算域转为多面体,简化了计算;

2) 显式积分公式利于在计算机上实现,能方便地推广到高维空间,逐次降维法无法推广;

3) 随着面元数量的增加,并行和串行计算消耗时间都随之线性增长,在测试计算机上并行计算的效率约为串行效率的 3.7 倍;

4) 矩的计算结果随着网格的细化而逐渐趋于收敛.

参考文献 (References)

[1] Johnson F T, Samant S S, Bieterman M B, et al. TRANAIR-a computer code for transonic analyses of arbitrary configurations, NASA-CR-4348[R]. Washington, D. C.: NASA, 1987.

[2] 洪国雄. 惯性张量的物理意义[J]. 大学物理, 1989, 12(1): 7-8.

Hong G X. The physical meaning of the inertia tensor[J]. College Physics,1989,12(1):7-8(in Chinese).

[3] 秦莉,杨明,郭庆. 遗传算法在质量矩导弹姿态控制中的应用[J]. 北京航空航天大学学报,2007,33(7):769-772.
Qin L,Yang M,Guo Q. Movingmass attitude control law based on genetic algerithm[J]. Journal of Beijing University of Aeronautics and Astronautics,2007,33(7):769-772(in Chinese).

[4] 徐水源. 惯性矩与惯性张量的关系[J]. 黄石教育学院学报,2005,22(4):90-92.
Xu S Y. The relationship of the moment of inertia and inertia tensor[J]. Journal of Huangshi Education College,2005,22(4):90-92(in Chinese).

[5] Soerjadi R. On the computation of the moments of a polygon,with some applications[M]. Netherlands:Stevin Laboratory,1968:43-58.

[6] Mirtich B. Fast and accurate computation of polyhedral mass properties[J]. Journal of Graphics Tools,1996,1(2):31-50.

[7] Sheynin S,Tuzikov A,Vasiliev P. Efficient computations of body moments[J]. Информационные Процессы,2002,2(1):22-23.

[8] Sheynin S A,Tuzikov A V. Explicit formulae for polyhedra moments[J]. Pattern Recognition Letters,2001,22(10):1103-1109.

[9] Tuzikov A V,Sheynin S A,Vasiliev P V. Computation of volume and surface body moments[J]. Pattern Recognition,2003,36(11):2521-2529.

[10] Hall M. Combinatorial theory[M]. New York:John Wiley & Sons,1998:166.

[11] VanRossum G,Drake F L. Python language reference manual[M]. United Kingdom:Network Theory,2003.

[12] VanRossum G,Drake F L. The python language reference[M]. Beaverton,OR,USA:Python Software Foundation,2010.

[13] Chandra R. Parallel programming in OpenMP[M]. San Francisco:Morgan Kaufmann,2001.

[14] Dagum L,Menon R. OpenMP:an industry standard API for shared-memory programming[J]. Computational Science & Engineering,IEEE,1998,5(1):46-55.

[15] Fredericks W J,Antcliff K R,Costa G,et al. Aircraft conceptual design using vehicle sketch pad[C]//48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition. Reston:American Institute of Aeronautics and Astronautics Inc,2010:1-17.

Explicit moment integration algorithm and its application

FU Zihao, GONG Guanghong*

(School of Automation Science and Electrical Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: The calculation of moment is often used in finite element method, volume calculation, moment of inertia calculation, etc. A discrete method of the computational domain in three-dimensional space was proposed firstly based on the superposition of moment. An explicit formula was derived in three-dimensional space and then extended to n-dimensional space, which can be easily implemented on the computer. Secondly, a parallel algorithm of moment calculation was designed and implemented with mixed Fortran and Python. Thirdly, a zero-order and second-order moment was calculated in a multi-fidelity example. The efficiency of the algorithm was compared with a serial algorithm and a successive dimensionality reduction algorithm. Meanwhile, efficiency analysis and error analysis were presented. The result shows that the explicit moment integration algorithm can be easily implemented with programs and runs faster than the serial algorithm. It is highly parallel and can also be easily extended to a higher dimensional space. The algorithm is highly parallel, whose error mainly comes from the discrete process of the computational domain.

Key words: moment; explicit moment formula; parallel computing; successive dimensionality reduction; discrete computational domain

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0289

基于局部径向二值模式的三维人脸识别

全下文载

刘青，孙军华*
（北京航空航天大学 仪器科学与光电工程学院，北京 100191）

摘 要：针对人脸识别中局部特征的提取，提出了局部径向二值模式（LRBP，Local Radial Binary Pattern），并将其用于三维人脸识别. 首先，对经过预处理的人脸深度图像进行区域划分；然后用局部径向二值模式提取子区域的特征序列，并将其链接在一起构成三维人脸的特征向量；最后，利用 Fisherface 方法对三维人脸特征向量进行训练和识别. 在中国科学院自动化研究所三维人脸数据库中选取样本，利用 LRBP 对其进行识别，结果表明该方法在基本不损失识别率的前提下，可以有效提高识别的效率.

关 键 词：模式识别；三维人脸识别；局部特征；二值模式；局部径向二值模式
中图分类号：TP391.4
文献标识码：A **文章编号**：1001-5965(2015)04-0732-05

近些年，随着计算机及其应用技术的迅速发展，生物特征识别作为新兴的信息认证技术正逐步发展起来. 人脸识别是其中的一个分支，它是一种利用人的面部特征作为识别信息的身份验证技术. 最初，该技术主要应用于二维图像，但由于二维图像本身易受到多种因素干扰，使得这一技术的发展受到制约. 随着三维成像技术的发展，人脸识别被逐步引入到三维图像中来. 三维图像不易受光照等因素的影响，因此比二维图像更适合用于人脸识别[1-4].

基于面部特征的算法是人脸识别中常用的方法，SIFT 特征[5]、Point Signature[6] 描述符和曲率等特征都可用于人脸识别. 局部二值模式（LBP）由 Ojala 等[7] 提出，最初是用它来提取图像的纹理信息. 后来被一些研究人员用于面部特征提取[8-9]，并引入到二维人脸识别中[10-12]，且得到了不错的效果. Guo 等[13] 提出的 CLBP 算子由 3 部分组成：局部灰度差的符号部分（CLBP_S）、幅值部分（CLBP_M）及中心像素灰度（CLBP_C）. 其中，CLBP_S 相当于基本的 LBP 算子. 在单一算子中，LBP 的特征提取能力最好，组合而成的 CLBP 算子的特征表征能力超过 LBP. Yan 等[14] 将不同采样半径的 LBP 算子结合在一起构成了 Multi-Radius LBP，并将其与现有的描述符 Gabor 特征、LBP 和 EOH 进行了比较. Gabor 特征和 LBP 在局部特征提取方面的效果相当且可以互补，两者结合后的结果较单独使用都有了明显提高，而 Multi-Radius LBP 在现有描述符中的识别结果最好. 然而以上方法在提高描述符的局部特征表现能力的同时也大幅增加了算子的维数，忽视了高维算子对运算效率的影响. 本文针对这一问题提出了局部径向二值模式，并将其应用到三维人脸识别中.

1 局部二值模式介绍

1.1 LBP

局部二值模式（LBP）以中心点像素值作为阈值，对邻域像素进行二值化处理. 其表达式为

收稿日期：2014-05-21；**录用日期**：2014-09-26；**网络出版时间**：2014-10-20 16:15
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0289.html
基金项目：国家自然科学基金资助项目（61275162）
作者简介：刘青（1989—），女，河北邯郸人，硕士生，602919926@163.com
* **通讯作者**：孙军华（1975—），男，湖北荆门人，副教授，sjh@buaa.edu.cn，主要研究方向为光电精密测量、机器视觉和图像理解.

引用格式：*刘青，孙军华. 基于局部径向二值模式的三维人脸识别[J]. 北京航空航天大学学报，2015，41(4)：732-736.*
Lin Q，Sun J H. 3D face recognition based on local radial binary pattern[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：732-736 (in Chinese).

$$L_{\mathrm{BP}(P,R)}=\sum_{p=0}^{P-1}S(g_p-g_c)2^p \tag{1}$$

$$S(x)=\begin{cases}1 & x\geqslant 0\\ 0 & x<0\end{cases}$$

g_c 为中心像素值；g_p 为邻域像素值；P 为邻域点个数；R 为邻域半径.

通常 LBP 算子取 3×3 邻域，将标准的 LBP 算子扩展为圆形邻域后可得到任意半径和邻域点数的算子. 图 1 给出了圆形邻域下 (P,R) 取不同值时算子邻域示意图.

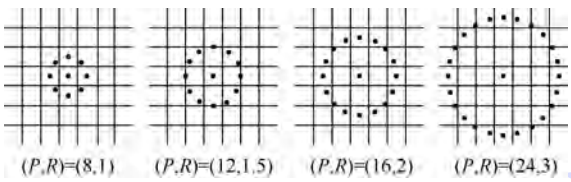


图 1　LBP 圆形邻域示意图

Fig. 1　Circularly symmetric neighbor sets for LBP

当邻域点数为 P 时，基本 LBP 算子的模式数为 2^P（$P=8$，模式数 $=256$），且模式数会随着 P 以幂次增长. 随后，Ojala 等[15]提出了 Uniform Pattern（统一模式），大大缩减了原有模式数种类. 统一模式的描述方式如式(2)：

$$L^{U2}_{\mathrm{BP}(P,R)}=\begin{cases}\sum_{i=0}^{P-1}S(g_i-g_c) & U(L_{\mathrm{BP}(P,R)})\leqslant 2\\ P+1 & \text{其他}\end{cases} \tag{2}$$

$$U(L_{\mathrm{BP}(P,R)})=|S(g_{P-1}-g_c)-S(g_0-g_c)|+\sum_{i=1}^{P-1}|S(g_i-g_c)-S(g_{i-1}-g_c)|$$

根据式(2)的表达，统一模式是二进制序列中 0 和 1 变换的次数不大于 2 的模式，大于 2 则为非统一模式. 统一模式有较强的分类能力，并且包含了图像的大部分结构信息. 但将所有非统一模式归为一类，仍会使部分有用信息丢失.

1.2　MS_LBP

单一 LBP 算子提取的结构信息是有限的，将不同尺度算子结合可以得到更多有用信息. 通常，先分别计算不同尺度 LBP 算子的特征向量，再将这些向量串接在一起得到一个高维特征向量. 该向量在表征局部结构信息方面更有优势，但却是以维数的增加作为代价的.

2　局部径向二值模式

为提取更多中心点邻域结构信息，且不增加特征向量维数，本文提出了局部径向二值模式（LRBP）. 其不仅可以更有效地提取局部结构信息，而且特征向量维数低，减小了后续计算的压力.

2.1　LRBP 生成

以中心像素作为原点，沿直线向外辐射的方向定义为径向. 第一径向 r 为水平向右的方向. 以 r 为起始，$2\pi/P$（P 为单周邻域点数，即同一采样圆上的邻域点数）为间隔选取其他径向 $\boldsymbol{r}_i$（$i=2,3,\cdots,P$）. 将半径为 $R_1,R_2,\cdots,R_n$ 的采样圆与各径向 $\boldsymbol{r}_i$（$i=2,3,\cdots,P$）的交点选为邻域点，每个径向上有 n 个点，共计 $n\times P$ 个点. 然后，以中心点像素值作为阈值，对所取的 $n\times P$ 个邻域点进行二值化处理. 按照靠近中心点处为高位，外侧为低位的原则，将各个径向上的二进制序列转化为十进制数，并将该数作为中心点在径向 r 上的特征数. LRBP 可用图 2 进行描述.

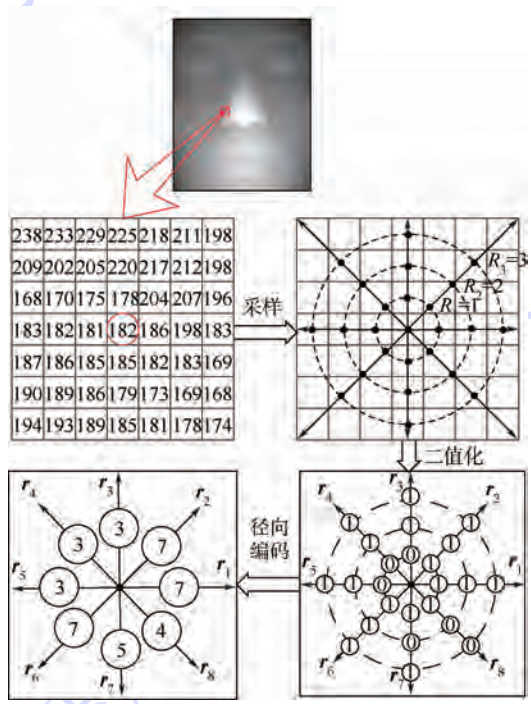


图 2　LRBP 示意图

Fig. 2　Schematic diagram of LRBP

LRBP 较 MS_LBP 的一个重要优势在于：在采样半径和邻域点数相同的情况下，由 LRBP 得到的特征向量的维数较低. 这是因为其沿径向对二值化结果进行编码后得到的模式数少，从而经过直方图统计之后的特征向量维数低. 表 1 给出了在采样半径 R_1,R_2,R_3 保持不变，P 取不同值时 LRBP 和 MS_LBP 的维数（采用统一模式）.

表 1　LRBP 和 MS_LBP 的维数差别

Table 1　Dimensional differences between LRBP and MS_LBP

算子	P		
	8	10	12
MS_LBP	177	279	405
LRBP	64	80	96

2.2 基于 LRBP 的三维人脸识别

与传统 LBP 一样,利用 LRBP 进行三维人脸识别也是通过提取人脸图像的直方图特征向量来完成的. 为去掉噪声和冗余数据的干扰,将原始三维点云数据转化为深度图前需要进行滤波和面部分割等处理. 然后利用 Li 等[16]的方法获得人脸深度图像. 通常,在提取特征向量之前要将人脸图像划分成若干个子图,这里采用等分方式,记分割的块数为 M. 对子图中的任意一点,利用局部径向二值模式计算得到其 P 个十进制特征数. 然后对子图中所有点在相同径向上的特征数进行直方图统计,得到序列 $h_i(i=1,2,\cdots,P)$. 然后将各方向的序列连接在一起,即得到该子图的特征向量 $\boldsymbol{H}_j=(h_1,h_2,\cdots,h_P)(j=1,2,\cdots,M)$. 记整个人脸图像的特征向量为 $\boldsymbol{L}$,则 $\boldsymbol{L}=(H_1,H_2,\cdots,H_M)$. 由于 $\boldsymbol{L}$ 的维数较高,所以采用降维和分类效果较好的 Fisherface[17]方法对特征向量进行训练和识别. 假设样本库中有 Q 个人的人脸数据,每人有 $n_i(i=1,2,\cdots,Q)$ 个训练样本,则整个训练集样本数为 $N(N=n_1+n_2+\cdots+n_Q)$. 每个训练样本的第 $j(j=1,2,3)$ 维特征向量记作 $\boldsymbol{x}_i^{kj}$,其中 i,k 表示第 i 个人的第 k 个样本. 训练流程如图 3 所示.

图 3 样本训练流程图

Fig. 3 Block diagram of training

完成样本训练后,将待识别样本的第 j 维特征向量向矩阵 $\boldsymbol{W}^j\boldsymbol{U}^j$ 作投影,得到投影向量 $\boldsymbol{\eta}^j$;利用余弦函数度量识别样本 $\boldsymbol{\eta}^j$ 与训练样本 $\boldsymbol{y}_i^j$ 的相似度. 对于向量 $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\cdots,\alpha_n)$ 和 $\boldsymbol{\beta}=(\beta_1,\beta_2,\cdots,\beta_n)$,余弦相似度 $S_{\cos}$ 定义为

$$S_{\cos}=\frac{\langle\boldsymbol{\alpha},\boldsymbol{\beta}\rangle}{\|\boldsymbol{\alpha}\|_2\|\boldsymbol{\beta}\|_2}=\frac{\sum_{i=1}^{n}\alpha_i\beta_i}{\|\boldsymbol{\alpha}\|_2\|\boldsymbol{\beta}\|_2} \tag{3}$$

3 实 验

3.1 实验设置

利用中国科学院自动化研究所 CASIA-3D FaceV1 三维人脸数据库对 LRBP 的表征和识别能力进行验证. 该数据库有 123 名对象,包含光照、表情和姿态的变化. 从每个对象中选取 15 幅图像,其中包括 5 幅中性表情和 10 幅不同表情(微笑、大笑、生气、惊吓和闭眼)的图像,每种表情图像各 2 张,总共 1 845 幅三维人脸图像.

为了评价表情变化对识别的影响,将实验分为 2 部分. 第 1 部分测试 LRBP 对中性表情的识别能力. 其中训练集合共 246 张图像,为每个对象的2 幅中性表情图像,这个集合用来训练 LRBP. 每个对象剩余的 3 张中性表情图像构成了测试集合,共计 369 张. 该集合作为测试集合用来测试中性表情下的识别情况. 2 张训练图像按照排列组合的方式选取,共 $C_5^2=10$ 种选择方式. 对每种选择都在相同的条件下进行多次重复性实验,最终的结果为各识别率的平均值. 实验的第 2 部分是测试 LRBP 对表情变化的识别能力. 其中训练集合为每个对象的 5 幅中性表情图像,共 615 张;测试集合为每个对象的 10 幅表情变化图像,共1 230 张.

实验前先采用文献[16]的方法将三维点云数据转化为深度图像,图像大小为 100 像素 × 80 像素,并采用 5 × 5 等分方式进行分割. 再按照上述数据分组方式,利用 LRBP 求取实验所用每幅深度图像的特征向量. 最后利用 Fisherface 方法对特征向量进行训练和识别. 实验均在 MATLAB 中实现,计算机 CPU 2.6 GHz,内存为 2 GB.

3.2 实验结果

3.2.1 识 别 率

在中性表情实验部分,为了检验参数 (P,R) 对实验结果的影响,每种方法分别进行了 9 组实验. 表 2 给出了 MS_LBP 和 LRBP 在 (P,R) 取不同值时的识别率.

表 2 中性表情识别率

Table 2 Recognition rates of neutral expressions

算子	$R=(R_1,R_2,\cdots,R_n)$	P			平均识别率/%
		8	10	12	
LRBP	(1,2,3)	96.26	96.48	96.29	95.90
	(1,2,3,4)	97.23	97.45	97.34	97.03
	(1,2,3,4,5)	97.67	98.05	97.99	97.79
MS_LBP	(1,2,3)	97.45	98.13	98.24	97.83
	(1,2,3,4)	97.67	98.18	98.32	97.98
	(1,2,3,4,5)	97.72	98.26	98.40	98.05

从表 2 的结果可以看出,对于中性表情的识别,LRBP 的识别率略低于 MS_LBP,平均低

北航学报 赠阅

1.04%，相差并不大. 在参数(P,R)改变时，两者的结果均有变化，这说明 P 和 R 对识别结果都是有影响的. 在 P 保持不变时，两者的识别率均随着邻域半径 R 的增加而提高，在 R = (1,2,3,4,5)时识别率最高；这是因为邻域半径数目的增多使得局部结构特征信息增加，从而提高了识别率. 在 R 保持不变时，LRBP 在 P = 10 时识别率较高，最高值为 98.05%；而 MS_LBP 在 P = 12 时识别率较高，最高值为 98.40%.

表 3 给出了 P = 8 时 LRBP 和 MS_LBP 对表情变化集合的识别结果.

从表 3 结果可以看出，在有表情识别方面，LRBP 与 MS_LBP 存在一定差距，识别率平均低 4.52%. 这可能是由于 LRBP 提取的低维特征向量造成了一定的邻域特征信息丢失，从而导致识别率降低. 从表中还可以看出，表情为“微笑”和“闭眼”时的识别率较高，与中性表情的识别结果较接近；而当表情为“大笑”“生气”和“吃惊”时，识别率则大幅降低. 这是因为“微笑”和“闭眼”两种表情造成的面部变形较小，与中性表情模板相似度高；而另外 3 种表情造成的面部变形较大，尤其是嘴部变形严重，从而影响了识别率. 与中性表情识别结果类似，两者在相同表情下的识别率都随着邻域半径数量的增加而有所提高.

表 3　表情变化实验结果

Table 3　Experimental results of expression varieties

算子	$R=(R_1,R_2,\cdots,R_n)$	微笑	大笑	生气	吃惊	闭眼	平均识别率/%
LRBP	(1,2,3)	94.31	69.11	76.42	76.42	93.50	81.95
	(1,2,3,4)	95.94	75.61	76.42	78.05	95.12	84.23
	(1,2,3,4,5)	97.56	75.61	79.67	78.05	95.93	85.36
MS_LBP	(1,2,3)	96.75	80.49	82.11	82.93	95.94	87.64
	(1,2,3,4)	97.97	81.30	82.52	83.74	96.34	88.37
	(1,2,3,4,5)	99.19	82.11	82.93	84.55	96.75	89.11

3.2.2　运算时间

为了检验 LRBP 的识别效率，对 LRBP 和 MS_LBP 在特征向量提取以及训练和识别时所用时间进行了统计.

图 4 给出了 LRBP 和 MS_LBP 对 1 845 幅人脸深度图像进行特征提取的总时间示意图. 由图可以看出两者在 P 相同的情况下，R 的增加会使特征提取时间变长，这是由特征向量维数增加导致的. 另外一个很明显的现象是随着 P 的增加，MS_LBP 提取特征的时间急剧增加，而 LRBP 的时间则变化较缓，P 为 8，10 和 12 时，LRBP 提取特征向量所需时间比 MS_LBP 平均缩短了 25.63%，58.06% 和 83.84%，这是因为 LRBP 提取的特征向量维数比 MS_LBP 的维数低.

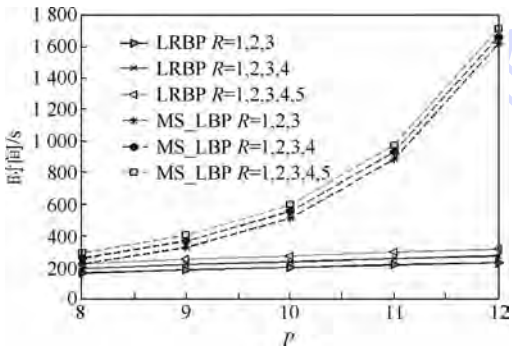


图 4　特征向量提取时间

Fig. 4　Time of extracting feature vectors

图 5 给出了对表情变化集合特征向量的训练和识别时间. 从图中可以看出，在相同参数条件下，LRBP 所用的训练和识别时间明显少于 MS_LBP. P 为 8，10 和 12 时，LRBP 比 MS_LBP 平均节省时间分别为 16.49%，24.33% 和 34.46%.

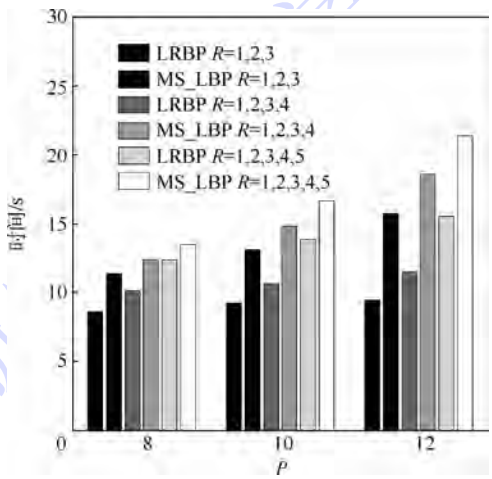


图 5　特征向量训练和识别时间

Fig. 5　Time of training and recognizing feature vectors

4　结　论

本文在 LBP 基础上提出了一种新的图像局部特征提取方法，经实验验证表明：①与 LBP 相比，在采样半径和邻域点数相同的情况下，LRBP 得到的特征向量的维数明显降低，可减轻后续计算的压力；②LRBP 在取得较好识别率的同时可大大缩减完成整个识别过程所用的时间.

致谢 感谢中国科学院自动化研究所提供的 CASIA-3D FaceV1 数据库.

参考文献（References）

[1] 明悦. 基于不变性特征的三维人脸识别研究[D]. 北京:北京交通大学,2013.
Ming Y. 3D face recognition based on invariant features[D]. Beijing:Beijing Jiaotong University,2013(in Chinese).

[2] 叶长明. 三维人脸识别中若干关键问题研究[D]. 合肥:合肥工业大学,2012.
Ye C M. Key issues in 3D face recognition[D]. Hefei:Hefei University of Technology,2012(in Chinese).

[3] Ardabilian M,Szeptycki P,Huang D,et al. 3D face recognition [J]. Signal and Image Processing for Biometrics,2012,30(2):89-115.

[4] Sharma P B,Goyani M M. 3D face recognition techniques-a review[J]. International Journal of Engineering Research and Applications (IJERA),2012,45(3):787-793.

[5] Lowe D. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision,2004,60(2):91-110.

[6] Chua C S,Han F,Ho Y K. 3D human face recognition using point signature[C]//Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition. Grenoble:IEEE,2000:233-238.

[7] Ojala T,Pietikäinen M,Harwood D. A comparative study of texture measures with classification based on featured distributions [J]. Pattern Recognition,1996,29(1):51-59.

[8] Zhang G,Wang Y. Multimodal 2D and 3D facial ethnicity classification[C]//Proceedings of Fifth IEEE International Conference on Image and Graphics. Piscataway,NJ:IEEE,2009:928-932.

[9] 孙鹤. 基于人脸面部特征的性别分类研究[D]. 上海:上海交通大学,2008.
Sun H. Gender classification based on human face facial features [D]. Shanghai:Shanghai Jiaotong University,2008(in Chinese).

[10] Zhao G Y,Pietikainen M. Dynamic texture recognition using local binary patterns with an application to facial expressions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2007,29(6):915-928.

[11] Vázquez H M,Reyes E G,Molleda Y C. A new image division for LBP method to improve face recognition under varying lighting conditions[C]//Proceedings of 19th International Conference on IEEE of Pattern Recognition. Piscataway,NJ:IEEE,2008:1-4.

[12] Shan C,Gong S,McOwan P W. Facial expression recognition based on local binary patterns:a comprehensive study[J]. Image and Vision Computing,2009,27(1):803-816.

[13] Guo Z,Zhang D. A completed modeling of local binary pattern operator for texture classification[J]. IEEE Transactions on Image Processing,2010,19(6):1657-1663.

[14] Yan S,Wang H,Tang X O,et al. Exploring feature descriptors for face recognition[C]//Proceedings of IEEE International Conference on Acoustics Speech and Signal. Piscataway,NJ:IEEE,2007:629-632.

[15] Ojala T,Pietikainen M,Maenpaa T. Multi-resolution grayscale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2002,24(7):971-987.

[16] Li L,Xu C H,Tang W,et al. 3D face recognition by constructing deformation invariant image[J]. Pattern Recognition Letters,2008,29(5):1596-1602.

[17] Belhumeur P N,Hespanha J P,Kriegman D. Eigenfaces vs. fisherfaces:recognition using class specific linear projection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1997,19(7):711-720.

3D face recognition based on local radial binary pattern

LIU Qing, SUN Junhua*

(School of Instrumentation Science and Opto-electronics Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: An operator named local radial binary pattern (LRBP) was proposed for extracting local features in face recognition. The binary sequence encoding scheme of the LRBP is different from that of the local binary pattern. Firstly, the proposed LRBP operator was used in 3D face recognition. 3D face depth images were preprocessed and divided into subregions. Then the signature sequences of the subregions were extracted by the LRBP operator. The feature vectors of a 3D face depth image were obtained by connecting the signature sequences of all the subregions of the image. Finally, the 3D face feature vectors were trained and recognized using the Fisherface method. Experiments were conducted using the 3D face database of Institute of Automation, Chinese Academy of Sciences. The results show that the proposed method can effectively promote the efficiency of 3D face recognition without reducing the recognition rates.

Key words: pattern recognition; 3D face recognition; local feature; binary pattern; local radial binary pattern

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0305

电磁航天器编队动力学建模与运动规划方法

邵龙飞, 师鹏*, 赵育善

(北京航空航天大学 宇航学院, 北京 100191)

摘 要: 基于电磁航天器作用原理,利用拉格朗日方法,建立电磁航天器"绳系"动力学模型.基于偏差线性化动力学模型,以切向、径向编队为例,分析编队构型保持稳定性,设计构型保持控制律;将双星电磁航天器编队(EMFF, Electromagnetic Formation Flight)重构运动规划问题转化为标准优化问题,利用高斯伪谱优化方法进行求解.并提出序列控制策略,将该双星模型扩展应用于多星电磁编队构型重构问题,转化为多阶段运动规划问题利用多阶段优化方法进行求解.仿真结果表明本文的动力学建模方法和控制方法是可行的.

关 键 词: 电磁航天器; 编队飞行; 构型重构; 序列控制; 运动规划

中图分类号: V412.4

文献标识码: A **文章编号**: 1001-5965(2015)04-0737-07

为解决常规推力航天器编队的不足(推进剂的消耗影响编队飞行的寿命;推力器排出的羽流可能会对邻近航天器造成冲击,或对其光学器件造成污染,影响成像等),人们提出了采用电磁作用系统,通过航天器之间电磁力及电磁力矩来实现编队控制[1].

电磁航天器的研究始于2000年[1],Miller和Kong[2-4]等首先提出了电磁编队飞行的概念,对其优点、理论基础、实施方案、可行性及设计等进行了分析.文献[5]建立了近地轨道电磁编队动力学模型,设计了自适应姿轨控制律.文献[6]考虑电磁编队位置跟踪的输入延时和扰动不确定因素,设计了滑模变结构控制律.杨乐平等[7-9]对电磁编队保持及构型稳定性进行了研究.文献[8]基于Kane方法建立了双星电磁编队6自由度姿轨耦合动力学模型,分析了构型稳定性.文献[9]分析了三星电磁编队相对构型及稳定性.文献[10]基于人工势能函数方法,对电磁编队运动轨迹的规划与控制进行了研究.Inampudi等[11-12]给出了近地轨道和拉格朗日点附近的双星库仑编队动力学模型,对编队轨迹优化进行研究.2013年,SPPL实验室建立了共振感应近场产生系统(RINGS, Resonant Inductive Near-field Generation System),在国际空间站进行了在轨实验,验证了电磁力推进和无线能源传送技术[13-14].

本文建立双星电磁航天器"绳系"动力学,并对其构型稳定性及保持进行分析,将电磁编队重构转化为运动规划问题,提出序列控制策略来解决高非线性多星电磁编队问题,最后通过多星电磁编队仿真验证模型及控制的有效性.

1 电磁航天器编队动力学模型

1.1 电磁系统模型

电磁航天器通过安装一套三轴正交通电线圈,改变线圈电流产生电磁作用力,来实现航天器编队相对位置控制.基于电磁远场模型,电磁航天

收稿日期: 2014-05-28; **录用日期**: 2014-09-01; **网络出版时间**: 2014-09-19 17:59
网络出版地址: www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0305.html
基金项目: 国家自然科学基金资助项目(11102007); 中央高校基本科研业务费资助项目(YWF-14-YHXY-012)
作者简介: 邵龙飞(1989—),男,陕西汉中人,硕士生,shaolongfei@sina.cn
* **通讯作者**: 师鹏(1981—),男,陕西咸阳人,讲师,shipeng@buaa.edu.cn,主要研究方向为航天器动力学与控制.

引用格式: 邵龙飞, 师鹏, 赵育善. 电磁航天器编队动力学建模与运动规划方法[J]. 北京航空航天大学学报, 2015, 41(4): 737-743. Shao L F, Shi P, Zhao Y S. Dynamics modeling and motion programming for electromagnetic formation flight[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(4): 737-743 (in Chinese).

器 B 在 $\boldsymbol{l}$ 位置处产生的磁场强度[5]为

$$\boldsymbol{B}_B(\boldsymbol{l})=\frac{\mu_0}{4\pi}\left(\frac{3\boldsymbol{\mu}_B\cdot\boldsymbol{l}}{\|\boldsymbol{l}\|^5}\boldsymbol{l}-\frac{\boldsymbol{\mu}_B}{\|\boldsymbol{l}\|^3}\right) \tag{1}$$

式中，$\boldsymbol{\mu}_B$ 为 B 的磁偶极子强度；$\mu_0=4\pi\times10^{-7}$ N/A^2 为真空磁导率. 电磁航天器 A 在 B 磁场中受到的电磁力[5]为

$$\boldsymbol{F}_{AB}=-\frac{3\mu_0}{4\pi}\left(-\frac{\boldsymbol{\mu}_A\cdot\boldsymbol{\mu}_B}{\|\boldsymbol{l}\|^5}\boldsymbol{l}-\frac{\boldsymbol{\mu}_A\cdot\boldsymbol{l}}{\|\boldsymbol{l}\|^5}\boldsymbol{\mu}_B-\frac{\boldsymbol{\mu}_B\cdot\boldsymbol{l}}{\|\boldsymbol{l}\|^5}\boldsymbol{\mu}_A+5\frac{(\boldsymbol{\mu}_A\cdot\boldsymbol{l})(\boldsymbol{\mu}_B\cdot\boldsymbol{l})}{\|\boldsymbol{l}\|^7}\boldsymbol{l}\right) \tag{2}$$

式中，$\boldsymbol{\mu}_A$ 为 A 的磁偶极子强度.

由电磁力解算线圈磁矩是一个复杂的非线性问题，一般可以通过添加约束来简化计算：固定某个电磁航天器的磁矩（如使 $\boldsymbol{\mu}_B$ 始终沿着 AB 连线方向，大小保持不变）或者限制双星能量消耗均衡（如使 $\boldsymbol{\mu}_A=\boldsymbol{\mu}_B$，则存在解析解）.

电磁力和电磁力矩是耦合产生的，会引起电磁航天器姿轨耦合. 为了突出研究电磁编队构型重构问题，本文暂不考虑姿态耦合，认为已通过其他姿控技术保证姿态稳定或机动任务.

1.2 双星电磁航天器相对动力学建模

在不考虑地磁场和其他摄动影响的情况下，星间电磁力属于内力，电磁控制不会改变系统质心位置，假设编队系统质心保持 Kepler 轨道运行.

如图 1 所示，春分点地心惯性坐标系记作 S_i；编队质心第 2 轨道坐标系记作 S_o；“绳系”坐标系 S_t 固连在电磁航天器 A，B 间假想的系绳上，可由 S_o 系先绕 y 轴转 α 角，再绕 x 轴转 β 角得到（将 A，B 看作质点，忽略 z_t 轴的旋转变化）. 相关位置矢量 $\boldsymbol{r}_{O_c}$，$\boldsymbol{r}_A$，$\boldsymbol{r}_B$ 等定义见图 1. S_o 系到 S_t 系坐标变换矩阵为

$$\boldsymbol{L}_{to}=\boldsymbol{L}_x(\beta)\boldsymbol{L}_y(\alpha)$$

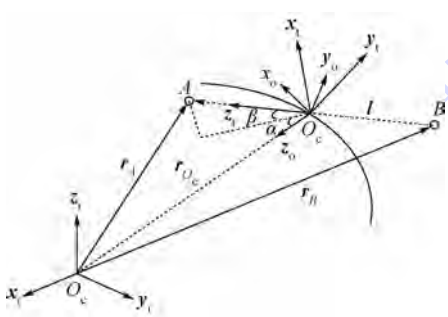


图 1 电磁航天器编队几何关系示意图

Fig. 1 Geometric model of EMFF system

设编队系统拉格朗日函数为

$$L=T-V_g-V_e$$

式中，T 为动能；V_g 为重力势能；V_e 为电磁场势能.

取广义坐标 $\boldsymbol{q}=[\alpha\quad\beta\quad l]^{\mathrm{T}}$，拉格朗日动力学方程为

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial\dot{\boldsymbol{q}}}-\frac{\partial L}{\partial\boldsymbol{q}}=\boldsymbol{R} \tag{3}$$

式中，$\boldsymbol{R}$ 为非有势广义力，取 $\boldsymbol{R}=\boldsymbol{0}$.

A，B 相对系统质心位置矢量为 $\boldsymbol{l}_1$ 和 $\boldsymbol{l}_2$，质量为 m_1 和 m_2，满足 $m_1\boldsymbol{l}_1+m_2\boldsymbol{l}_2=\boldsymbol{0}$，令 $\boldsymbol{l}=\boldsymbol{l}_1-\boldsymbol{l}_2$；即可得 $\boldsymbol{l}_1$ 及 $\boldsymbol{l}_1$ 在 S_o 系下相对变化率为

$$\begin{cases}\boldsymbol{l}_1^o=\boldsymbol{L}_{ot}\boldsymbol{l}_1^t=\\ \dfrac{m_2 l}{m_1+m_2}[\sin\alpha\cos\beta\quad-\sin\beta\quad\cos\alpha\cos\beta]^{\mathrm{T}}\\ \dfrac{\mathrm{d}\boldsymbol{l}_1}{\mathrm{d}t}\Big|_o=\begin{bmatrix}\dot{x}_1\\ \dot{y}_1\\ \dot{z}_1\end{bmatrix}=\dfrac{m_2}{m_1+m_2}\cdot\\ \begin{bmatrix}\dot{l}\sin\alpha\cos\beta+l\cos\alpha\cos\beta\dot{\alpha}-l\sin\alpha\sin\beta\dot{\beta}\\ -\dot{l}\sin\beta-l\cos\beta\dot{\beta}\\ \dot{l}\cos\alpha\cos\beta-l\sin\alpha\cos\beta\dot{\alpha}-l\cos\alpha\sin\beta\dot{\beta}\end{bmatrix}\end{cases} \tag{4}$$

根据速度合成定理，A 的绝对运动速度为

$$\frac{\mathrm{d}\boldsymbol{r}_A}{\mathrm{d}t}=\frac{\mathrm{d}(\boldsymbol{r}_{O_c}+\boldsymbol{l}_1)}{\mathrm{d}t}\Big|_o+\boldsymbol{\Omega}\times(\boldsymbol{r}_{O_c}+\boldsymbol{l}_1)=\begin{bmatrix}\dot{x}_1-\dot{\theta}z_1+\dot{\theta}r_{O_c}\\ \dot{y}_1\\ \dot{z}_1+\dot{\theta}x_1-\dot{r}_{O_c}\end{bmatrix} \tag{5}$$

式中，质心轨道角速度 $\boldsymbol{\Omega}=[0\quad-\dot{\theta}\quad0]^{\mathrm{T}}$.

同理可求出 B 的绝对速度，双星电磁航天器编队系统的动能可以整理为

$$T=\frac{1}{2}m_1\|\dot{\boldsymbol{r}}_A\|^2+\frac{1}{2}m_2\|\dot{\boldsymbol{r}}_B\|^2=\frac{1}{2}\bar{m}(\dot{l}^2+l^2(\dot{\beta}^2+\cos^2\beta(\dot{\alpha}-\dot{\theta})^2))+\frac{1}{2}(m_1+m_2)(\dot{\theta}^2r_{O_c}^2+\dot{r}_{O_c}^2) \tag{6}$$

式中

$$\bar{m}=\frac{m_1m_2}{m_1+m_2}$$

系统重力势能（泰勒展开，保留二阶项）为

$$V_g=-\frac{\mu m_1}{\|\boldsymbol{r}_{O_c}+\boldsymbol{l}_1\|}-\frac{\mu m_2}{\|\boldsymbol{r}_{O_c}+\boldsymbol{l}_2\|}=-\frac{\mu}{r_{O_c}}(m_1+m_2)+\frac{\mu}{r_{O_c}^3}\frac{\bar{m}l^2}{2}(1-3\cos^2\beta\cos^2\alpha) \tag{7}$$

编队系统电磁场势能为

$$V_e = V_{AB}(\boldsymbol{\mu}_A, \boldsymbol{l}) = -\boldsymbol{\mu}_A \boldsymbol{B}_B(\boldsymbol{l}) \tag{8}$$

由式(3)可整理得双星电磁航天器近地编队相对动力学模型为

$$\begin{cases} \ddot{\alpha}\cos^2\beta - 2(\dot{\alpha}-\dot{\theta})\dot{\beta}\sin\beta\cos\beta + \\ \quad 2\dfrac{\dot{l}}{l}(\dot{\alpha}-\dot{\theta})\cos^2\beta + \\ \quad 3\dfrac{\mu}{r_{O_c}^3}\sin\alpha\cos\alpha\cos^2\beta - \ddot{\theta}\cos^2\beta = \dfrac{Q_\alpha}{\bar{m}l^2} \\ \ddot{\beta} + 2\dfrac{\dot{l}}{l}\dot{\beta} + \\ \quad \left[(\dot{\alpha}-\dot{\theta})^2 + 3\dfrac{\mu}{r_{O_c}^3}\cos^2\alpha\right]\sin\beta\cos\beta = \dfrac{Q_\beta}{\bar{m}l^2} \\ \ddot{l} - \left[\dot{\beta}^2 + (\dot{\alpha}-\dot{\theta})^2\cos^2\beta + \right. \\ \quad \left.\dfrac{\mu}{r_{O_c}^3}(3\cos^2\alpha\cos^2\beta - 1)\right] l = \dfrac{Q_l}{\bar{m}} \end{cases} \tag{9}$$

式中，$Q_\alpha = F_\alpha l, Q_\beta = F_\beta l, Q_l = F_l$，为广义力. 为简化计算，设 $\boldsymbol{\mu}_B$ 始终沿着 z_t 轴方向，$\boldsymbol{\mu}_B^t = [0\ 0\ \mu_{Bl}]^T$，电磁力在 S_t 系下可简化为

$$\boldsymbol{F}_{AB}^t = \begin{bmatrix} F_\alpha \\ F_\beta \\ F_l \end{bmatrix} = \frac{3\mu_0\mu_{Bl}}{4\pi l^4}\begin{bmatrix} \mu_{A\alpha} \\ \mu_{A\beta} \\ -2\mu_{Al} \end{bmatrix} \tag{10}$$

2　动力学模型分析

2.1　动力学特性分析

系统模型式(9)具有强非线性和耦合性，使得分析控制过程较为复杂. 但是式(9)具有以下优势：

1) 以双星电磁航天器编队构型参数 α, β, l 为变量，有利于分析编队整体构型的变化规律. 由 α, β 定义，可得双星“绳系”系统相对于 S_o 系的角速度为

$$\boldsymbol{\omega}^t = \dot{\boldsymbol{\beta}}^t + \boldsymbol{L}_x(\beta)\dot{\boldsymbol{\alpha}}^t = [\dot{\beta}\quad \dot{\alpha}\cos\beta\quad -\dot{\alpha}\sin\beta]^T \tag{11}$$

可以直接分析编队构型运动情况.

2) 可用于电磁航天器绳系编队分析.

前文建模过程中，双星电磁航天器间并不存在真实的系绳的连接. 双星间若存在真实系绳连接，式(9)中第 3 个关于 l 式可由绳系装置进行解耦：可以减少绳系方向电磁消耗，节约电能；便于构型重构与保持控制；电磁双星交会过程中，距离过近，电磁远场模型误差较大，可利用系绳低速拉回，完成交会任务.

2.2　模型线性化与稳定性分析

考虑 α, β 和 l 存在小偏差：

$$\begin{cases} \alpha = \alpha_0 + \Delta\alpha \\ \beta = \beta_0 + \Delta\beta \\ l = l_0 + \Delta l \end{cases} \tag{12}$$

以双星电磁航天器近地圆轨道径向、切向编队构型保持为例，研究系统式(9)构型保持稳定性.

$$\begin{cases} \alpha \to \alpha_0; \dot{\alpha}, \ddot{\alpha} \to 0 \\ \beta \to \beta_0; \dot{\beta}, \ddot{\beta} \to 0 \\ l \to l_0; \dot{l}, \ddot{l} \to 0 \end{cases} \tag{13}$$

为便于系统径向编队保持($\alpha_0 = 0°, \beta_0 = 0°$, $l = l_0$)稳定性分析与控制，将式(12)代入式(9)，泰勒展开进行线性化可得：

$$\begin{cases} \Delta\ddot{\alpha} - 2\dfrac{\Omega}{l_0}\Delta\dot{l} + 3\Omega^2\Delta\alpha = u_\alpha \\ \Delta\ddot{\beta} + 4\Omega^2\Delta\beta = u_\beta \\ \Delta\ddot{l} + 2\Omega l_0\Delta\dot{\alpha} - 3\Omega^2\Delta l - 3\Omega^2 l_0 = u_l \end{cases} \tag{14}$$

显然，式(14)中，$\Delta\beta$ 是解耦的，其零控状态临界稳定；$\Delta\alpha$ 和 Δl 相耦合，其零控特征方程为 $\lambda^4 + 4\Omega^2\lambda^2 - 9\Omega^4 = 0$，可得 $\lambda^2 = \sqrt{13}\Omega - 2\Omega^2$. 近地轨道恒有 $\lambda^2 > 0$，则 λ 存在实部为正的根.

根据李雅普诺夫线性化定理，可得原系统径向编队构型不稳定. 需要设计控制律来保持构型稳定. 对式(14)选取反馈控制：

$$\begin{cases} u_\alpha = -2\dfrac{\Omega}{l_0}\Delta\dot{l} + 3\Omega^2\Delta\alpha - f_1\Delta\dot{\alpha} - f_2\Delta\alpha \\ u_\beta = 4\Omega^2\Delta\beta - f_3\Delta\dot{\beta} - f_4\Delta\beta \\ u_l = 2\Omega l_0\Delta\dot{\alpha} - 3\Omega^2\Delta l - 3\Omega^2 l_0 - f_5\Delta\dot{l} - f_6\Delta l \end{cases} \tag{15}$$

根据赫尔维茨稳定判据，二阶线性系统稳定充要条件为特征方程各项系数为正. 将控制 $\boldsymbol{u}$ 代入式(14)整理，$f_1 \sim f_6 > 0$ 时，就可保证系统式(14)严格稳定，从而原系统式(9)在零点附近局部渐进稳定.

同理可对切向编队保持($\alpha_0 = 90°, \beta_0 = 0°$, $l = l_0$)模型线性化、稳定分析及设计控制律.

仿真算例 1　设 $m_A = m_B = 100$ kg，考虑电磁系统的可实现性，$\boldsymbol{\mu}_B$ 始终沿着 z_t 方向，大小为 10^5 H/m，$\boldsymbol{\mu}_A$ 的各个轴分量 $\mu_i (i = \alpha, \beta, l)$ 不超过 4×10^4 H/m，系统质心初始轨道根数见表 1. 切向编队 $l_0 = 25$ m，$\alpha_0 = \pi/2$ rad，$\beta_0 = 0$ rad；径向编队 $l_0 = 25$ m，$\alpha_0 = \beta_0 = 0$ rad. 初始偏差均为：$\Delta l = 1$ m，$\Delta\alpha = -0.1$ rad，$\Delta\beta = 0.1$ rad.

北京航空航天大学学报

北航学报 赠阅

表 1 初始时刻编队质心轨道根数

Table 1 Initial orbital elements of mass center of EMFF system

参数	a/km	e	i/(°)	Ω/(°)	ω/(°)	θ/(°)
取值	6878.1	0	0	0	0	0

仿真结果见图 2 和图 3,仿真曲线表明当仿真时间较短、α 和 β 偏差较小时,线性化模型能很好地代替非线性模型. 利用设计的控制律,500 s 内将 l,α,β 误差控制到 0,有良好的鲁棒性.

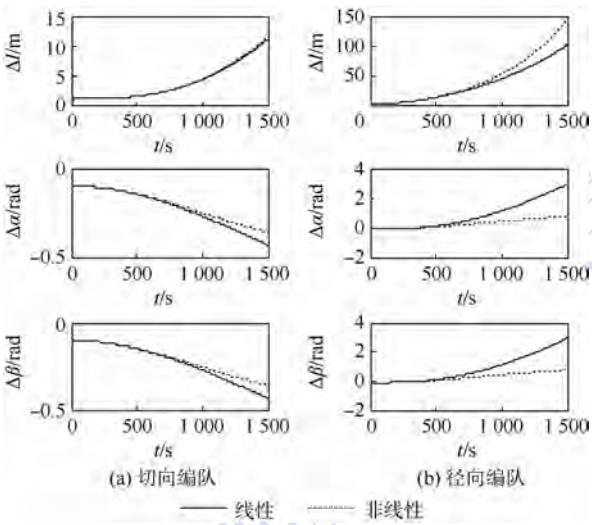


图 2 电磁编队线性化与非线性动力学模型比较

Fig. 2 Comparison between linear and nonlinear dynamics models for EMFF

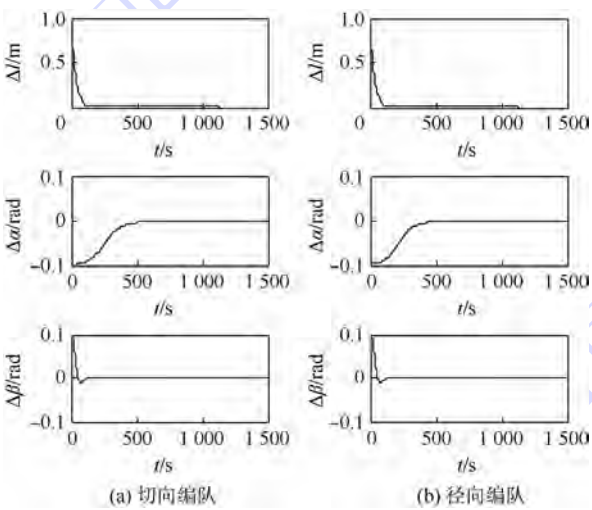


图 3 电磁编队构型保持状态误差变化曲线

Fig. 3 State parameter error variations of EMFF system

3 电磁航天器编队运动规划

3.1 基于高斯伪谱法的电磁航天器运动规划

电磁航天器编队重构问题,实际上可以转化为运动轨迹规划问题进行求解.

定义状态变量为 $\boldsymbol{x}=[\alpha \quad \dot{\alpha} \quad \beta \quad \dot{\beta} \quad l \quad \dot{l}]^{\mathrm{T}}$,控制磁矩为 $\boldsymbol{u}=[\mu_\alpha \quad \mu_\beta \quad \mu_l]^{\mathrm{T}}$,则“绳系”动力学系统模型式(9)可整理为 $\dot{\boldsymbol{x}}=f(\boldsymbol{x},\boldsymbol{u})$;考虑电磁系统的可实现性及防碰撞,约束条件为 $|\mu_i|\leqslant\mu_{\max}$;$|l|\geqslant l_{\min}$,初始和终端约束条件为 $\boldsymbol{x}(t_0)=\boldsymbol{x}_0$,$\boldsymbol{x}(t_{\mathrm{f}})=\boldsymbol{x}_{\mathrm{f}}$. 电磁航天器运动优化性能指标可以选为:

1) 时间性能指标.

$$J=\int_{t_0}^{t_{\mathrm{f}}}\mathrm{d}t \tag{16}$$

2) 重构过程平缓的性能指标.

$$J=\int_{t_0}^{t_{\mathrm{f}}}\ddot{l}^2\,\mathrm{d}t \tag{17}$$

3) 系统能量性能指标.

电磁编队重构过程中,设 $\boldsymbol{\mu}_B$ 保持不变,只需考虑 A 星电磁能消耗情况. 带电线圈磁矩 $\boldsymbol{\mu}$ 可表示为 $\boldsymbol{\mu}=NIS\boldsymbol{n}$,其中 N 为线圈匝数,I 为电流,S 为线圈面积,$\boldsymbol{n}$ 为载流平面法线方向单位向量,服从右手定则. 带电线圈固装在电磁航天器本体上,磁矩 $\boldsymbol{\mu}$ 在 S_{t} 系和本体系的投影存在如下关系:

$$\boldsymbol{\mu}^{\mathrm{t}}=\boldsymbol{L}_{\mathrm{tb}}\boldsymbol{\mu}^{\mathrm{b}}\Rightarrow\|\boldsymbol{\mu}^{\mathrm{t}}\|^2=\|\boldsymbol{\mu}^{\mathrm{b}}\|^2=\sum_{i=1}^{3}N_i^2I_i^2S_i^2 \tag{18}$$

电功率为 $P=I^2R$,利用 $\boldsymbol{\mu}^{\mathrm{t}}$ 和 $\boldsymbol{\mu}^{\mathrm{b}}$ 计算出的电功率之和是相等的. 所以系统能量性能指标可选为

$$J=\int_{t_0}^{t_{\mathrm{f}}}(P_{A\alpha}+P_{A\beta}+P_{Al})\,\mathrm{d}t=$$
$$K_{\mathrm{e}}\int_{t_0}^{t_{\mathrm{f}}}(\mu_{A\alpha}^2+\mu_{A\beta}^2+\mu_{Al}^2)\,\mathrm{d}t \tag{19}$$

基于式(9),将双星电磁航天器编队运动规划问题转化成标准优化问题. 高斯伪谱优化方法具有求解精度高和收敛速度快等优点[15],所以本文采用高斯伪谱方法对该优化控制问题进行求解,求得有限点处的控制近似解,再利用数值插值获得电磁航天器在重构过程中的控制磁矩 $\boldsymbol{u}(t)$,具体算法及推导过程参见文献[15].

仿真算例 2 假设 $m_A=100$ kg 和 $m_B=200$ kg,l 由 25 m 重构变为 15 m,t_0 和 t_{f} 时刻其他构型状态都为 0. 为了避免碰撞,限定 l 不小于 10 m. 初始时刻 AB 编队系统质心轨道根数见表 1. 高斯伪谱法优化指标采用电能消耗最优.

仿真结果如图 4 和图 5 所示,两星距离由 25 m平滑地变化到 15 m,速度不超过 0.01 m/s,α,β 及其变化率满足 t_{f} 时刻状态要求. 由于采用能量最优为优化指标,各轴磁矩变化都比较平滑.

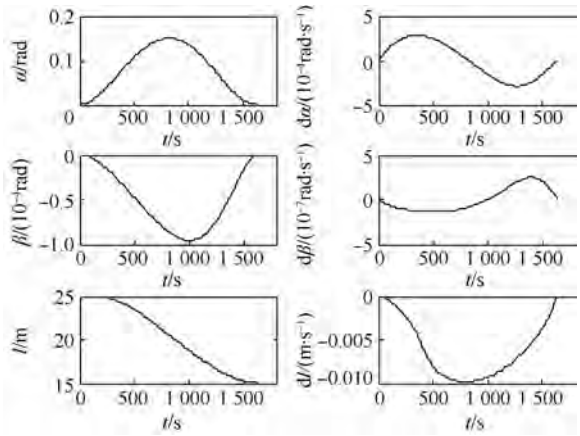


图 4　双星电磁编队构型状态参数变化

Fig. 4　State parameter variations of two-craft EMFF reconfiguration

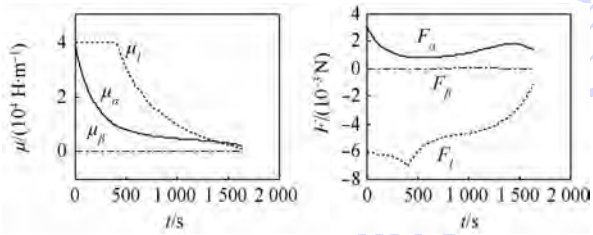


图 5　电磁编队控制磁矩与电磁控制力

Fig. 5　Electromagnetic control (magnetic dipole and force) of two-EMFF reconfiguration

3.2　序列控制

式(9)形式上只适用于双星电磁航天器编队，本文对是否可以利用式(9)完成复杂的多电磁航天器编队控制任务进行研究. 提出序列控制策略，按照某个顺序，通过两两控制，逐一将各电磁航天器控制到目标状态.

设电磁航天器编队系统由 N 个航天器组成，其相对位置矢量和质量分别为 $\boldsymbol{r}_i$ 和 m_i. 由于电磁作用不改变编队系统质心位置，即

$$\sum_{i=1}^{N} m_i \boldsymbol{r}_i^0 = \sum_{i=1}^{N} m_i \boldsymbol{r}_i^{\mathrm{f}} \tag{20}$$

不失一般性，选择“1 号”航天器通过电磁力作用，逐个将其余控制到目标位置：

$$\begin{cases} m_1 \boldsymbol{r}_1^0 + m_2 \boldsymbol{r}_2^0 = m_1 \boldsymbol{r}_1^{1,2} + m_2 \boldsymbol{r}_2^{\mathrm{f}} \\ m_1 \boldsymbol{r}_1^{1,2} + m_3 \boldsymbol{r}_3^0 = m_1 \boldsymbol{r}_1^{1,2,3} + m_3 \boldsymbol{r}_3^{\mathrm{f}} \\ \vdots \\ m_1 \boldsymbol{r}_1^{1,2,\cdots,N-1} + m_N \boldsymbol{r}_N^0 = m_1 \boldsymbol{r}_1^{1,2,\cdots,N} + m_N \boldsymbol{r}_N^{\mathrm{f}} \end{cases} \tag{21}$$

经过分析可得 $\boldsymbol{r}_1^{1,2,\cdots,N} = \boldsymbol{r}_1^{\mathrm{f}}$，最终“1 号”航天器自然能够到达其期望位置. “1 号”航天器在重构过程中可自由选择. 由此，序列控制的重点在于如何选择控制对、控制优先级和确定控制方案的标准. 例如，可以把控制时间最短、最节省能量或其他特殊任务约束、环境条件、航天器状态等因素作为约束指标. 另外，电磁航天器相对作用距离较短(100 m 以内)，为了避免星间碰撞，重构过程必须考虑防碰撞策略.

例如图 6，三星电磁系统编队构型重构过程的序列控制次序选择为：$[t_0, t_{\mathrm{m}}]$，利用 C 将 A 控制到目标状态；$[t_{\mathrm{m}}, t_{\mathrm{f}}]$，利用 C 将 B 控制到目标状态. 从而通过序列控制策略将多星电磁航天器编队运动规划问题转化为多阶段运动规划问题. 基于 Rao 等[16]给出求解多段优化问题的 GPOPS (Gauss Pseudospectral Optimization Software)算法，本文将多阶段运动规划转化为多阶段优化问题进行求解，算法推导详见文献[16].

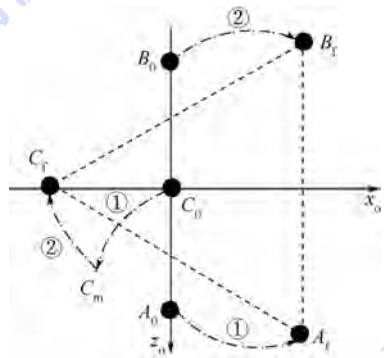


图 6　构型重构序列控制示意图

Fig. 6　Sequence control of multi-reconfiguration maneuver

仿真算例 3　考虑 A, B, C 三星系统，$m_A = m_B = 100\,\mathrm{kg}$，$m_C = 200\,\mathrm{kg}$，轨道根数见表 1，重构过程如图 6所示；$\boldsymbol{\mu}_C$ 始终沿着相应的 z_{t} 方向，恒为 8×10^4 H/m；$\boldsymbol{\mu}_A$ 和 $\boldsymbol{\mu}_B$ 在 S_{t} 系各个轴磁矩 $\mu_i (i = \alpha, \beta, l)$ 最大不超过 4×10^4 H/m. 为了避免碰撞，限定 A, B, C 两两距离 l 要不少于 10 m. 采用高斯伪谱法进行序列控制轨迹规划，选时间最优为优化指标.

仿真结果如图 7 ~ 图 9 所示，编队重构过程分为两个阶段(仿真结果为 $t_0 = 0\,\mathrm{s}$，$t_{\mathrm{m}} = 470\,\mathrm{s}$，$t_{\mathrm{f}} = 1\,568.5\,\mathrm{s}$)：$[t_0, t_{\mathrm{m}}]$ 段 A, C 相互作用，使 A 到达目标状态；$[t_{\mathrm{m}}, t_{\mathrm{f}}]$ 段 B, C 相互作用，使 A 到达目标状态. 控制磁矩几乎都达到了最大 4×10^4 H/m，呈现出 Bang-Bang 控制的特点，这与选时间最优为性能指标相吻合.

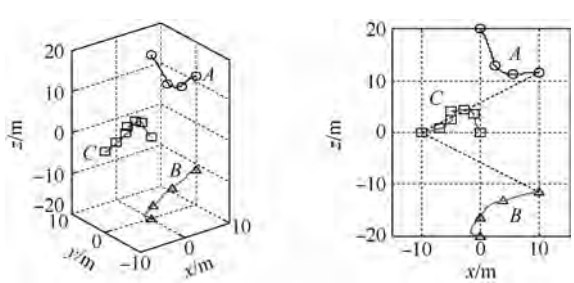


图 7　电磁航天器构型重构优化路径示意图

Fig. 7　Optimal trajectories of multi-EMFF reconfiguration

北航学报 赠阅

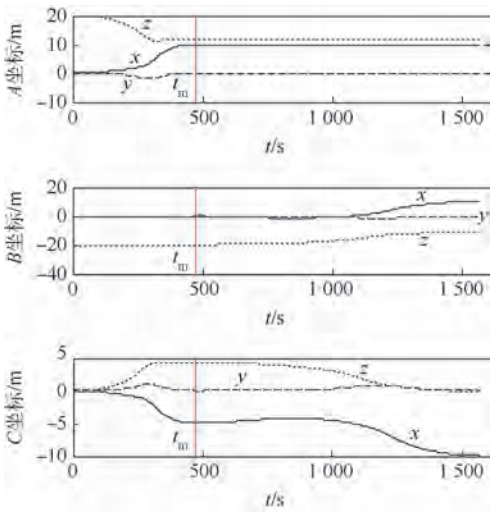


图 8 A,B 和 C 在重构变换过程中坐标变化

Fig. 8 Position coordinates of multi-EMFF reconfiguration

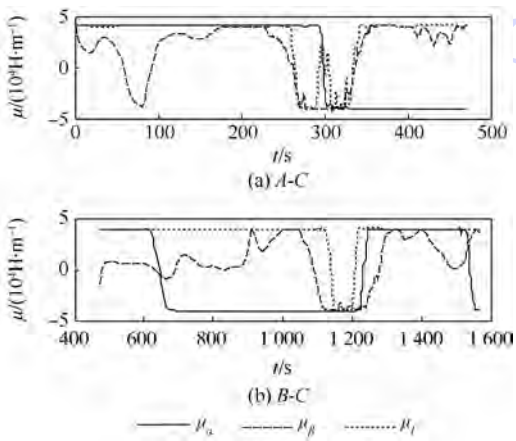


图 9 A 和 B 在重构变换过程中控制磁矩的变化

Fig. 9 Histories of magnetic moments in reconfiguration

4 结 论

本文建立了双星电磁航天器“绳系”动力学模型,利用线性化模型分析了电磁编队构型的稳定性,提出了序列控制策略,将多星电磁编队运动规划转化为多阶段优化问题.

1) 短时间、小偏差下,偏差线性化模型可以很好地代替原模型,可以简化模型的分析与控制;

2) 通过序列控制策略,利用双星模型可以有效地解决复杂高非线性多星电磁编队控制问题.

参考文献 (References)

[1] 王龙,杨乐平,许军校. 电磁编队飞行与电磁交会对接关键技术及进展[J]. 装备指挥技术学院学报,2009,20(1):74-78.
Wang L, Yang L P, Xu J X. The key technology and development of electromagnetic formation flight and electromagnetic rendezvous docking[J]. Journal of the Academy of Equipment Command & Technology, 2009, 20(1): 74-78(in Chinese).

[2] Miller D W, Sedwick R J, Eliaset L, et al. NIAC phase I final report: electromagnetic formation flight[J]. Final Raport Massachusetts Institute of Technology, 2002, 121: 163-179.

[3] Kong E, Kwon D W, Schweighartet S A, et al. Electromagnetic formation flight for multisatellite arrays[J]. Journal of Spacecraft and Rockets, 2004, 41(4): 659-666.

[4] Ahsun U. Dynamics and control of electromagnetic satellite formations[D]. Massachusetts: Massachusetts Institute of Technology, 2007.

[5] Ahsun U, Miller D W, Ramirez L. Control of electromagnetic satellite formations in near-earth orbits[J]. Journal of Guidance, Control, and Dynamics, 2010, 33(6): 1883-1891.

[6] 苏建敏,董云峰. 电磁卫星编队位置跟踪滑模变结构控制[J]. 宇航学报,2011,32(5):1093-1099.
Su J M, Dong Y F. Sliding mode variable structure control for electromagnetic satellite formation station-tracking[J]. Journal of Astronautics, 2011, 32(5): 1093-1099(in Chinese).

[7] Cai W W, Yang L P, Zhu Y W, et al. Formation keeping control through inter-satellite electromagnetic force[J]. Science China Technological Sciences, 2013, 56(5): 1102-1111.

[8] 黄涣,杨乐平,朱彦伟,等. 双星电磁编队的动力学平衡态稳定性与控制[J]. 国防科技大学学报,2013,35(3):12-17.
Huang H, Yang L P, Zhu Y W, et al. Stability and control of dynamics equilibrium for two-spacecraft electromagnetic formation[J]. Journal of National University of Defense Technology, 2013, 35(3): 12-17(in Chinese).

[9] Huang H, Zhu Y W, Yang L P, et al. Stability and shape analysis of relative equilibrium for three-spacecraft electro- magnetic formation[J]. Acta Astronautica, 2014, 94(1): 116-131.

[10] 苏建敏,董云峰. 利用人工势函数法的卫星电磁编队控制[J]. 北京航空航天大学学报,2012,38(2):213-217.
Su J M, Dong Y F. Artificial potential function method for satellite electromagnetic formation control[J]. Journal of Beijing University of Aeronautics and Astronautics, 2012, 38(2): 213-217(in Chinese).

[11] Inampudi R, Hanspeter S. Optimal reconfigurations of two-craft coulomb formation in circular orbits[J]. Journal of Guidance, Control, and Dynamics, 2012, 35(6): 1805-1815.

[12] Inampudi R, Hanspeter S. Orbit radial dynamic analysis of two-craft Coulomb formation at libration points[J]. Journal of Guidance, Control, and Dynamics, 2014, 37(2): 682-691.

[13] Alinger D J. System analysis and design for the resonant inductive near-field generation system(RINGS)[D]. Maryland: University of Maryland College Park, 2013.

[14] Porter A K, Alinger D J, Sedwick R J, et al. Dual-purpose resonate actuators for electromagnetic formation flight and wireless power transfer[C]//AIAA Guidance, Navigation, and Control Conference. Reston: AIAA, 2014: 1-18.

[15] Benson D. A Gauss pseudospectral transcription for optimal control[D]. Massachusetts: Massachusetts Institute of Technology, 2005.

[16] Rao A V, Benson D A, Darby C, et al. Algorithm 902: GPOPS, a MATLAB software for solving multiple-phase optimal control problems using the Gauss pseudospectral method[J]. ACM Transactions on Mathematical Software, 2010, 37(2): 22: 1-22: 39.

北航学报 赠阅

Dynamics modeling and motion programming for electromagnetic formation flight

SHAO Longfei, SHI Peng*, ZHAO Yushan

(School of Astronautics, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: Based on the mechanism of electromagnetic spacecraft, the relative translational "tethered" dynamics model for two-craft electromagnetic formation flight (EMFF) in low earth orbit was developed by using Lagrange theory. The linearized equations for motion and stability analysis of two-craft radial and along-track formation were investigated. To stabilize the formation structure, the feedback control law was studied. The motion programming of electromagnetic formation could be transformed into a standard optimal problem, which could be perfectly solved by the Gauss pseudospectral method. In order to deal with the reconfiguration maneuver of multi-craft electromagnetic formation, the sequence control strategy was proposed. Then the multi-craft problem was transformed into a multi-phase motion planning problem. Thus, the problem of multi-phase optimal control could be solved by Gauss pseudospectral optimization software (GPOPS). The simulation result shows that the methods of dynamics modeling and motion planning proposed in this paper are valid.

Key words: electromagnetic spacecraft; formation flight; reconfiguration maneuver; sequence control; motion programming

北京航空航天大学学报

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0319

基于分段非线性 Arrhenius 的贮存寿命评估方法

周洁，姚军*，宋燕

（北京航空航天大学 可靠性与系统工程学院，北京 100191）

摘 要：针对对数线性阿伦尼斯模型适用于温度在较小范围内变化的情况，给出了一种阿伦尼斯模型的修正方法.结合步退应力加速寿命试验失效率高、样本量少、试验设备简单的优点，给出了一种在工程实践中利用分段非线性阿伦尼斯模型对步退应力加速贮存试验进行评估的方法.首先，基于分段非线性阿伦尼斯模型建立步退应力下失效数据折算的数学模型.其次，建立了基于全样本的极大似然函数.最后利用遗传算法解非线性方程组，得到分段非线性阿伦尼斯模型的参数估计.以雷达高度表的步退应力加速贮存试验进行实例分析，将基于分段非线性阿伦尼斯模型、基于对数线性阿伦尼斯模型的加速贮存寿命评估方法得到的计算结果与产品外场得到的平均寿命、失效率以及可靠度函数进行对比分析，基于分段非线性阿伦尼斯模型的加速贮存寿命评估方法得到的结果误差在5%范围内.验证了论文中基于分段非线性阿伦尼斯模型的步退应力加速贮存试验评估方法的有效性以及准确性.

关 键 词：分段非线性阿伦尼斯模型；步退应力；加速贮存试验；寿命评估；极大似然函数

中图分类号：TB114.3

文献标识码：A **文章编号**：1001-5965(2015)04-0744-07

长期以来，高可靠、长寿命电子产品的量化评估一直是可靠性工程领域的一个技术难题.随着科学技术的快速发展，可靠性水平的不断提高，这一问题日益突出.对于高可靠性、长寿命的电子元器件来说，主要通过加速寿命试验进行寿命评估.而加速寿命试验的评估则需要借助于加速模型[1-2]，电子设备在贮存期间，会受到温度、湿度、振动等环境因素的影响，然而，在整个贮存期间主要的影响因素是温度.阿伦尼斯模型已被广泛应用于温度作为加速应力的加速寿命试验的评估中[3].Nelson 给出了指数分布下、威布尔分布下以及对数正态分布下阿伦尼斯模型用于描述产品的寿命与温度之间的关系[4-6]；文献[7]对某电子产品施加温度步进应力加速寿命试验，并运用阿伦尼斯模型、基本 GM(1,1)模型和残差修正 GM(1,1)模型对温度步进应力条件下的特征寿命进行点估计，证明阿伦尼斯模型并不完全适用于实际试验情况；文献[8]就阿伦尼斯方程和由其得出的阿伦尼斯加速因子之间存在的矛盾进行了分析研究，并提出了修正的阿伦尼斯加速因子；文献[9]针对产品在高温下的非工作失效率接近于其工作失效率，提出了阿伦尼斯修正模型，但这两种修正方法均没有应用到步进(退)加速试验的寿命评估中.

目前国内外相关研究中，探讨较多的是恒定应力加速试验评估方法，也有部分步进应力加速试验评估方法[10-12]，鲜有步退应力加速试验评估方法的相关论述[13].对于高可靠、长寿命电子产

收稿日期：2014-06-04；**录用日期**：2014-08-25；**网络出版时间**：2014-09-18 10:40
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0319.html
作者简介：周洁(1990—)，女，山东潍坊人，硕士生，zhouj90@126.com
* **通讯作者**：姚军(1971—)，男，陕西西安人，副教授，yao2jun1@sina.com，主要研究方向为可靠性与环境试验.

引用格式：*周洁，姚军，宋燕. 基于分段非线性 Arrhenius 的贮存寿命评估方法[J]. 北京航空航天大学学报，2015，41(4)：744-750.*
Zhou J，Yao J，Song Y. Storage life evaluation method based on segmented nonlinear Arrhenius model[J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：744-750 (in Chinese).

品,虽然步进应力加速试验理论方法较为成熟,统计精度较高,却存在所需样本量较大、试验时间较长及在做步进应力加速贮存试验过程中常会遇到较低应力水平下产品的损伤效果不明显的问题等不足,严重影响到试验效率.步退加速寿命试验与步进加速寿命试验相比,存在失效率高、样本量少、试验设备简单的优点[13-17],进而成为加速贮存试验在工程应用中的一个重要的研究方向.文献[10-11]给出了在指数分布场合下利用阿伦尼斯模型对步进应力加速寿命试验进行统计分析的方法;文献[12]给出了在指数分布场合下利用步进应力加速寿命试验的极大似然估计方法,并证明非线性方程组有唯一解;文献[17-18]从多个角度通过 Monte-Carlo 仿真模拟对比分析了步进应力和步退应力加速试验的试验效率,论证了步退应力试验的效率优势和步退应力试验方法的有效性.但这几种方法的缺点是假设产品在整个加速应力范围内产品的寿命与应力之间均服从对数线性阿伦尼斯模型,未对阿伦尼斯模型的适用情况进行分析做出修正.

本文针对产品在高温下的非工作失效率接近于其工作失效率[9],而对数线性阿伦尼斯模型并不适用于拟合温度较高时产品的寿命与温度应力之间的关系,给出了描述寿命与温度关系的分段非线性阿伦尼斯模型.与对数线性阿伦尼斯模型相比,无需产品退化速率为对数线性化的假设,算法相对简单,因此提高了该模型的工程适用性.通过某雷达高度表的加速贮存试验评估结果表明,与对数线性阿伦尼斯模型相比,采用本文修正后的分段非线性阿伦尼斯模型能够得到更加准确的贮存寿命评估结果.

1　分段非线性阿伦尼斯模型

1.1　模型假设

为了不失一般性,采用阿伦尼斯模型进行加速贮存试验寿命评估需要做如下假设[1]:

假设 1　在正常应力水平 S_0 和加速应力水平 $S_1,S_2,\cdots,S_k$ 下产品的寿命分布都服从指数分布,在应力 S_i 下产品的寿命分布函数为

$$F_i(t)=1-\mathrm{e}^{-t/\theta_i}\quad t\geqslant 0,i=0,1,\cdots,k \tag{1}$$

式中,θ_i 为 S_i 下产品的平均寿命,失效密度函数为

$$f_i(t)=\lambda_i\mathrm{e}^{-\lambda_i t}\quad t\geqslant 0,i=0,1,\cdots,k \tag{2}$$

假设 2　产品的平均寿命 θ_i 与所用的加速应力水平 S_i 之间服从阿伦尼斯模型的关系:

$$\ln\theta_i=a+b\xi(S_i)\quad i=0,1,\cdots,k \tag{3}$$

假设 3　产品的剩余寿命仅依赖于当时已累积失效部分和当时的应力水平,而与累积方式无关.

1.2　加速模型

阿伦尼斯模型[1]:

$$\theta=A\cdot\mathrm{e}^{\frac{E_a}{kT}} \tag{4}$$

式中,θ 为产品平均寿命;A 为常数;E_a 为失效机理的激活能;$k=8.617\times10^{-5}$ eV,表示 Boltzman 常数;T 为绝对温度,K.

由于在加速试验中,产品会经历高温下的加速试验,温度变化范围较大,在较低的温度范围内,产品的非工作失效率基本保持不变,而在高温范围内,产品的非工作失效率渐进接近产品的工作失效率[9],因此并不能判定阿伦尼斯模型在整个温度变化范围内都是对数线性关系(对 $1/T$).如果仍然在整个温度范围内用对数线性阿伦尼斯模型来处理问题,就会出现较大偏差,有必要引进一种精度更高的寿命评估方法.因此,在高温时,提出下面的对数非线性阿伦尼斯修正模型[9]:

$$\theta=A\cdot\mathrm{e}^{\frac{E_a}{kT}}\cdot\mathrm{e}^{-\left(\frac{T}{T_m}\right)^P} \tag{5}$$

式中,T_m 为产品失效机理发生改变的临界温度;P 为形状参数.对式(4)和式(5)两边取对数,得到分段非线性阿伦尼斯模型:

$$\theta_i=\begin{cases}\exp(a+b\varphi_i) & T_0\leqslant S_i\leqslant T_{med}\\ \exp(a+b\psi(\varphi_i)) & T_{med}<S_i\leqslant T_m\end{cases} \tag{6}$$

式中,$a=\ln A$ 和 $b=\dfrac{E_a}{k}$ 且 $b>0$ 为待估参数;$\varphi_i=1/S_i$ 为已知函数;T_0 为正常应力下的温度;$\psi(\varphi_i)=\varphi_i-\dfrac{1}{b}\cdot(T_m\cdot\varphi_i)^{-P}$;$T_{med}$ 为失效率发生较大变化的临界值.

根据指数分布下失效率与平均寿命的关系[1],得到:

$$\lambda_i=\begin{cases}\exp(-a-b\varphi_i) & T_0\leqslant S_i\leqslant T_{med}\\ \exp(-a-b\psi(\varphi_i)) & T_{med}<S_i\leqslant T_m\end{cases} \tag{7}$$

1.3　步退应力加速贮存试验

设 S_1 为试验初始加速应力水平,且 $S_1>S_2>\cdots>S_{k-1}>S_k$,$t_i(i=1,2,\cdots,k-1)$ 表示应力 S_i 到 S_{i+1} 的转换时刻,$t_0=0$ 为试验开始时刻,t_k 为试验结束时刻.记第 i 个等级下第 j 个失效样本的失效时间为 $t_{ij}(i=1,2,\cdots,k;j=1,2,\cdots,r_i)$,对于定数截尾试验,有 $t_{ij}=t_k$.试验总失效数为 $r=\sum_{i=1}^{k}r_i$.

以 $\Delta_i = t_i - t_{i-1}$ 表示在 S_i 下的试验时间.

2 贮存寿命评估模型

2.1 试验时间折算数学模型

步退应力试验的失效数据除最高应力水平以外都是多个应力水平共同作用的结果,即只有最高应力下的失效数据才是产品真正的寿命. 因此如何对试验时间进行折算从而分离出完整寿命数据是步退应力试验分析的关键问题[19].

如果用 $G(t \mid S_k)$ 表示 $S_1, S_2 \cdots, S_k$ 应力水平下截止到 t 时刻产品的试验时间折算到 S_k 下的等效时间,则:

1) 应力水平 S_1 作用下第 j 个故障产品的试验时间为 $G_{1j}(t_{1j} \mid S_1) = t_{1j}, 1 \leqslant j \leqslant r_1$;

2) 应力水平 $S_i, 2 \leqslant i \leqslant k$ 作用下第 j 个故障产品的试验时间[12]为

$$G_{ij}(t_{ij} \mid S_i) = \sum_{h=1}^{i-1} \Delta_h \cdot \tau_{hi} + t_{ij} - t_{i-1}$$
$$2 \leqslant i \leqslant k, 1 \leqslant j \leqslant r_i$$

式中, $\sum_{h=1}^{i-1} \Delta_h \cdot \tau_{hi}$ 为第 i 个应力水平下第 j 个故障产品在应力水平 S_i 下的补偿量. 进而得到第 i 个应力水平下第 j 个故障产品的试验时间的模型为

$$\begin{cases} G_{1j}(t_{1j} \mid S_1) = t_{1j} & 1 \leqslant j \leqslant r_1 \\ G_{2j}(t_{2j} \mid S_2) = \Delta_1 \cdot \tau_{12} + t_{2j} - t_1 & 1 \leqslant j \leqslant r_2 \\ \vdots \\ G_{kj}(t_{kj} \mid S_k) = \sum_{h=1}^{k-1} \Delta_h \cdot \tau_{hk} + t_{kj} - t_{k-1} & 1 \leqslant j \leqslant r_k \end{cases} \tag{8}$$

2.2 评估模型

2.2.1 产品的寿命分布

由假设 1 和 3 可知产品在步退应力加速寿命试验下的分布函数[1]为

$$F(G_k(t_k \mid S_k)) = 1 - \mathrm{e}^{-\lambda_k \cdot G_k(t_k \mid S_k)} =$$
$$1 - \exp\{-\mathrm{e}^{-a-b\psi_k} \cdot G_k(t_k \mid S_k)\} =$$
$$1 - \exp\left\{-\mathrm{e}^{-a-b\psi_k} \cdot \left(\sum_{h=1}^{k-1} \Delta_h \cdot \tau_{hk} + t_k - t_{k-1}\right)\right\} \tag{9}$$

第 i 个等级下第 j 个失效样本的失效概率密度函数[1]为

$$f(G_{ij}(t_{ij} \mid S_i)) = \lambda_i \cdot \mathrm{e}^{-\lambda_i G_{ij}(t_{ij} \mid S_i)} \tag{10}$$

2.2.2 极大似然估计模型

对于低温($T_0 < T \leqslant T_{\mathrm{med}}$)下产品平均寿命的极大似然估计的建模及求解可参考文献[12],在此只对高温($T_{\mathrm{med}} < T < T_{\mathrm{m}}$)下产品平均寿命的极大似然估计建模并进行求解.

对于 n 个产品进行步退应力加速贮存寿命试验,建立全样本的极大似然估计模型[12]如下:

$$L = \left[\prod_{i=1}^{k}\prod_{j=1}^{r_i} f(G_{ij}(t_{ij} \mid S_i))\right] \cdot [1 - F(G_k(t_k \mid S_k))]^{n-r} \tag{11}$$

结合前面的寿命分布函数和失效概率密度函数,得到模型的对数形式如下:

$$\ln L = \sum_{i=1}^{k}\sum_{j=1}^{r_i}\left\{(-a - b\psi_i) - \mathrm{e}^{-a}\left[(t_{ij} - t_{i-1})\mathrm{e}^{-b\psi_i} + \sum_{h=1}^{i-1}\Delta_h \mathrm{e}^{-b\psi_h}\right]\right\} -$$
$$\mathrm{e}^{-a}(n-r)\left[\Delta_k \mathrm{e}^{b\psi_k} + \sum_{i=1}^{k-1}\Delta_i \cdot \mathrm{e}^{-b\psi_i}\right] =$$
$$-ra - b\sum_{i=1}^{k} r_i\psi_i - \mathrm{e}^{-a}\sum_{i=1}^{k} u_i \mathrm{e}^{-b\psi_i} \tag{12}$$

式中

$$u_i = \sum_{j=1}^{r_i} t_{ij} - r_i \cdot t_{i-1} + \left(n - \sum_{h=1}^{i} r_h\right) \cdot (t_i - t_{i-1})$$
$$i = 1, 2, \cdots, k$$

利用该对数似然方程,分别对参数 a, b 和 P 求偏导数,建立非线性方程组如下:

$$\begin{cases} \dfrac{\partial \ln L}{\partial a} = -r + \mathrm{e}^{-a}\sum_{i=1}^{k} u_i \mathrm{e}^{-b\psi_i} = 0 \\ \dfrac{\partial \ln L}{\partial b} = -\sum_{i=1}^{k} r_i\varphi_i + \mathrm{e}^{-a}\sum_{i=1}^{k} u_i\varphi_i \mathrm{e}^{-b\psi_i} = 0 \\ \dfrac{\partial \ln L}{\partial P} = -\sum_{i=1}^{k} r_i (T_{\mathrm{m}} \cdot \varphi_i)^{-P} \ln(T_{\mathrm{m}} \cdot \varphi_i)^{-1} - \\ \qquad \mathrm{e}^{-a}\sum_{i=1}^{k} u_i\left(-\dfrac{1}{T_{\mathrm{m}} \cdot \varphi_i}\right)\mathrm{e}^{-b\psi_i} = 0 \end{cases} \tag{13}$$

利用遗传算法进行特征提取,图 1 为基于遗传算法的实现框架[20].

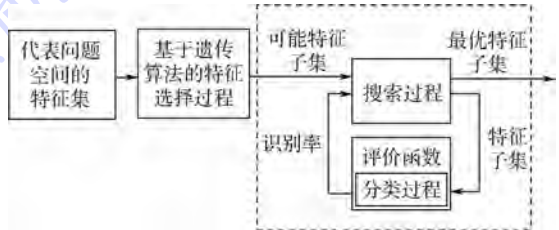


图 1 基于遗传算法的实现框架

Fig. 1 Implementation framework based on genetic algorithm

基于遗传算法的求解过程[20]如下.

1) 确定决策变量和约束条件:

决策变量为 a, b, P,约束条件为 $a < 0, b > 0, P > 0$.

2) 确定目标函数:

$$\min f(a,b,P)=\left(\frac{\partial \ln L}{\partial a}\right)^2+\left(\frac{\partial \ln L}{\partial b}\right)^2+\left(\frac{\partial \ln L}{\partial P}\right)^2$$
$$a<0,b>0,P>0 \tag{14}$$

3) 求解目标函数,获得最优解.

2.2.3　正常应力水平下平均寿命 θ_0 的极大似然估计

利用遗传算法对式(13)求解得到待估计参数的估计值为$(\hat{a},\hat{b},\hat{P})$,结合式(6)进而得到正常应力水平下参数 θ_0 的极大似然估计(MLE,Maximum Likelihood Estimation)为

$$\hat{\theta}_0=\mathrm{e}^{\hat{a}+\hat{b}/S_0} \tag{15}$$

3　应用算例

3.1　雷达高度表步退应力加速贮存试验

某雷达高度表属于电子类产品,由主机与收、发天馈装置组成,主要元器件类别包含:电感器、电容器(钽电解、磁介质)、金属膜电阻器、连接器(SMA、低频航空插头)、微波单片、集成电路、导线、半柔性电缆等. 其寿命服从指数分布,从外场的试验数据分析已知其平均故障间隔时间(MTBF)为 604 800 h. 根据试验应力的提高应以不改变产品在正常工作水平下的失效机理为准则,通过预试验已知雷达高度表在 165℃ 的试验温度下均为电源模块失效,失效原因为高温使电源模块内印制板垂直向上方向上的聚氨酯橡胶发生较大形变,致使其灌封保护的个别焊点开裂,造成电源模块失效,确定 $T_m=438$ K. 且通过预试验得到雷达高度表在 $T>353$ K 时失效率较高,故确定 $T_{med}=353$ K. 由于温度是影响电子产品在贮存期性能的主要应力,选择试验加速应力为温度应力,加速模型选择阿伦尼斯模型.

现欲考察雷达高度表在正常贮存温度($S_0=298$ K)条件下的寿命,由于雷达高度表是典型的高可靠长寿命产品,在常应力下其失效非常缓慢,用定数截尾(截尾数为 1)步退应力加速寿命试验方法对其贮存寿命进行评估.

相应试验参数为:试验样本 $n=6$,步退应力等级 $k=3$,加速应力为 $S_1=388$ K,$S_2=378$ K,$S_3=368$ K,每个应力等级下的失效数均为 1,即 $r_i=1,i=1,2,3$. 各应力等级下故障产品的故障时间(即各应力等级转换时刻)分别为 $t_1=2\,160$ h,$t_2=5\,420$ h,$t_3=11\,820$ h,则各应力水平下的试验时间分别是 $\Delta_1=t_1=2\,160$ h,$\Delta_2=t_2-t_1=3\,260$ h,$\Delta_3=t_3-t_2=6\,400$ h.

3.2　贮存寿命评估

3.2.1　基于分段非线性阿伦尼斯模型的贮存寿命评估结果

结合式(13)和试验数据,得到加速应力 $S_i(i=1,2,3)$ 下试样的故障前时间折算到加速应力水平 S_i 下的试验时间为 $t_i^*=(t_i-t_{i-1})+\sum_{h=1}^{i-1}\Delta_h \mathrm{e}^{b(\psi(\varphi_i)-\psi(\varphi_h))}$,利用遗传算法求解,得到加速模型中的参数估计值为 $\hat{a}=-5.08$,$\hat{b}=5\,488.83$,$\hat{P}=3$,将估计结果$(\hat{a},\hat{b},\hat{P})$代入式(6)进而得到加速方程为

$$\theta_i=\begin{cases}\exp(-5.08+5\,488.83\varphi_i) & T_0\leqslant S_i\leqslant T_{med}\\ \exp(-5.08+5\,488.83\psi(\varphi_i)) & T_{med}<S_i\leqslant T_m\end{cases} \tag{16}$$

取正常贮存温度 $S_0=298$ K,得到正常贮存条件下产品的平均寿命 $\theta_1=620\,880$ h,失效率 $\lambda_1=1/\theta_1=1.61\times10^{-6}$.

3.2.2　基于对数线性阿伦尼斯模型的贮存寿命评估结果

根据文献[12],加速应力 $S_i(i=1,2,3)$ 下试样的故障前时间折算到加速应力水平 S_i 下的试验时间为 $t_i^*=(t_i-t_{i-1})+\sum_{h=1}^{i-1}\Delta_h \mathrm{e}^{b(\varphi_i-\varphi_h)}$,得到基于对数线性阿伦尼斯模型的步退应力加速贮存试验中加速模型的参数估计值 $\hat{a}=-2.28$,$\hat{b}=4\,552.78$.

将估计结果$(\hat{a},\hat{b})$代入式(5)得到产品在自然贮存条件下的平均寿命的方程:$\theta_2=\exp(-2.28+4\,552.78\varphi_0)$,进而得到产品在自然贮存条件下的平均寿命 $\theta_2=441\,435$ h,失效率为 $\lambda_2=1/\theta_2=2.27\times10^{-6}$.

3.3　结果分析

1) 将基于分段非线性阿伦尼斯模型的加速贮存寿命评估方法得到的计算结果、基于对数线性阿伦尼斯模型加速贮存寿命评估方法得到的计算结果与产品的实际平均寿命和失效率进行对比分析,其误差结果如下式所示:

$$S(\theta_1)=\frac{620\,880-604\,800}{604\,800}=0.026\,6 \tag{17}$$

$$S(\theta_2)=\frac{441\,435-604\,800}{604\,800}=-0.270\,1 \tag{18}$$

$$S(\lambda_1)=\frac{1.61\times10^{-6}-1.65\times10^{-6}}{1.65\times10^{-6}}=-0.024\,2 \tag{19}$$

北航学报 赠阅

$$S(\lambda_2)=\frac{2.27\times10^{-6}-1.65\times10^{-6}}{1.65\times10^{-6}}=0.3758 \tag{20}$$

从式(17)~式(20)可以看出,基于对数线性阿伦尼斯模型评估得到的平均寿命 θ_2 相对于实际平均寿命的误差为 27.01%,失效率的误差为 37.58%,评估结果偏差较大. 而基于修正的分段非线性阿伦尼斯模型得到的平均寿命 θ_1 相对于实际平均寿命的误差仅为 2.66%,失效率的误差仅为 2.42%,相对误差均控制在 5% 之内,其评估结果精确度远远高于基于对数线性阿伦尼斯模型评估得到的结果,评估结果精确度较高.

2) 结合式(6),得到基于分段非线性阿伦尼斯模型得到的失效率和基于对数线性阿伦尼斯模型得到的失效率随不同温度应力的变化趋势如图 2所示.

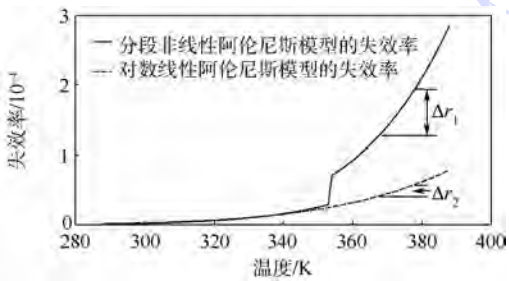


图 2 失效率随温度变化的趋势图

Fig. 2 Trend of failure rate changing with temperature

从图 2 可以看出,在低温阶段,基于对数线性阿伦尼斯模型的加速贮存寿命评估方法得到的失效率 r_2 和基于分段非线性阿伦尼斯模型的加速贮存寿命评估方法得到的失效率 r_1 随温度变化均较平缓,而在高温时,如从 368~378 K,失效率 r_1 的变化率为 $\frac{\Delta r_1}{378-368}=5.4179\times10^{-6}$,失效率 r_2 的变化率为 $\frac{\Delta r_2}{378-368}=1.6039\times10^{-6}$,即失效率 r_1 的变化率远大于失效率 r_2 的变化率,基于分段非线性阿伦尼斯模型的加速贮存寿命评估方法更加符合电子产品在较低的温度范围时非工作失效率较小,而在高温范围时,其非工作失效率较大并渐进接近产品的工作失效率的现象.

3) 将基于分段非线性阿伦尼斯模型的加速贮存寿命评估方法得到的可靠度 R_1、基于对数线性阿伦尼斯模型的加速贮存寿命评估方法得到的可靠度 R_2 与产品外场得到的可靠度 R_0 进行对比分析,可得其可靠度函数的对比如图 3 所示.

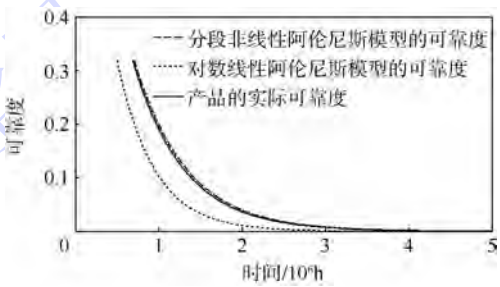


图 3 不同方法得到的可靠度随时间变化趋势的对比

Fig. 3 Comparison of reliability trends over time got by different ways

从图 3 中可以看出,基于分段非线性阿伦尼斯模型得到的可靠度函数 R_1 与实际可靠度函数 R_0 几乎重合,吻合度相对较高,进一步验证了本文中基于分段非线性阿伦尼斯模型的加速贮存寿命评估方法的有效性以及准确性.

3.4 Monte-Carlo 仿真

为进一步说明方法的合理性,利用 Monte Carlo 仿真试验数据并分析结果. 根据贮存失效分布函数,产品的寿命服从指数分布,$F(t)=1-e^{-t/\theta}$,从变换抽样方法可得 $X=1-e^{-t/\theta}$. 由于 X 与 $1-X$ 均为区间[0,1]上服从均匀分布的随机数,可得 $T=-\theta\ln X$ 即为服从参数 θ 的指数分布的随机变量. 然后利用 Monte Carlo 仿真,得到仿真试验数据及相应的评估结果如表 1 所示.

表 1 Monte Carlo 仿真及评估结果

Table 1 Monte Carlo simulation and evaluation results

序号	应力	仿真失效时间/h	θ/h	基于分段非线性		基于对数线性	
				θ_1/h	偏差/%	θ_2/h	偏差/%
1	$T_m=433$ K, $T_{med}=408$ K $S_1=428$ K, $S_2=418$ K, $S_3=408$ K	288 624 912	120 960	125 860	4.05	106 265	−12.15
2	$T_m=418$ K, $T_{med}=363$ K $S_1=388$ K, $S_2=378$ K, $S_3=368$ K	1 243 2 587 3 763	345 600	362 681	4.94	286 969	−16.96
3	$T_m=431$ K, $T_{med}=388$ K $S_1=418$ K, $S_2=408$ K, $S_3=398$ K	2 184 3 192 4 320	2 420 000	2 562 538	5.89	1 950 800	−19.39

从表 1 可以看出,基于分段非线性阿伦尼斯模型的步退应力加速贮存试验评估方法得到的评估结果与仿真的偏差百分比控制在 5% 左右,其精度明显高于基于对数线性阿伦尼斯模型的步退应力加速贮存试验评估方法所得到的结果.

4　结　论

1) 本文针对对数线性阿伦尼斯模型不能准确地拟合寿命与应力之间的关系,结合步退应力加速寿命试验的统计分析问题进行讨论,给出了一种基于分段非线性阿伦尼斯模型的步退应力加速贮存试验评估方法.

2) 通过对雷达高度表的步退应力加速贮存试验评估得到的平均寿命、失效率以及可靠度函数的对比分析,可以看出本文所研究的基于分段非线性阿伦尼斯模型的加速贮存寿命评估方法的结果误差在 5% 范围内,具有较好的适用性和较高的精确度.本文的基于分段非线性阿伦尼斯模型的加速贮存寿命评估方法能够有效地解决步退应力加速贮存寿命试验的统计分析问题,为武器装备定寿与延寿工程奠定了基础.

参考文献 (References)

[1] 茆诗松,王玲玲. 加速寿命试验[M]. 北京:科学出版社,1997:16-22.
Mao S S, Wang L L. Accelerated life testing[M]. Beijing: Science Press, 1997:16-22(in Chinese).

[2] 张春华,陈循,温熙森. 加速寿命试验技术综述[J]. 兵工学报,2004,25(4):485-490.
Zhang C H, Chen X, Wen X S. A review of accelerated life testing technology[J]. Acta Armamentarii, 2004, 25(4):485-490(in Chinese).

[3] Hartler G. Parameter estimation for the arrhenius model[J]. IEEE Transactions on Reliability, 1986, 35(4):414-418.

[4] Nelson W. Accelerated testing-statistical model, test plans and data analysis[M]. New York: John Wiley & Sons, 1990:496-502.

[5] Nelson W. Analysis of accelerated life test data-part I: the arrhenius model and graphical methods[J]. IEEE Transactions on Electrical Insulation, 1971, 6(4):165-181.

[6] Nelson W. Accelerated life testing step-stress models and data analysis[J]. IEEE Transactions on Reliability, 1980, 29(2):103-108.

[7] 李健,汪金华,陆陪永. 温度步进应力加速寿命试验研究[J]. 电子产品可靠性与环境试验,2007,25(1):1-4.
Li J, Wang J H, Lu P Y. Research of temperature stepped stress accelerated life testing[J]. Electronic Product Reliability and Environmental Testing, 2007, 25(1):1-4(in Chinese).

[8] 李进,李传日. 加速寿命试验中修正阿伦尼斯加速因子的研究[J]. 电子产品可靠性与环境试验,2009(S1):38-42.
Li J, Li C R. A study of the modified Arrhenius acceleration factor in acceleration life test[J]. Electronic Product Reliability and Environmental Testing, 2009(S1):38-42(in Chinese).

[9] Coit D W, Priore M G. Impact of nonoperating periods on equipment reliability, RADC-1984-22232[R]. Rome: IIT Research Institute, 1984.

[10] 茆诗松. 指数分布场合下步进应力加速寿命试验的统计分析[J]. 应用数学学报,1985,8(3):311-316.
Mao S S. Statistical analysis of accelerated life testing-step-stress models under the exponential distribution case[J]. Acta Mathematicae Applicatae Sinica, 1985, 8(3):311-316(in Chinese).

[11] 仲崇新,张志华. 指数分布场合定时和定数截尾步进应力加速寿命试验和统计分析[J]. 应用概率统计,1991,7(1):52-59.
Zhong C X, Zhang Z H. Type-I censored and type-II censored step stress accelerated life test and statistical analysis under exponential distribution[J]. Chinese Journal of Applied Probability and Statistics, 1991, 7(1):52-59(in Chinese).

[12] 费鹤良,张学新. 指数分布场合下步进应力加速寿命试验的极大似然估计[J]. 应用数学学报,2004,17(3):398-404.
Fei H L, Zhang X X. The maximum likelihood estimation of step stress accelerated life testing under exponential distribution situation[J]. Acta Mathematicae Applicatae Sinica, 2004, 17(3):398-404(in Chinese).

[13] 李宪珊. 步退应力加速试验方法及其应用[J]. 强度与环境,2012,39(6):59-62.
Li X S. Step down stress accelerated testing method and application[J]. Structure & Environment Engineering, 2012, 39(6):59-62(in Chinese).

[14] 张春华. 步降应力加速寿命试验的理论和方法[D]. 长沙:国防科技大学,2002.
Zhang C H. Theory and method of step down stress accelerated life testing[D]. Changsha: National University of Defense Technology, 2002(in Chinese).

[15] 张春华,陈循,温熙森. 步降应力加速寿命试验(上篇)——方法篇[J]. 兵工学报,2005,26(5):661-665.
Zhang C H, Chen X, Wen X S. Step-down-stress accelerated life testing-methodology[J]. Acta Armamentarii, 2005, 26(5):661-665(in Chinese).

[16] 张春华,陈循,温熙森. 步降应力加速寿命试验(下篇)——统计分析篇[J]. 兵工学报,2005,26(5):666-669.
Zhang C H, Chen X, Wen X S. Step-down-stress accelerated life testing-statistical analysis[J]. Acta Armamentarii, 2005, 26(5):666-669(in Chinese).

[17] 徐广,王蓉华. 步降应力加速寿命试验的效率分析[J]. 上海师范大学学报:自然科学版,2008,37(5):468-475.
Xu G, Wang R H. Efficiency analysis of step-down-stress accelerated life testing[J]. Journal of Shanghai Normal University: Natural Sciences, 2008, 37(5):468-475(in Chinese).

[18] Wang R H, Sha N J, Gu B Q, et al. Comparison analysis of efficiency for step-down and step-pp stress accelerated life testing[J]. IEEE Transactions on Reliability, 2012, 61(2):590-603.

[19] 赵宇,杨军,马小兵. 可靠性数据分析[M]. 北京:国防工业出版社,2011:25-26.

北航学报 赠阅

Zhao Y, Yang J, Ma X B. Data analysis of reliability[M]. Beijing: National Defence Industry Press, 2011: 25-26 (in Chinese).

[20] 雷英杰,张善文,李续武,等. 遗传算法工具箱及应用[M]. 西安:西安电子科技大学出版社,2014:18-19.
Lei Y J, Zhang S W, Li X W, et al. Genetic algorithm toolbox and its application[M]. Xi'an: Xi'an University of Electronic Science and Technology Press, 2014: 18-19 (in Chinese).

Storage life evaluation method based on segmented nonlinear Arrhenius model

ZHOU Jie, YAO Jun*, SONG Yan

(School of Reliability and Systems Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: Considering that logarithm linear Arrhenius model is only suitable for a small temperature range, a correction method for the Arrhenius model was given. As the step-down-stress accelerated life testing has the advantages of higher failure rate, less sample size, and simpler testing equipment, an evaluation method used in engineering practice for step-down-stress accelerated storage life testing was provided based on the segmented nonlinear Arrhenius model. A mathematical model for failure data conversion under step down stress based on the segmented nonlinear Arrhenius model was first established. The maximum likelihood function of the whole samples was given. The segmented nonlinear Arrhenius model parameter estimation could then be obtained by employing the genetic algorithm to solve nonlinear equations. Taking the step-down-stress accelerated storage life testing of radar altimeter as an example for analysis, the results of average life expectancy, the failure rate and reliability function based on logarithm linear Arrhenius model accelerated storage life evaluation method, the results based on segmented nonlinear Arrhenius model accelerated storage life evaluation method, and the results from field data were compared. The relative error of the results based on the segmented nonlinear Arrhenius model accelerated storage life evaluation method is within 5%. The efficiency and accuracy of the step-down-stress accelerated storage life evaluation method based on the segmented nonlinear Arrhenius model is verified.

Key words: segmented nonlinear Arrhenius model; step down stress; accelerated storage testing; life evaluation; maximum likelihood function

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2014.0499

基于多尺度梯度及深度神经网络的汉字识别

潘炜深，金连文*，冯子勇

（华南理工大学 电子与信息学院，广州 510641）

摘 要：介绍了一种基于多尺度滑动窗的方法提取文字的梯度直方图特征，并结合深度神经网络对印刷体汉字进行识别. 针对梯度直方图的空间关系，使用可伸缩的滑动窗对图像进行分割，在不同尺度上获取文字的特征信息，有效融合汉字的全局特征和局部分块特征. 实验采用5层的深度神经网络模型对国标一级3 755个印刷体汉字进行分类，并应用Dropout技术防止训练过拟合，提高神经网络的泛化能力. 实验准确率达到98.292%，有较好的识别性能，验证了本文多尺度梯度特征及深度神经网络模型在文字识别上的有效性.

关 键 词：多尺度滑动窗；梯度直方图；深度神经网络；泛化能力；汉字识别
中图分类号：TP391.4
文献标识码：A **文章编号**：1001-5965(2015)04-0751-06

汉字识别技术的研究是社会科技发展的关键因素. 在高度信息化的时代，如何让计算机高效地识别出如此之多的汉字，特别是印刷体汉字信息，是汉字识别领域的重要问题. 印刷体汉字识别有着很高的实用价值，在中文信息处理、文本行识别和自动化办公等领域都有重大的理论意义和应用价值. 印刷体汉字识别最早可追溯到20世纪60年代，IBM公司的Casey和Nagy利用简单模板匹配法识别出1 000个印刷体汉字[1]. 20世纪70年代初，日本研究员也开始研究汉字识别技术并成功研制出能识别2 000～4 000个印刷体汉字和特定手写体汉字的设备. 中国的光学字符识别(OCR，Optical Character Recognition)研究起源于20世纪70年代末，主要针对数字、字母和符号的识别. 20世纪80年代末，中国已在文字识别领域上取得了丰硕的成果，在汉字建模和识别方法上都有所创新，OCR产品开始走向市场[2]. 如今人工智能技术飞速发展，机器学习与模式识别等领域出现了很多新的成就和突破，如神经网络、自然语言处理等. 如何吸取这些先进的技术，把他们更好地运用到文字识别中，是一个需要长期研究和努力的方向[3].

特征提取和选择是汉字识别的关键. 随着汉字分类技术研究的深入，文字的特征表达和描述已经取得了长足的发展. 总体上汉字特征分为统计特征和结构特征. 统计特征建立在二值或灰度点阵图像之上，它将点阵图像从整体上进行处理，在较高的层次上进行统计，如全局变换特征和笔画穿透数特征. 这类特征易于提取且往往具有良好的鲁棒性和较好的抗噪声能力，但对局部信息的敏感度不够，区分相似字的能力较差. 结构特征则是字符的本质特征. 它描述了字符的拓扑结构，

收稿日期：2014-04-28；**录用日期**：2014-11-27；**网络出版时间**：2014-11-27 17:41
网络出版地址：www.cnki.net/kcms/doi/10.13700/j.bh.1001-5965.2014.0499.html
基金项目：国家自然科学基金资助项目(61075021)；国家科技支撑计划资助项目(2013BAH65F01，2013BAH65F04)；广东省科技计划资助项目(2012A010701001)
作者简介：潘炜深(1990—)，男，广东惠州人，硕士生，panweishen2009@gmail.com
*通讯作者：金连文(1968—)，男，贵州都匀人，教授，lianwen.jin@gmail.com，主要研究方向为手写汉字识别、机器学习、模式识别、云计算等.

引用格式：*潘炜深，金连文，冯子勇. 基于多尺度梯度及深度神经网络的汉字识别[J]. 北京航空航天大学学报，2015，41(4)：751-756. Pan W S，Jin L W，Feng Z Y. Recognition of Chinese characters based on multi-scale gradient and deep neural network [J]. Journal of Beijing University of Aeronautics and Astronautics，2015，41(4)：751-756 (in Chinese).*

对字体变形影响较小,对相似字的区分效果好,如汉字特征点特征,但这类特征易受噪声等干扰. 为更好地结合各自的优点,本文采用一种适用于汉字分类的特征描述子,通过多尺度滑动窗的方式提取文字的全局特征和局部特征. 滑动窗的思想由法国研究员 Dalal 和 Triggs 在静态图像和视频行人检测问题上提出,利用滑动窗口技术计算图像局部区域的梯度方向直方图来构成特征,对图像几何和光学的变形都能保持很好的不变性[4].

神经网络(ANN)是人类在对其大脑神经网络认识理解的基础上构造的能够实现自动信息处理的机器系统[5]. 由于其具有有效的学习、模仿人脑智能等方面的能力而广泛地被应用于模式识别、图像处理等领域. 文字识别是一个模拟人类认知过程的问题,因此,将神经网络应用于文字识别一直是体现神经网络性能的一个重要应用领域[6-8]. 本文使用一个 5 层的深度神经网络模型对提取好的特征进行分类识别,同时使用 Dropout 技术防止训练过拟合,提高神经网络的泛化能力.

1 文字特征提取

特征提取是印刷体汉字识别系统最为关键的环节之一. 为了在得到文字的拓扑结构信息的同时又具有较好的抗噪性,本节介绍一种基于多尺度滑动窗的方式提取文字的梯度方向信息,使用尺度可伸缩的滑动窗对图像进行分割,在每个滑动窗内采用梯度直方图的方式进行表达,最后把所有滑动窗内的特征描述子按一定的权重大小联合起来形成最后的特征矢量的方法. 这种特征提取方法不仅计算代价低,而且引入了空间位置信息,可以有效获取汉字的全局特征和局部分块特征.

1.1 梯度信息

为了获取文字的梯度方向直方图,首先需要计算图像每个像素的梯度大小和方向,捕捉文字的大致轮廓. 本文利用 5×5 的高斯梯度算子计算原图像梯度,并统计 12 个不同方向的梯度大小. 图 1 表示“阿”字的梯度幅值信息,通过梯度的幅值可以捕捉文字的轮廓和纹理. 图 2 显示了“阿”字的梯度方向分布信息,不同的颜色代表不同的梯度方向. 因此,文字的梯度为汉字分类提供了一个非常有效的鉴别信息.

1.2 多尺度梯度直方图

为了尽可能多地获取文字整体的轮廓纹理信息,同时提取局部的分块特征,提高对相似字的区分效果,采用多尺度滑动窗的方式提取文字梯度信息. 多尺度滑动窗示意图如图 3 所示,从中可见,有 3 种尺度大小分别为 16×16,32×32,48×48 的红色矩形窗,利用滑动窗提取图像不同区域的梯度信息,然后统计不同区域的梯度方向,形成梯度方向直方图,最后把不同尺度的直方图按照一定的权重比例级联起来,形成特征矢量.

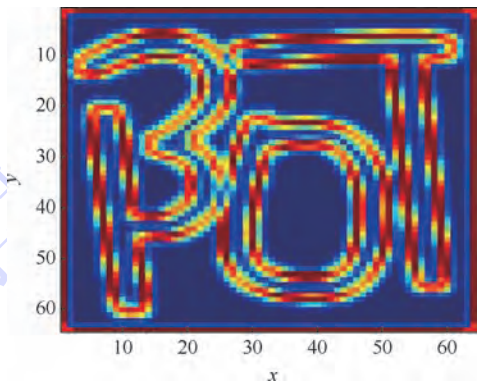


图 1 梯度幅值可视化

Fig. 1 Visualization of gradient magnitudes

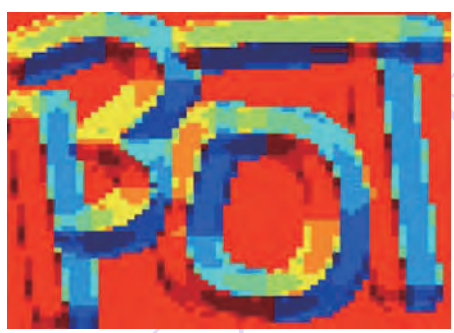


图 2 梯度方向可视化

Fig. 2 Visualization of gradient directions

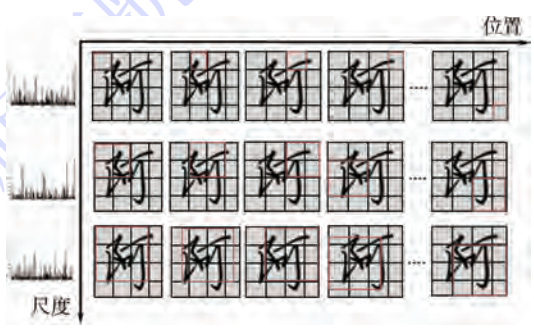


图 3 多尺度滑动窗示意图

Fig. 3 Multi-scale sliding windows

由于滑动窗在滑动的过程中以一定的步长进行移动,因此,窗与窗之间存在交叠(overlap). 滑动窗之间的交叠示意图如图 4 所示,从中可见,红色虚线区域部分就是由于选取的步长小于滑动窗的宽度而产生的. 交叠区域的特征将会以不同结果多次出现在最后的特征向量中. 这种重叠机制

往往可以改善特征的性能，但同时也会增加特征的维度.

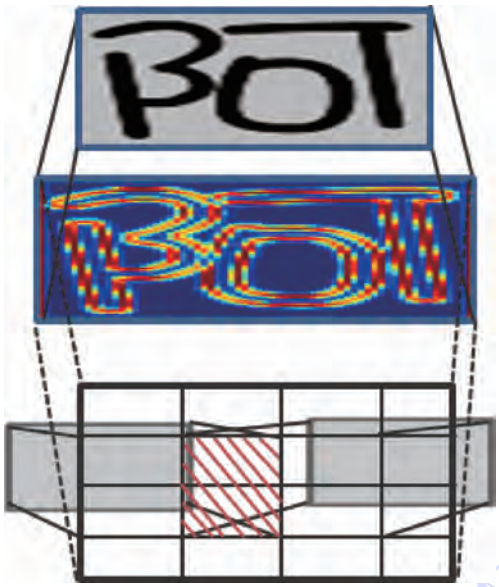


图 4　滑动窗之间的交叠示意图

Fig. 4　Overlap of the sliding windows

2　深度神经网络模型

神经网络是一种模拟人脑神经元细胞的网络结构，它是由大量简单的神经元相互连接成的自适应非线性动态系统. 相比于传统的多层感知器，本文采用整流线性单元(Rectified Linear Units)作为网络的激活函数，用 Soft-max 函数作为网络的输出，并以交叉熵(Cross-entropy)衡量训练的损失. 使用经典的随机梯度下降(SGD)法求解目标函数的最优解，用反向传播 BP 算法不断调整网络的权值. 为了防止训练过拟合，使用 Dropout 技术，提高网络的泛化能力.

2.1　传递函数

传统的人工神经网络使用 Sigmoid 和 tanh 等非线性函数作为激活函数，把神经元的输出映射在[0,1]或[-1,1]之间，但这些激活函数在训练时比较慢，在传递的过程中信息丢失较多，识别效果往往不太理想. 本文使用一种在受限玻尔兹曼机(RBMs)模型里广泛使用的整流线性单元函数[9]：$f(z) = \max(z,0)$ 作为网络的传递函数. 这种激活函数能使相关的特征信息在多层网络传递过程中得到更大的保留，与传统的 Sigmoid 和 tanh 传递函数相比，计算更简单，训练更快. Sigmoid 和 ReLU 函数如图 5 所示.

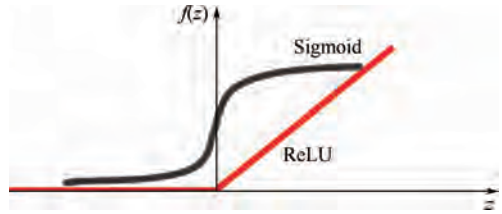


图 5　Sigmoid 和 ReLU 函数

Fig. 5　Sigmoid and ReLU functions

2.2　Soft-max 回归

相对于 Logistic[10] 回归，Soft-max[11] 回归是解决多类分类问题的一个有效方法. 对于有 k 类，每类有 m 个样本的数据集合：

$$\left\{(x^{(1)},y^{(1)}),(x^{(2)},y^{(2)}),\cdots,(x^{(m)},y^{(m)})\right\},$$
$$y^{(m)} \in \left\{1,2,\cdots,k\right\} \tag{1}$$

给定一个输入 x，对于每一个类别 j 估算出概率值 $p(y=j|x)$，假如用一个 k 维的输出向量来表示 k 个概率值，则函数应为：

$$\boldsymbol{h}_{\boldsymbol{\theta}}(x)=\begin{bmatrix} p(y^{(i)}=1 \mid x^{(i)};\boldsymbol{\theta}) \\ p(y^{(i)}=2 \mid x^{(i)};\boldsymbol{\theta}) \\ \vdots \\ p(y^{(i)}=k \mid x^{(i)};\boldsymbol{\theta}) \end{bmatrix}= \frac{1}{\sum_{j=1}^{k} \mathrm{e}^{\boldsymbol{\theta}_j^{\mathrm{T}}x^{(i)}}} \begin{bmatrix} \mathrm{e}^{\boldsymbol{\theta}_1^{\mathrm{T}}x^{(i)}} \\ \mathrm{e}^{\boldsymbol{\theta}_2^{\mathrm{T}}x^{(i)}} \\ \vdots \\ \mathrm{e}^{\boldsymbol{\theta}_k^{\mathrm{T}}x^{(i)}} \end{bmatrix} \tag{2}$$

式中，$\boldsymbol{h}_{\boldsymbol{\theta}}(x)$ 为模型的输出；i 为样本序号；$\boldsymbol{\theta}$ 为模型参数. $\sum_{j=1}^{k} \mathrm{e}^{\boldsymbol{\theta}_j^{\mathrm{T}}x^{(i)}}$ 是归一化因子，使得所有概率为 1. 神经网络模型采用交叉熵作为目标代价函数，其形式如下：

$$\ell = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)}=j\} \ln \frac{\mathrm{e}^{\boldsymbol{\theta}_j^{\mathrm{T}}x^{(i)}}}{\sum_{j=1}^{k} \mathrm{e}^{\boldsymbol{\theta}_j^{\mathrm{T}}x^{(i)}}}\right] \tag{3}$$

根据反向传播 BP 算法，将训练样本的误差从输出层反向传播到输入层，用随机梯度下降法求解目标函数的最优解，并在传播的过程中不断调整网络的权值，实现训练.

2.3　Dropout 技术

Dropout 是 Hinton 教授在深度学习领域里提出来的一个重要技巧，其思想主要是在模型训练时随机让网络某些隐含层节点输出值为零，这种如同在图像中加入噪声的方式能防止模型在训练过程中出现过拟合，提高神经网络的泛化能力[12-14]. 对于每次输进来的样本，由于其 Dropout 的随机性，每个样本对应的网络结构都不相同，这些不同的网络结构同时又共享隐含节点的权值，使得不同的样本对应不同的模型. 本文对输入层和所有的隐藏层都采用了 Dropout. Dropout 如图 6 所示，从中可见，深色的神经元表示被随机选为

Dropout 的节点单位.

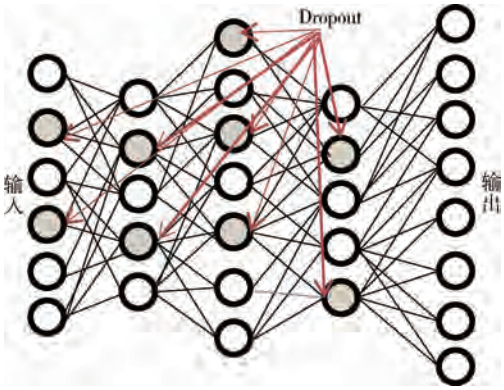


图 6 Dropout 示意图

Fig. 6 Illustration of Dropout

3 实 验

实验使用了一个 5 层的深度神经网络作为模型分类器. 由于数据规模大,网络参数众多,训练过程比较缓慢,对此借用 CUDAMat[15] 矩阵库,使用 GPU 进行加速,对国标一级 3 755 个印刷体共 375 500 个汉字进行测试.

3.1 数 据 库

实验数据共有 100 套字,包含宋体、楷体、黑体、钢笔体和圆体等各种字体. 每套字有 3 755 个样本. 随机选 70 套作为训练,10 套做校正,20 套做测试. "阿"字的 100 套不同风格的字体如图 7 所示.

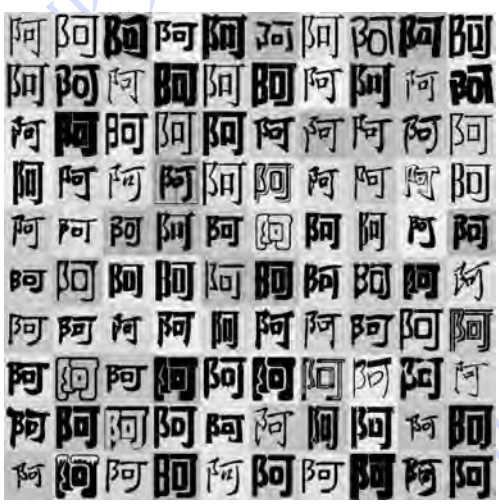


图 7 "阿"字的 100 套字

Fig. 7 One hundred fonts of "阿"

3.2 特征性能对比

为比较本文多尺度滑动窗梯度特征跟其他特征之间的性能差异,选用另外 6 种常用的特征如局部二值模式(LBP, Local Binary Pattern)[16-17]、GIST[18]、梯度方向直方图(HOG, Histogram of Oriented Gradient)、灰度特征、梯度特征和卷积神经网络(CNN, Convolutional Neural Netuorks)自学习出来的特征进行对比评估. LBP 特征不具有旋转和尺度不变性,但对于局部亮度的变化比较敏感,是一种直方图类型的纹理特征. GIST 是利用不同方向和空间分辨率的Gabor滤波器对图像采样滤波的特征. HOG 是一种全局特征,通过统计像素梯度方向直方图获得矢量描述. 灰度特征则直接把图像的像素灰度值作为特征矢量进行输入. 八方向梯度特征更多应用在文字上,把汉字点阵图像按 8 个方向进行分解. 卷积神经网络是一种常用的特征学习方法,能从大量的底层数据中获得有效的特征表示. 实验采用包含 2 层卷积层,2 层聚合层和 1 层全连接层的网络结构进行特征自学习. 特征对比实验数据如表 1 所示.

表 1 特征对比实验

Table 1 Experiments of different features

特征	识别率/%	时间/min
灰度	86.266	47.30
LBP	87.513	46.64
GIST	92.754	57.67
CNN 自学习特征	93.368	462.857
HOG	94.501	46.19
梯度	95.157	45.35
本文梯度特征	98.361	47.04

由表 1 看出,本文的多尺度滑动窗梯度特征效果更好. 相比于一般的梯度特征,虽然也有较高的识别率 95.157%,但基于多尺度的特征融合了文字更多的梯度信息,因此,其识别率更高,更有利于文字的识别分类. 由于手动提取的特征其参数已经预先固定,训练过程只需微调连接层的权重,故其运行时间较快,复杂度较低. 而卷积神经网络由于卷积层和聚合层的存在,在训练过程中需要不断对卷积和聚合模板的参数和权重进行调节,因此,其复杂度更高,运行时间更长.

3.3 Dropout 实验

下面采用本文提出的多尺度梯度特征在 5 层深度神经网络结构下对不同的 Dropout 率进行测试. 为公平对比,统一对所有的输入层和隐藏层进行 Dropout. 实验结果如表 2 所示.

表 2 不同 Dropout 率效果

Table 2 Results of different Dropout rates

Dropout 率	训练集识别率/%	测试集识别率/%
0	99.998	96.810
0.05	99.960	97.373
0.10	99.847	97.668
0.15	99.657	97.929
0.20	99.341	98.012
0.25	98.871	98.229
0.30	98.121	98.292
0.40	95.241	98.361

由表 2 可知，当 Dropout 率为 0 时，训练集识别率接近 100%，但测试集的识别率只有 96.810%，说明此时网络在训练过程中已出现过拟合. 随着 Dropout 率的不断增加，训练识别率逐渐降低，但测试识别率却不断上升，因此，网络的推广能力越来越强. 当 Dropout 率为 0.40 时，虽然训练集的识别率只有 95.241%，但此时的测试集的识别率达到最高，为 98.361%. 实验表明，Dropout 技术在一定程度上可以防止网络出现过拟合，提高神经网络的泛化能力.

同时本文还将探究在不同的网络层上进行 Dropout 的效果. 以下实验使用 0.40 作为 Dropout 率，数据如表 3 所示.

表 3　不同 Dropout 层的结果

Table 3　Results of different Dropout layers

Dropout 层	训练识别率/%	测试错误率/%
—	99.998	3.19
输入层	99.643	2.32
隐藏层	98.695	2.03
输入层 + 隐藏层	95.241	1.64

当同时对输入层和所有的隐藏层进行 Dropout 时，效果比仅仅在输入层或仅仅在隐藏层 Dropout 时要好，相比于没有 Dropout 的情况，系统的错误率更是下降了 48.589%. 为了最大限度地优化网络的泛化性能，本文将对输入层和所有的隐藏层采用 Dropout.

最后为验证系统的鲁棒性，本文共做了 7 组试验. 每组都选用不同的 70 套训练样本、10 套校验样本和 20 套测试样本.

多组实验结果如表 4 所示，从中可见，多组实验后系统平均识别率为 98.292%，标准差几乎接近零，可见本文的印刷体识别系统有很高的准确率的同时具备很强的鲁棒性.

表 4　多组对比实验

Table 4　Comparison of multiple experiments

组数	识别率/%
第 1 组	98.361
第 2 组	98.294
第 3 组	98.177
第 4 组	98.208
第 5 组	98.360
第 6 组	98.336
第 7 组	98.310
均值	98.292
标准差	0.06753

4　结　论

采用了一种基于多尺度滑动窗的方法提取文字的梯度信息，并结合深度神经网络对印刷体汉字进行分类识别. 实验结果表明：

1）本文介绍的多尺度滑动窗梯度直方图特征相比于其他特征，有更好的分类效果.

2）同时在神经网络的输入层和隐藏层使用 Dropout 技术，防止了网络训练出现过拟合，增强深度神经网络的泛化性能.

3）通过选用不同的训练集和测试集进行试验，证明本系统具有较强的鲁棒性.

今后的研究工作将从以下几个方面进行改进和提高：

1）在特征提取上，尝试选择不同的特征参数，探究不同参数对汉字识别效果的影响，同时实现多种有效特征的融合，以提高汉字识别准确率.

2）探究适合文字识别的深度神经网络结构，尝试利用深度学习如深度置信网络或自动编码机等方法去学习更有效的特征.

参考文献（References）

[1] Mori K, Masuda I. Advances in recognition of Chinese characters [C]//Proceedings of the Fifth International Conference on Pattern Recognition. Miami: IEEE Computer Society Press, 1980: 692-702.

[2] 丁晓青. 汉字识别研究的回顾[J]. 电子学报, 2002, 30(9): 1364-1368.
Ding X Q. Chinese character recognition: a review[J]. Journal of Acta Electronica Sinica, 2002, 30(9): 1364-1368(in Chinese).

[3] 荆涛, 王仲. 光学字符识别技术与展望[J]. 计算机工程, 2003, 29(2): 1-2.
Jing T, Wang Z. A survey of optical character recognition[J]. Computer Engineering, 2003, 29(2): 1-2(in Chinese).

[4] Dalal N, Triggs B. Histograms of oriented gradients for humandetection[C]//IEEE Conference on Computer Vision and Pattern Recognition. San Diego, CA: IEEE Computer Society Press, 2005: 886-893.

[5] Islam A, Hasan M R, Rahaman R, et al. Designing ANN using sensitivity & hypothesis correlation testing[C]//Computer and Information Technology. Dhaka: IEEE Computer Society Press, 2007: 1-6.

[6] Soulie F F, Viennet E, Lamy B. Multi-modular neural network architectures: applications in optical character and human face recognition[J]. International Journal of Pattern Recognition and Artificial Intelligence, 1993, 7(4): 721-755.

[7] Guyon I. Applications of neural networks to character recognition [J]. International Journal of Pattern Recognition and Artificial Intelligence, 1991, 5(1-2): 353-382.

[8] Chang H D, Wang J F, Kuo S C. A Bayesian neural network for separating similar complex handwritten Chinese characters[J]. Pattern Recognition Letters, 1994, 15(4): 403-408.

[9] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines[C]//Proceedings of the 27th International Conference on Machine Learning. Haifa: International Machine

Learning Society,2010:807-814.
[10] Hosmer Jr D W,Lemeshow S. Applied logistic regression[M]. Hoboken:John Wiley & Sons,2004:31-43.
[11] Duan K,Keerthi S S,Chu W,et al. Multi-category classification by soft-max combination of binary classifiers[M]. Berlin:Springer-Verlag Berlin Heidelberg,2003:125-134.
[12] Hinton G E,Srivastava N,Krizhevsky A,et al. Improving neural networks by preventing co-adaptation of feature detectors[EB/OL]. [2014-04-14]. http://arxiv.org/abs/1207.0580.
[13] Krizhevsky A,Sutskever I,Hinton G E. Image net classification with deep convolutional neural networks[J]. Neural Information Processing Systems,2012,25(2):1097-1105.
[14] Dahl G E,Sainath T N,Hinton G E. Improving deep neural networks for LVCSR using rectified linear units and dropout[C]// IEEE International Conference on Acoustics,Speech and Signal Processing(ICASSP). Vancouver,BC:IEEE Computer Society Press,2013:8609-8613.
[15] Volodymyr M. Cudamat:a CUDA-based matrix class for python. Rep. UTML-TR-2009-004[R]. Toronto:University of Toronto,2009.
[16] Ojala T,Pietikäinen M,Harwood D. Performance evaluation of texturemeasures with classification based on Kullback discrimination of distributions[C]//International Conference on Pattern Recognition. Jerusalem,Israel:IEEE Computer Society Press,1994:582-585.
[17] Ojala T,Pietikäinen M,Harwood D. A comparative study of texture measures with classification based on featured distributions[J]. Pattern Recognition,1996,29(1):51-59.
[18] Siagian C,Itti L. Gist:a mobile robotics application of context-based vision in outdoor environment[C]//Computer Vision and Pattern Recognition-Workshops. San Diego,CA:IEEE Computer Society Press,2005:88.

Recognition of Chinese characters based on multi-scale gradient and deep neural network

PAN Weishen, JIN Lianwen*, FENG Ziyong

(Department of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China)

Abstract: The method to extract the gradient histogram feature of the Chinese characters with a multi-scale sliding window and to recognize the printed Chinese characters with deep neural network was presented. In order to acquire the spatial information of the gradient histogram, a retractable sliding window technique was proposed for segmenting the images and getting the gradient feature information from different scales which can effectively combine all the global features and local block features of Chinese characters. The experiment was carried out by using a 5-layer deep neural network to classify 3 755 categories of printed Chinese characters. A Dropout technique was applied so as to prevent over-fitting training and to improve the generalization ability of the neural network. The accuracy of the experiment reaches 98.292%, which has better recognition performance and demonstrates that the method of applying a multi-scale gradient feature and deep neural network model on the recognition of Chinese characters is effective.

Key words: multi-scale sliding window; gradient histogram; deep neural network; generalization ability; recognition of Chinese characters

北航学报
赠 阅

《北京航空航天大学学报》征稿简则

《北京航空航天大学学报》是北京航空航天大学主办的以航空航天科学技术为特色的综合性自然科学学术期刊(月刊)。本刊以反映航空航天领域研究成果与动态、促进学术交流、培养科技人才和推动科技成果向社会生产力转化为办刊宗旨。本刊为中国自然科学技术核心期刊,并被 Ei Compendex 等国内外权威文献检索系统收录。本刊向国内外公开发行,为进一步提高办刊质量和所刊出文章的学术水平,特制定本简则。

1 论文作者及内容

1.1 本刊面向海内外所有学者。

1.2 主要刊载与航空航天科学技术有关的材料科学及工程、飞行器设计与制造、宇航科学与工程、信息与电子技术、控制技术和自动化工程、流体力学和动力工程、计算机科学及应用技术、可靠性工程与失效分析等领域的研究文章。航空航天科学技术民用方面以及具有航空航天工程背景的应用数学、应用物理、应用力学和工程管理等方面的文章也在本刊优先考虑之列。

2 来稿要求

2.1 论文应具有创新性、科学性、学术性和可读性。

2.2 论文为原创作品,尚未公开发表过,并且不涉及泄密问题。若发生侵权或泄密问题,一切责任由作者承担。

2.3 主题明确,数据可靠,图表清晰,逻辑严谨,文字精练,标点符号正确。

2.4 文稿撰写顺序:中文题名(一般不超过 20 个汉字),作者中文姓名、单位、所在城市、邮政编码,中文摘要(包括目的、方法、结果及结论),中文关键词(5~8 个),中图分类号,英文题名,作者英文姓名、单位、所在城市、邮政编码、国别,英文摘要,英文关键词,引言,正文,参考文献。首页下角注明基金项目名称及编号,作者信息。

2.5 作者请登录本刊网页进行在线投稿。

3 稿件的审核、录用与版权

3.1 来稿须经专家两审和主编、编委讨论后决定刊用与否。

3.2 若来稿经审查后认定不宜在本刊发表,将及时告知作者。如果在投稿满 3 个月后仍未收到本刊任何通知,作者有权改投它刊。在此之前,请勿一稿多投,否则一切后果自负。

3.3 来稿一经刊登,即赠送单行本。

3.4 来稿一经作者签字并在本刊刊出,即表明所有作者都已经认可其版权转至本刊编辑部。本刊在与国内外文献数据库或检索系统进行交流及合作时,不再征询作者意见。

邮寄地址:100191 北京市海淀区学院路 37 号 北京航空航天大学学报编辑部
办公地点:北京航空航天大学办公楼 405,407,409 房间
电 话:(010)82315594,82338922,82314839,82315426
E-mail:jbuaa@ buaa. edu. cn
http://bhxb. buaa. edu. cn
http://www. buaa. edu. cn

北京航空航天大学学报

北航学报
赠 阅

《北京航空航天大学学报》
第五届编辑委员会

主　任（主　编）：赵沁平

（以下按姓氏笔划为序）

副主任（副主编）：丁希仑　王少萍　孙志梅　李秋实　李焕喜　杨嘉陵
苗俊刚　相　艳　徐立军　钱德沛　曹晋滨

编　委：马殿富　王　琪　王　聪　邓小燕　王青云　王荣明　刘　宇
刘　红　江　洁　刘　强　闫　鹏　朱天乐　刘铁钢　齐铂金
陈万春　邹正平　苏东林　杨世春　沈成平　邱志平　宋知人
杨树斌　张晓林　杨晓奕　杨继萍　李惠峰　吴新开　张瑞丰
杨照华　宋凝芳　周　锐　林宇震　林贵平　战　强　姚仰平
胡庆雷　赵秋红　段海滨　赵巍胜　席　平　郭　宏　徐　洁
徐世杰　郭洪波　康　锐　翟　锦　熊华钢

北京航空航天大学学报
Beijing Hangkong Hangtian Daxue Xuebao
（原《北京航空学院学报》）
（月刊　1956 年创刊）
第 41卷 第4期　2015年4月

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUAA)
(Monthly, Started in 1956)
Vol.41 No.4 April 2015

主管单位　中华人民共和国工业和信息化部
主办单位　北京航空航天大学
主　　编　赵沁平
编辑出版　《北京航空航天大学学报》编辑部
邮　　编　100191
地　　址　北京市海淀区学院路 37 号
印　　刷　北京科信印刷有限公司
发　　行　北航文化传媒集团
发行范围　国内外发行
联系电话　（010）82315594　82338922　82314839
电子信箱　jbuaa@buaa.edu.cn

Administrated by　Ministry of Industry and Information Technology of the People's Republic of China
Sponsored by　Beijing University of Aeronautics and Astronautics (BUAA)
(Beijing 100191, P. R. China)
Chief Editor　Zhao Qinping
Edited and Published by　Editorial Board of JBUAA
Printed by　Beijing Kexin Printing Co., Ltd.
Distributed by　BUAA Culture Media Group Limited
Telephone　(010) 82315594　82338922　82314839
E-mail　jbuaa@buaa.edu.cn
http://bhxb.buaa.edu.cn

刊　　号　ISSN 1001-5965 / CN 11-2625/V
国内定价　20.00 元／期

ISSN 1001-5965
9 771001 596151 04>