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

2018-01
Vol.44 No.1

北京航空航天大学学报

第 44 卷　第 1 期　（总第 299 期）　2018 年 1 月

北航学报
赠阅

目　次

地标自主导航的高精度地标库建立方法 …… 杨博，张典律，赵晓涛（1）
考虑单元界面关联特性的微定位平台刚度建模 …… 宫金良，郑洋洋，张彦斐（10）
基于 TVP-VAR 模型的多模式交通需求耦合分析 …… 马晓磊，孙硕，丁川，王云鹏（18）
面向高分辨率遥感影像的地形辐射校正方法 …… 丁一帆，尤红建，张浩，陈双军，许斌，孙韬（27）
原子自旋陀螺气室加热电磁噪声抑制实验研究 …… 周斌权，郝杰鹏，梁晓阳，全伟，刘刚（36）
实时单核和谐周期分区系统时间窗口分配算法 …… 尚利宏，谭特，周密（43）
共轴刚性双旋翼非定常气动干扰载荷分析 …… 谭剑锋，孙义鸣，王浩文，林长亮（50）
基于抛物方程的低空空域监测雷达城市环境地杂波强度分布建模 ……
…… 雷鹏，冉志强，王俊，刘晓敏（63）
基于分布式动力的翼身融合飞机整流罩气动设计 …… 项洋，吴江浩，熊峻江（71）
频率步进 SAR 虚拟阵列模型成像方法 …… 冉达，尹灿斌，贾鑫（82）
飞翼无人机机动飞行非线性鲁棒控制方法 …… 李继广，陈欣，李亚娟，张榕（89）
液氮贮箱常压停放实验与数值仿真 …… 李佳超，梁国柱（99）
星载 GNSS-R 辅助海洋盐度探测方法 …… 王峰，杨东凯（108）
一种增量并行式动态图异常检测算法 …… 韩涛，兰雨晴，肖利民，刘艳芳（117）
GNSS 海面反射信号的三维建模方法 …… 祁永强，张波，杨东凯，张彦仲，张健敏（125）
纠错输出编码的留一误差界估计 …… 薛爱军，王晓丹（132）
纤维金属层板金属层应变测量及应力预测方法 ……
…… 孟维迎，谢里阳，胡杰鑫，吕骁，秦波，王博文（142）
基于傅里叶域卷积表示的目标跟踪算法 …… 朱日东，杨小远，王敬凯（151）
基于 CT 的泡沫铝三维细观模型重建及应用 …… 李侯贞强，张亚栋，张锦华，姜春琳（160）
脉冲星角位置对脉冲模板的影响及其削弱策略 …… 信世军，郑伟，王奕迪（169）
无人机室内视觉/惯导组合导航方法 …… 王亭亭，蔡志浩，王英勋（176）
孔边倒角裂纹当量初始缺陷尺寸的确定方法 …… 孙晓娜，贺小帆，李玉海（187）
基于空间混联机构的人体肩部骨骼运动模型 …… 聂超，宋智斌，戴建生（196）
随机优化的改进交叉熵方法 …… 任超，张航，李洪双（205）

期刊基本参数： CN 11-2625/V * 1956 * m * A4 * 214 * zh * P * ￥50.00 * 900 * 24 * 2018-01

（**编　辑**　娄　嘉　张　嵘　李　晶　张欣蔚　孙　芳　王艳梅）

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Vol. 44 No. 1 (Sum 299) January 2018

北航学报 赠阅

CONTENTS

A building method of high-precision landmark library for landmark autonomous navigation
…… YANG Bo, ZHANG Dianlyu, ZHAO Xiaotao (1)
Stiffness modeling of micro-positioning platform considering unit interface relevance characteristics
…… GONG Jinliang, ZHENG Yangyang, ZHANG Yanfei (10)
A coupling analysis of multimodal transportation demands based on TVP-VAR models
…… MA Xiaolei, SUN Shuo, DING Chuan, WANG Yunpeng (18)
Topographic correction method for high-resolution remote sensing images
…… DING Yifan, YOU Hongjian, ZHANG Hao, CHEN Shuangjun, XU Bin, SUN Tao (27)
Experimental study on electromagnetic noise suppression of atomic spin gyroscope heating chamber
…… ZHOU Binquan, HAO Jiepeng, LIANG Xiaoyang, QUAN Wei, LIU Gang (36)
Time windows distribution algorithm for real-time harmonic-period partition system on uniprocessor platform
…… SHANG Lihong, TAN Te, ZHOU Mi (43)
Analysis of rigid coaxial rotor unsteady interactional aerodynamic loads
…… TAN Jianfeng, SUN Yiming, WANG Haowen, LIN Changliang (50)
Parabolic equation based land clutter power map modeling for low-altitude surveillance radar in urban areas
…… LEI Peng, RAN Zhiqiang, WANG Jun, LIU Xiaomin (63)
Aerodynamic design of nacelle of blended-wing-body aircraft with distributed propulsion
…… XIANG Yang, WU Jianghao, XIONG Junjiang (71)
Imaging approach for frequency-stepped SAR with virtual array model
…… RAN Da, YIN Canbin, JIA Xin (82)
Nonlinear robust control method for maneuver flight of flying wing UAV
…… LI Jiguang, CHEN Xin, LI Yajuan, ZHANG Rong (89)
Experiment and numerical simulation of liquid nitrogen tank atmospheric ground parking
…… LI Jiachao, LIANG Guozhu (99)
Sea surface salinity determination method assisted by spaceborne GNSS-R
…… WANG Feng, YANG Dongkai (108)
Incremental and parallel algorithm for anomaly detection in dynamic graphs
…… HAN Tao, LAN Yuqing, XIAO Limin, LIU Yanfang (117)
Three-dimensional modeling method of GNSS sea surface reflection signal
…… QI Yongqiang, ZHANG Bo, YANG Dongkai, ZHANG Yanzhong, ZHANG Jianmin (125)
Leave-one-out error bounds estimation for error correcting output codes
…… XUE Aijun, WANG Xiaodan (132)
Strain measurement and stress prediction methods of metal layer in fiber metal laminates
…… MENG Weiying, XIE Liyang, HU Jiexin, LYU Xiao, QIN Bo, WANG Bowen (142)
Convolution representation-based object tracking algorithm in Fourier domain
…… ZHU Ridong, YANG Xiaoyuan, WANG Jingkai (151)
Reconstruction and application of three-dimensional mesoscopic model of aluminum foam based on CT
…… LI Houzhenqiang, ZHANG Yadong, ZHANG Jinhua, JIANG Chunlin (160)
Impact of pulsar angular position on pulse template and its compensation method
…… XIN Shijun, ZHENG Wei, WANG Yidi (169)
Integrated vision/inertial navigation method of UAVs in indoor environment
…… WANG Tingting, CAI Zhihao, WANG Yingxun (176)
Determination method of equivalent initial flaw size for crack initiated at hole chamfering
…… SUN Xiaona, HE Xiaofan, LI Yuhai (187)
A shoulder skeletal kinematic model based on spatial hybrid mechanism
…… NIE Chao, SONG Zhibin, DAI Jiansheng (196)
Stochastic optimization method based on improved cross entropy
…… REN Chao, ZHANG Hang, LI Hongshuang (205)

北京航空航天大学学报

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0021

地标自主导航的高精度地标库建立方法

杨博*，张典律，赵晓涛

（北京航空航天大学 宇航学院，北京 100083）

摘 要：利用地标信息的卫星自主导航方法是通过星载光学摄像机实时拍摄地标景象来计算卫星的位置、速度等参数，具有完全自主的特点，并且适用范围广泛，可用于能够周期性获得地面图像的航天器，但受到地标选取方法以及地标库建立策略的制约，影响到导航精度，阻碍了推广使用。针对上述问题，提出了建立全局及高性能的地标库，即采用全球地标控制点选取的原则选取性能优良的全球地标样本，并自动生成全局地标库。通过计算机仿真验证了该方法的有效性，结果表明，使用全球地标控制点选取地标样本，并自动生成全局地标库，可以快速而准确地为航天器进行高精度的自主导航，导航位置误差约为 99 m，速度误差约为 0.08 m/s。

关 键 词：卫星自主导航；地标导航；地标库；灰度场特征参数；最大相关系数（MCC）法
中图分类号：V448
文献标识码：A **文章编号**：1001-5965(2018)01-0001-09

基于地标的航天器自主导航是一种利用已标定的地面标志物进行自主定位的导航方式，并具有精度高、自主性强的特点。该自主导航系统的误差不随时间积累，且在每个周期内都能稳定获得足够的导航信息，因而可以辅助各类航空器、临近空间飞行器的惯性导航修正误差，或用于周期性地在轨对地观测卫星的几何精度校验。

1975 年，Kau[1] 提出了一种利用地面线型地标进行导航的方法。通过采用人工选取地标的方法，进行交互式地标导航工作[2-4]，需要做大量的重复性工作，并不适用于对大数据量导航图像进行实时自动处理的遥感卫星[5]。同时，人工选取地标的方法涉及大量的人工操作，缺乏一致性标准，工作人员的主观性会影响地标选取的质量，直接导致导航精度大幅度下降。另外，地标匹配误差对全局的遥感应用产生严重的影响[6-7]，因此自动的地标生成和识别是必要的[8]。

国内外科学家对自动地标生成技术进行了大量的研究。Emery 等[9] 于 2003 年提出了基于最大相关系数（Maximum Correlation Coefficient，MCC）的自动地标导航方法，由于该方法具有易于实现、计算速度快、精度高的优点，其已成为目前为止应用范围最为广泛的地标建立方法。郭强等[10] 优化了最大相关系数法的计算流程，降低了算法的复杂度。但是与人工地标导航类似，基于最大相关系数的自动地标导航方法仍然需要依据大量历史遥感资料，人工制定不同季节、不同光照条件的模板，同样依赖于地标选取人员的经验。美国空间科学与工程中心（Space Science and Engineering Center，SSEC）对静止轨道 GOES 卫星设计了自动地标导航系统[11]，认为当匹配模板的细节信息量不够时，Emery 等[9] 的最大相关系数法不适用。卢耀秋[12]、赵礼铮和白光弼[13] 分别在静止气象卫星和极轨气象卫星上使用了地标导航方法，但仍然是人工交互式的方法。杨磊等[8,14] 研究了基于最大相关系数法的自动地标导航方

收稿日期：2017-01-16；**录用日期**：2017-04-24；**网络出版时间**：2017-05-16 16:43
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170516.1643.002.html
* **通信作者**. E-mail：yangbo@buaa.edu.cn

引用格式：*杨博，张典律，赵晓涛. 地标自主导航的高精度地标库建立方法[J]. 北京航空航天大学学报，2018，44(1)：1-9.*
YANG B，ZHANG D L，ZHAO X T. A building method of high-precision landmark library for landmark autonomous navigation[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(1)：1-9 (in Chinese).

北航学报 赠阅

法,提出了将地面图像建立控制点并自动生成地标的方法,但其地标选取的范围也局限于海岸线和河流地标。

针对以上地标导航中地标建立的局限性问题,本文提出了一种基于全局地标信息的选取原则及导航方法,即将地标遥感图像的边界和粗糙纹理作为生成地标图像的基准,通过筛选全球模板建立地标库,利用建立的地标库进行自动地标匹配,进而有效地提高了自主地标导航的精度。

1 地标导航原理

图 1 为基于地标的卫星自主导航原理图。其工作过程是:将飞行器上预先装订的标准地标库中的模板与卫星实时拍摄图像进行地标匹配,使用最大相关系数法得到若干个已标定的地标在遥感图像上的位置坐标。Rosborough 等[15] 认为要实现准确的地标导航需要至少匹配得到 2 个地标。将地标在遥感图像上的坐标转化为视线矢量,结合轨道动力学模型进行导航解算,获得位置、速度等导航信息。

该导航方法使用的前提是在获取遥感卫星数据时,要求卫星姿态的预报或测量是准确的,否则卫星姿态的误差将会影响到位置误差及速度误差的测量。同时,遥感图像中存在大量不确定形状的云层,严重影响地标匹配,因此,在地标匹配之前需要对拍摄图像进行严格的云检测。

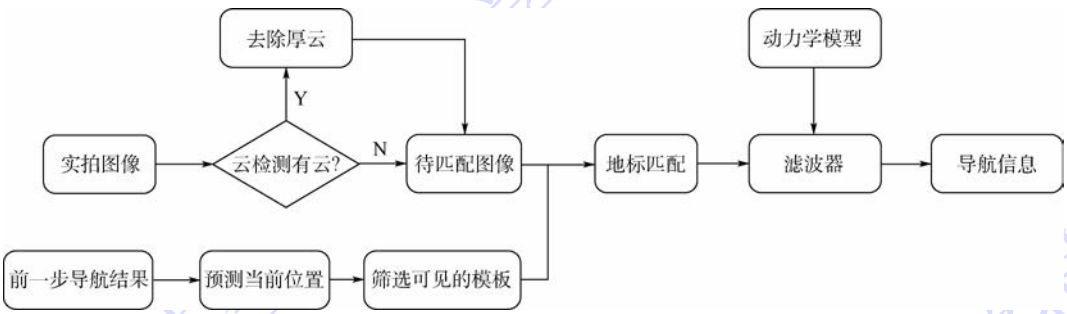


图 1 地标导航原理图

Fig. 1 Schematic of landmark navigation

2 全局地标库的建立与匹配算法

2.1 地标的选取原则

在基于地标的自主导航方案中,地标库的生成将直接影响到导航精度。全局地标库需要建立不同光照条件下的地标模板,而对于每一个地标,需要建立至少 8 个模板(白昼、黑夜,春、夏、秋、冬)[9]。这些地标图像需要具备以下 2 个特征:①地标图像具有清晰的结构特征;②地标图像与待匹配的卫星遥感图像具有相同的投影方式。

具体做法如下:

1)地标库主要由湖泊、河流、海岸线、岛屿组成,由于水域和陆地的灰度值差异很大,可以容易地获取水陆边界的结构特征作为地标。

2)为了减少误匹配的概率,选取水陆边界地标的结构特征必须是独特的,由于不确定卫星拍摄图像相对于模板的角度,所以不能使用水陆边界斜率作为特征。

3)使用边界中高曲率和拐点区域作为生成水陆地标的基础。使用世界数据银行(World Data Bank)和世界海岸线矢量库(World Vector Shoreline)中的数据[8],自动计算地标轮廓拐点,"T"型连接点作为控制点,用于扩展成地标,如图 2[8] 所示。

4)在非水陆地标区域,存在大量不规则灰度剧烈变化区域,无法筛选出高曲率点,但其中有大量的不规则天然粗纹理可以提供唯一的辨识特征[16],可以作为这一区域的主要结构特征,很好地补充海陆地标。

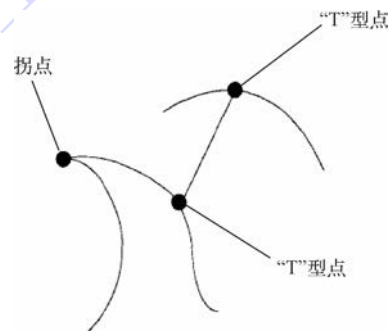


图 2 地面控制点的定义[8]

Fig. 2 Definition of control point on earth surface[8]

2.2 全局地面控制区的建立

将拥有足够多粗糙纹理的区域定义为控制区,并作为形成地标块的基础。在选取控制区的过程中,没有考虑地面纹理的具体特征,地形的粗糙程度体现的纹理已足以提供导航信息[17]。因

此,无需使用计算量大的灰度共生矩阵来描述图像纹理特性。设大小为 $M \times N$ 的图像 $f(x,y)$,选取 $m \times n$ 的计算局部窗口,用以下灰度场特征参数描述地面粗糙程度。

1)灰度标准差。描述灰度值偏离平均值的程度,可以宏观反映特征值整体的起伏程度。

$$\begin{cases} \sigma = \sqrt{\dfrac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(f(i,j)-\bar{f})^2} \\ \bar{f} = \dfrac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}f(i,j) \end{cases} \tag{1}$$

2)灰度粗糙度。描述灰度值变化的剧烈程度,粗糙度越大,局部的灰度起伏约剧烈。x 轴和 y 轴方向的粗糙度分别为

$$\begin{cases} r_x = \dfrac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n-1}|f(i,j)-f(i,j+1)| \\ r_y = \dfrac{1}{n(m-1)}\sum_{i=1}^{m-1}\sum_{j=1}^{n}|f(i,j)-f(i+1,j)| \end{cases} \tag{2}$$

3)局部灰度相关系数。描述局部的灰度相关度,如果地标本身相关性很大,将大大影响地标匹配的准确度。x 轴和 y 轴方向的局部灰度相关系数分别为

$$\begin{cases} R_x = \dfrac{1}{m(n-1)\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n-1}[(f(i,j)-\bar{f})(f(i,j+1)-\bar{f})] \\ R_y = \dfrac{1}{n(m-1)\sigma^2}\sum_{i=1}^{m-1}\sum_{j=1}^{n}[(f(i,j)-\bar{f})(f(i+1,j)-\bar{f})] \end{cases} \tag{3}$$

实际上,在表达地形的局部变化情况时,灰度粗糙度和局部灰度相关系数是一致的[17]。

图 3 为中哈边境天山山脉附近的遥感图像,是典型的具有大量混合地貌的遥感图像,除了通常可作为地标的水陆边界地貌之外,还包括了雪山、草原、森林、荒漠、河流等多种地貌,是全球具有代表性的模板图像。

根据 2.1 节中的选取地标原则,将图 3 灰度化,结果如图 4 所示。

以图 4 为基础,根据本节内容,取 m、n 的值为 5,计算图像的地形标准差和地形粗糙度,得到的标准差和粗糙度如图 5 ~ 图 6 所示。

将灰度标准差作为衡量局部地形的特征数量指标,并参考灰度粗糙度建立控制区域。若某一区域的灰度标准差和灰度粗糙度 2 个参数都大于某个阈值,则可以将这一区域作为候选控制区域。

具体在建立全球地标模板库时,选取灰度标准差阈值为 30,灰度粗糙度阈值为 6,可以得到图 7所示的候选控制区域(白色高亮)。

由图 7 看出,使用灰度差作为特征参数选取的控制区域内包含了大量的水陆边界地标,此类地标需要计算控制点后扩展为可用地标。

依据以上方法挑选出的控制区域主要包含以下 3 类:

1) 2 种易于区分的地形或地貌的边界地区,如海岸线、岛屿、沙漠边界等,如图 8 所示。



图 3　混合地貌遥感图像

Fig. 3　Remote sensing image of mixed landscape

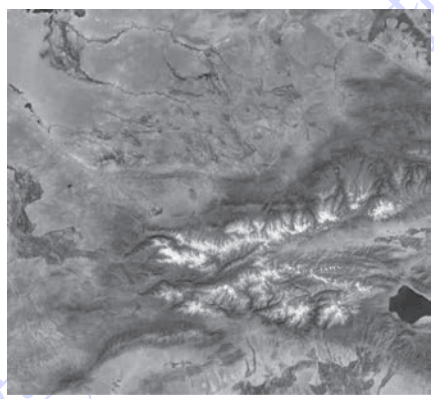


图 4　混合地貌遥感图像的灰度化结果

Fig. 4　Grayscale results of remote sensing image of mixed landscape

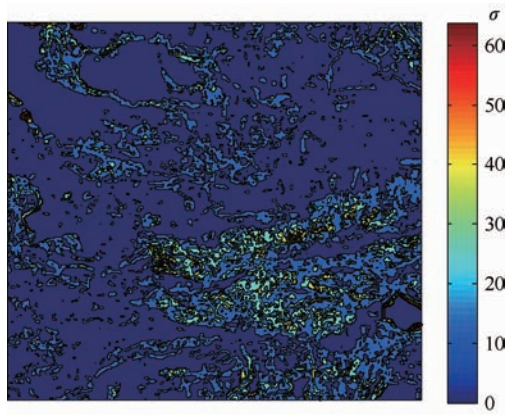


图 5　灰度标准差

Fig. 5　Standard deviation of grayscale

北航学报 赠阅

2）较为均匀地貌中出现较为明显变化的区域，如河流、湖泊、山脉等，如图 9 所示。

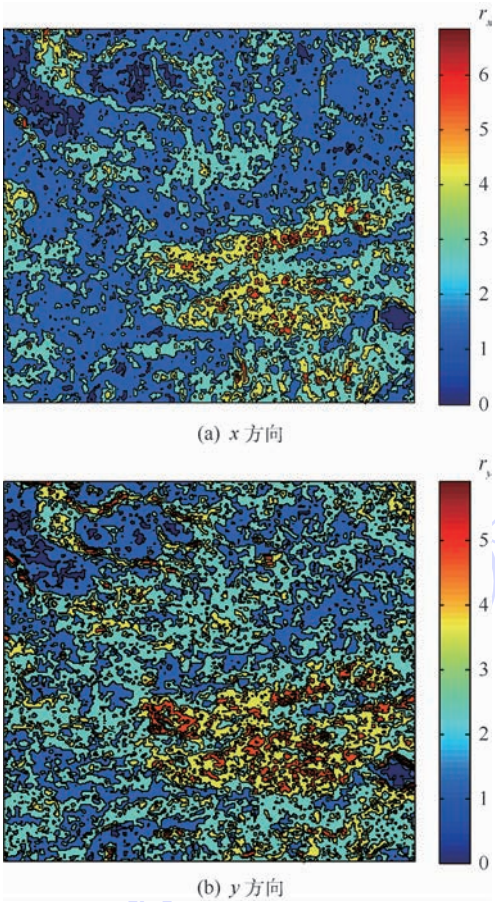


图 6 灰度粗糙度

Fig. 6 Roughness of grayscale

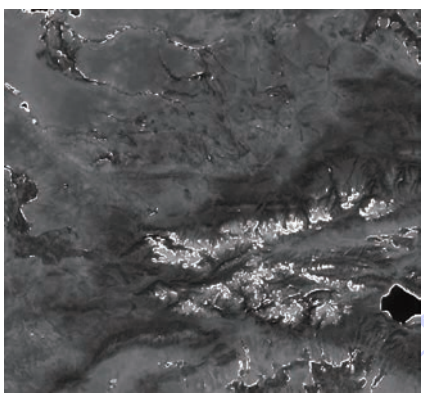


图 7 候选控制区域

Fig. 7 Selected ground control area



图 8 典型地貌边界地标

Fig. 8 Typical landmarks of topographic boundary

3）有较为明显的粗纹理地区，如戈壁、森林、丘陵等，如图 10 所示。

综上所述，首先计算控制区域并筛选出可用于生成地标的控制区域，再以这些控制区域为基础，扩展成为用于匹配的全局地标图像库（地标块大小为 30 × 30）。使用这一方法获得的地标库的地标数据量比单纯的水陆边界地标库增加了 43.97%。



图 9 典型地貌变化地标

Fig. 9 Typical landmarks of topographic diversification



图 10 典型纹理地标

Fig. 10 Typical texture landmarks

2.3 地标匹配算法

地标匹配算法采用灰度相关系数法，其具有运算速度快、准确率高的特点。

设大小为 $M \times N$ 的卫星图像 $f(x,y)$ 和大小为 $J \times K$ 的模板图像 $w(x,y)$，如图 11 所示。在图像 $f(x,y)$ 中寻找与模板 $w(x,y)$ 相匹配的区域，即将模板作为一个空间滤波器在卫星图像 $f(x,y)$ 中求取每个位置上的相关系数 $R(x,y)$。使用归一化最大互相关（normalized maximum cross correlation）系数表达二者的相关性，具体表示为

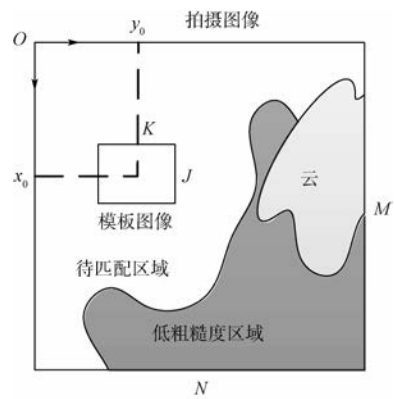


图 11 地标模板在卫星图像上的匹配

Fig. 11 Matching of landmark template on satellite image

北航学报 赠阅

$$R(x,y)=\sum_s\sum_t\left(f(s,t)-\bar{f}(s,t)\right)\left(w(x+s,y+t)-\bar{w}\right)\Big/\left\{\left[\sum_s\sum_t\left(f(s,t)-\bar{f}(s,t)\right)^2\right]^{1/2}\cdot\left[\sum_s\sum_t\left(w(x+s,y+t)-\bar{w}\right)^2\right]^{1/2}\right\} \tag{4}$$

$R(x,y)$值越接近 1,表示图像的相似度越高,匹配精度越高。同时应用 t 检验:当显著水平小于 0.05 为误匹配,此时相关系数 $R(x,y)$ 的阈值为 0.72;当相关系数 $R(x,y)$ 取到最大值 $R_m(x_m,y_m)$ 且超过 0.72 的位置,就是卫星图像与地标模板图像相匹配的位置[8]。

对全球模板进行匹配的计算量极大,需要在匹配之前对卫星拍摄图像进行处理,地标匹配的具体步骤如下:

1) 对拍摄图像进行严格的云检测,剔除厚云的部分。

2) 将剩余部分依照式(1)和式(2)计算灰度标准差和灰度粗糙度,本文选取标准差阈值为 15,粗糙度阈值为 1.2,剔除图像中标准差和粗糙度同时小于阈值的部分。

3) 在地标库中选择对应光照条件(白昼、黑夜,春、夏、秋、冬)部分的地标,依据卫星图像幅宽,选取坐标处于卫星星下点附近幅宽范围内的地标。

4) 将选出的地标与拍摄图像中剩余部分的每一个可能的位置进行相关系数计算,相关系数取到最大值且超过阈值 0.72 的位置就是地标匹配的位置。

3　地标导航

3.1　地标导航观测方程

对于给定的时刻,通过对地观测设备得到实时的拍摄图像,在与地标库中的地标进行匹配后,能得到若干个已知具体地理位置的地标在实时拍摄图像上的位置坐标。假设 P_i 为实时拍摄图像上匹配后得到的第 i 个地标的中心,在实时拍摄图像上的坐标为(x_i,y_i),如图 12 所示。

为推导计算方便,假定相机坐标系和卫星本体坐标系重合,即相机安装矩阵为单位矩阵,则地标在卫星本体坐标系中的单位地标视线矢量为地标中心在相机像平面上的坐标加上相机焦距,表示为

$$\boldsymbol{u}_i^{\mathrm{B}}=\frac{1}{\sqrt{(x_i d_{\mathrm{p}})^2+(y_i d_{\mathrm{p}})^2+f^2}}\begin{bmatrix}x_i d_{\mathrm{p}}\\ y_i d_{\mathrm{p}}\\ f\end{bmatrix} \tag{5}$$

式中:f 为 CCD 相机焦距;d_{p} 为 CCD 器件像元尺寸。转换到惯性坐标系为

$$\boldsymbol{u}_i^{\mathrm{I}}=\boldsymbol{T}_{\mathrm{IO}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_i^{\mathrm{B}} \tag{6}$$

式中:$\boldsymbol{A}$ 为卫星的姿态矩阵;$\boldsymbol{T}_{\mathrm{IO}}$ 为质心轨道坐标系到惯性坐标系的变换矩阵。

卫星除去在大面积的海洋上空运行外,都可以在实时拍摄图像中匹配得到多个地标,卫星和地标的空间关系图如 13 所示。

图 13 中,$\boldsymbol{R}_{\mathrm{V}}$ 为卫星位置,$\boldsymbol{R}_i$ 为匹配得到的第 i 个地标的位置,与单位地标视线矢量之间的关系可以表示为

$$\boldsymbol{R}_i=d\boldsymbol{u}_i+\boldsymbol{R}_{\mathrm{V}} \tag{7}$$

式中:地标位置矢量 $\boldsymbol{R}_i$ 由地标库已知。同时,地标视线矢量的长度 d 与单位地标视线矢量 $\boldsymbol{u}_i$ 以及地标位置矢量 $\boldsymbol{R}_i$ 满足余弦定理,如下:

$$d=\boldsymbol{R}_i\cdot\boldsymbol{u}_i+\sqrt{\boldsymbol{R}_i^2-\boldsymbol{R}_{\mathrm{V}}^2+(\boldsymbol{R}_i\cdot\boldsymbol{u}_i)^2} \tag{8}$$

式中:卫星位置 $\boldsymbol{R}_{\mathrm{V}}$ 可根据动力学一步估计获得。

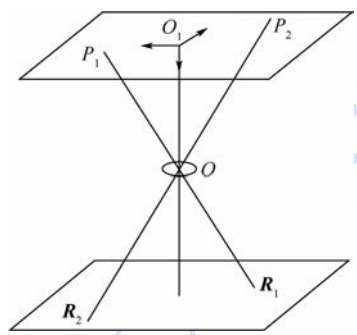


图 12　相机成像的几何关系

Fig. 12　Geometry of camera imaging

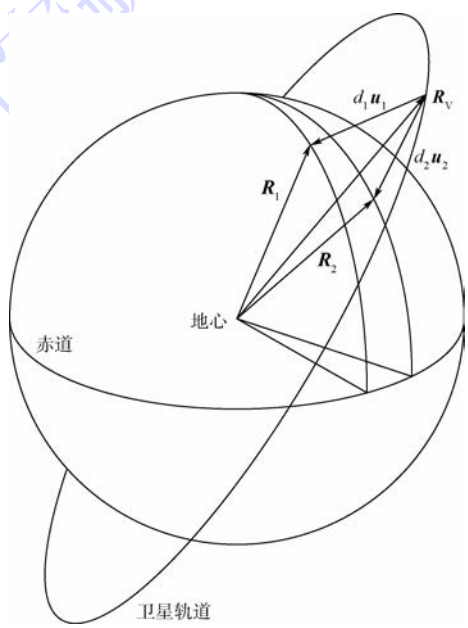


图 13　地标导航的相关矢量关系

Fig. 13　Related vector relation in landmark navigation

北航学报 赠阅

取相机坐标系下的地标单位视线矢量 $\boldsymbol{u}_i$ 作为量测量，由式(7)和式(8)推得量测方程为

$$\boldsymbol{Z}(k)=\boldsymbol{u}_i+\boldsymbol{v}=\frac{1}{d}(\boldsymbol{R}_i-\boldsymbol{R}_V)+\boldsymbol{v} \tag{9}$$

式中：$\boldsymbol{v}$ 为量测噪声。

3.2 卫星状态模型

选取历元(J2000.0)地心赤道惯性坐标系。只考虑引力场摄动二阶带谐项，其他摄动如日月引力摄动、太阳光压摄动、大气阻力摄动均为高阶小量。系统的状态方程为

$$\begin{cases}\dfrac{\mathrm{d}x}{\mathrm{d}t}=v_x\\ \dfrac{\mathrm{d}y}{\mathrm{d}t}=v_y\\ \dfrac{\mathrm{d}z}{\mathrm{d}t}=v_z\\ \dfrac{\mathrm{d}v_x}{\mathrm{d}t}=-\mu\dfrac{x}{r^3}\left[1-J_2\left(\dfrac{|\boldsymbol{R}_V|}{r}\right)^2\left(7.5\dfrac{z^2}{r^2}-1.5\right)\right]+\Delta F_x\\ \dfrac{\mathrm{d}v_y}{\mathrm{d}t}=-\mu\dfrac{y}{r^3}\left[1-J_2\left(\dfrac{|\boldsymbol{R}_V|}{r}\right)^2\left(7.5\dfrac{z^2}{r^2}-1.5\right)\right]+\Delta F_y\\ \dfrac{\mathrm{d}v_z}{\mathrm{d}t}=-\mu\dfrac{z}{r^3}\left[1-J_2\left(\dfrac{|\boldsymbol{R}_V|}{r}\right)^2\left(7.5\dfrac{z^2}{r^2}-1.5\right)\right]+\Delta F_z\end{cases}$$

$$r=\sqrt{x^2+y^2+z^2} \tag{10}$$

式中：x、y、z 分别为卫星三轴的位置；v_x、v_y、v_z 分别为卫星三轴的速度；μ 为引力常数；J_2 为摄动系数；ΔF_x、ΔF_y、ΔF_z 分别为除引力场摄动二阶带谐项外其他摄动因素影响之和在三轴的分量。

4 导航仿真

仿真条件为：轨道高度为 1 000 km；卫星参数如表 1 所示。表中：a 为轨道半长轴；e 为偏心率；i 为轨道倾角；ω 为近地点幅角；M_0 为平近点角。

卫星保持三轴稳定、对地定向。拍摄得到的图像参考风云系列卫星，即星下点分辨率为 1 100 m，图像幅宽为 2 900 km。假设卫星每 30 s 能获得一幅可用的图像，并且拍摄时刻均为当地时间的日间，考虑到仿真中匹配计算的时间，匹配所用的拍摄图像使用了拍摄图像中心 500 像素 ×500 像素的部分。导航地标库采用 2.2 节建立的地标库中的春季白天部分，使用文献[10]中经过优化的匹配算法，平均每 0.899 7 s 可以完成一次匹配。

导航计算中，滤波计算周期设为 5 s，并假设每步滤波最多使用 3 个匹配到的图像信息进行导航。

表 1 样例轨道的轨道要素

Table 1 Orbital elements of sample orbit

参 数	a/m	e	i/(°)	ω/(°)	M_0/(°)
数 值	7 478 137.0	0	45.0	0	0

图 14 为系统在 5 个周期内的仿真结果。其平均位置误差为 125.827 4 m(3σ)，平均速度误差为 0.104 9 m/s(3σ)。

将轨道高度从 500 ~ 2 000 km 取不同值进行仿真，仿真结果如表 2 所示。

仿真中，地标视线矢量是根据地标位置和模拟轨道中卫星的位置预先计算好的。从表 2 中数据看出，在地标导航中，轨道下降会使位置和速度误差明显减弱，这是因为轨道高度下降会使拍摄图像的分辨率提高，因而导航精度提高。

在上述仿真中，导航均限制使用 3 个地标信息，而实际对于幅宽为 2 900 km 的遥感图像而言，在仿真的 5 个周期内均有超过 70% 的图像可以获得至少 4 个导航地标信息。地标信息数量对导航精度有一定影响，如表 3 所示。

由表 3 看出，在导航中使用更多个地标信息数量并不能有效提高导航精度，但至少需要 2 个地标才能进行导航[15]，而更多的地标信息却可以加快数据的收敛，见图 15 中 3 组仿真的均方根误差 RMSE曲线。因此，当飞行器运行在容易获得

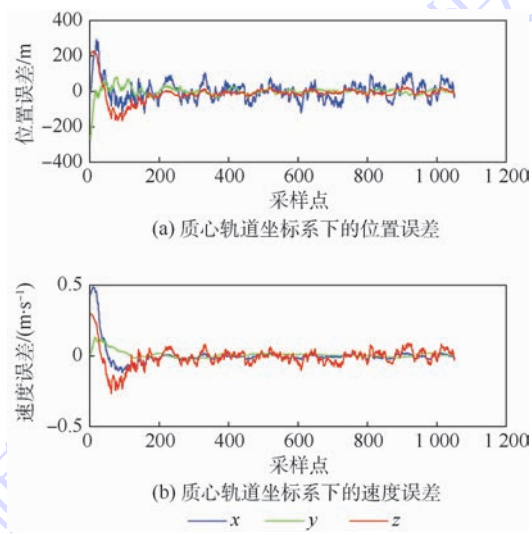


图 14 使用本文地标库进行导航的导航误差

Fig. 14 Navigation error of navigation with landmark library in the text

表 2 不同轨道高度仿真结果对比

Table 2 Comparison of simulation results of different orbital heights

轨道高度/km	平均位置误差(3σ)/m	平均速度误差(3σ)/(m·s^{-1})
500	62.713 4	0.058 2
800	87.832 8	0.080 3
1 100	99.125 5	0.086 9
1 400	123.505 0	0.109 4
1 700	165.607 7	0.130 0
2 000	192.363 7	0.141 5

地标的区域时,应以选择质量较高的地标为前提,适当控制数量以提高匹配速度;反之,即在难以获取地标信息的区域,一旦获取地标信息应当尽可能多的使用地标信息,使导航快速收敛。

图 15 为只用水陆地标导航解算结果,在相同的仿真条件下,由于缺少了大量的非水陆地标,即非全局地标,导致导航精度明显下降,平均位置误差为 144.638 8 m(3σ),平均速度误差为 0.128 2 m/s(3σ),仿真结果如图16所示。与使用本文全局地标库导航算法相比,导航位置精度和速度精度降低了 10% 和 11%。

同时,从 2 次仿真的 RMSE 曲线(见图 17)可见,收敛速度也受到了影响。

表 3　不同地标信息数量仿真结果对比

Table 3　Comparison of simulation results of different number of landmarks information

地标信息数量	平均位置误差(3σ)/m	平均速度误差(3σ)/(m·s^{-1})
2	99.917 7	0.088 2
3	99.125 5	0.086 9
4	98.898 4	0.085 2

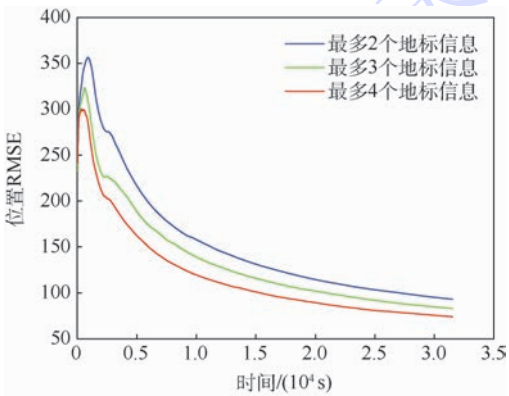


图 15　不同地标信息数量的位置 RMSE 对比

Fig. 15　Comparison of position RMSE for different number of landmark information

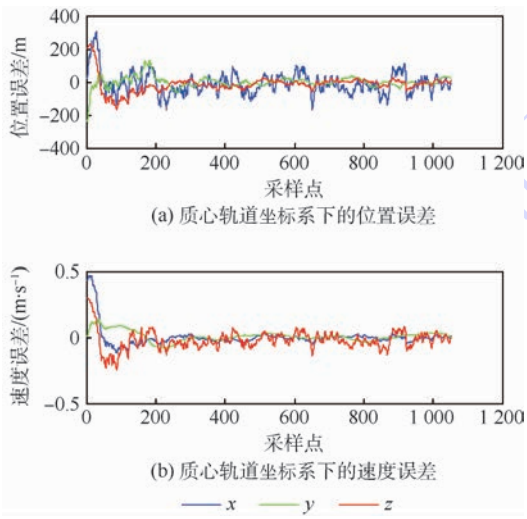


图 16　使用传统地标库进行导航的导航误差

Fig. 16　Navigation error of navigation with traditional landmark library

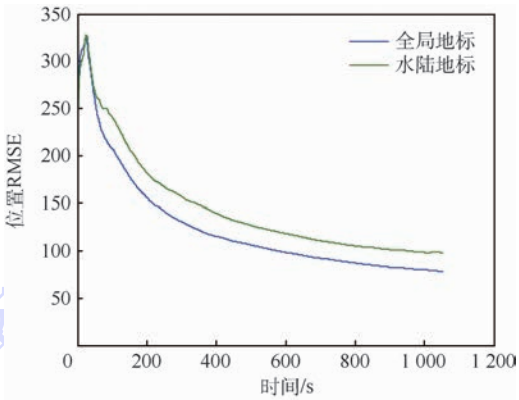


图 17　不同地标库的位置 RMSE 对比

Fig. 17　Comparison of position RMSE for different landmark libraries

5　结　论

本文针对利用地标的卫星自主导航方案中的瓶颈问题,即地标库的建立和匹配方法,提出了一种新的全局地标选取原则及匹配算法,具体工作如下:

1) 全局地标选取范围包括水陆边界以及扩展到有灰度差为特征参数的地貌区域,由此建立的地标库的数据量比单纯的水陆边界地标库增加了 43.97%。由于受季节影响较大,需要准备不同季节的地标模板,并且需要时常更新。

2) 采用拥有足够多粗糙纹理的区域的控制区域作为地标块的基础,不用考虑地面纹理的具体特征,因此可以大大减少描述图像纹理特性的灰度共生矩阵来描述图像纹理特性。

3) 使用全局地标库进行导航解算,相比于已有只用水陆边界的地标库,导航精度提高了 10%,相应速度也提高了 11%。

参考文献(References)

[1] KAU S P. Autonomous satellite orbital navigation using known and unknown earth landmarks[C]// Guidance and Control Conference. Reston: AIAA, 1975: 1-12.

[2] BACHMANN M, BENDIX J. An improved algorithm for NOAA-AVHRR image referencing[J]. International Journal of Remote Sensing, 1992, 13(16): 3205-3215.

[3] HO D, ASEM A. NOAA AVHRR image referencing[J]. International Journal of Remote Sensing, 1986, 7(6): 895-904.

[4] ILLERA P, DELGADO J A, CALLE A. A navigation algorithm for satellite images[J]. International Journal of Remote Sens-

北航学报 赠阅

ing,1996,17(3):577-588.

[5] MOIGNE J L,CAMPBELL W J,CROMP R F,et al. An automated parallel image registration technique based on the correlation of wavelet features[J]. IEEE Transactions on Geoscience and Remote Sensing,2002,40(8):1849-1864.

[6] TOWNSHEND J R,JUSTICE C O,GURNEY C E,et al. The impact of misregistration on change detection[J]. IEEE Transactions on Geoscience and Remote Sensing,1992,30(5):1054-1060.

[7] DAI X,KHORRAM S. The effects of image misregistration on the accuracy of remotely sensed change detection[J]. IEEE Transactions on Geoscience and Remote Sensing,1998,36(5):1566-1577.

[8] 杨磊,杨忠东. 极轨气象卫星自动地标导航方法[J]. 应用气象学报,2009,20(3):329-336.
YANG L,YANG Z D. The automated landmark navigation of the polar meteorological satellite[J]. Journal of Applied Meteorological Science,2009,20(3):329-336(in Chinese).

[9] EMERY W J,BALDWIN D,MATTHEWS D. Maximum cross correlation automatic satellite image navigation and attitude corrections for open-ocean image navigation[J]. IEEE Transactions on Geoscience and Remote Sensing,2003,41(1):33-42.

[10] 郭强,杨磊,赵现纲,等. 气象卫星图像导航的地标匹配算法研究与优化[J]. 计算机工程与应用,2013,49(24):152-156.
GUO Q,YANG L,ZHAO X G,et al. Research and optimization of landmark matching algorithm for meteorological satellite image navigation[J]. Computer Engineering and Applications,2013,49(24):152-156(in Chinese).

[11] KIM T,LEE T,CHOI H. Landmark extraction,matching and processing for automated image navigation of geostationary weather satellites[J]. Proceedings of SPIE,2005,5657(1):30-37.

[12] 卢耀秋. 静止气象卫星的地标导航计算方法[J]. 计算物理,1992,9(4):775-777.
LU Y Q. Landmark navigation method of geostationary meteorological satellite[J]. Chinese Journal of Computational Physics,1992,9(4):775-777(in Chinese).

[13] 赵礼铮,白光弼. 极轨气象卫星局部数据集的精地标导航[J]. 气象,1992,18(11):44-46.
ZHAO L Z,BAI G B. Navigation of precise landmarks in polar-orbit meteorological satellite local data[J]. Meteorological Monthly,1992,18(11):44-46(in Chinese).

[14] 杨磊,冯小虎,郭强,等. 风云二号气象卫星图像自动几何精校正[J]. 计算机工程与应用,2011,47(3):202-206.
YANG L,FENG X H,GUO Q,et al. Automatic geometric precision correction of Fengyun-2 meteorological satellite imagery[J]. Computer Engineering and Applications,2011,47(3):202-206(in Chinese).

[15] ROSBOROUGH G W,BALDWIN D G,EMERY W J,et al. Precise AVHRR image navigation[J]. IEEE Transactions on Geoscience and Remote Sensing,1994,32(3):644-657.

[16] 蒋春华. 浅析纹理防伪技术[J]. 中国防伪报道,2012(11):47-50.
JIANG C H. Analysis of texture anti-counterfeiting technology[J]. China Anti-Counterfeiting Report,2012(11):47-50(in Chinese).

[17] 郑彤,蔡龙飞,王志刚,等. 地形匹配辅助导航中匹配区域的选择[J]. 中国惯性技术学报,2009,17(2):191-196.
ZHENG T,CAI L F,WANG Z G,et al. Selection of matching area in terrain match aided navigation[J]. Journal of Chinese Inertial Technology,2009,17(2):191-196(in Chinese).

作者简介:

杨博 女,博士,副教授。主要研究方向:航天器天文自主导航技术。

张典律 男,硕士研究生。主要研究方向:导航制导与控制。

赵晓涛 男,硕士研究生。主要研究方向:导航制导与控制。

北京航空航天大学学报

北航学报 赠阅

A building method of high-precision landmark library for landmark autonomous navigation

YANG Bo*, ZHANG Dianlyu, ZHAO Xiaotao

(School of Astronautics, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: Satellite autonomous navigation based on landmark information can be used for all kinds of satellites which can periodically obtain images of earth surface for its high accuracy and independence. This method has been restricted by landmark selection method and the establishment strategy of landmark library, which affect the navigation precision and hinder the popularization and application of this navigation method. To solve these problems, a method for establishing global high-performance landmark library is proposed. The global landmark library could be automatically generated by using global landmark control areas to select high-quality global landmark samples. To verify the feasibility of the proposed method, a simulation program was designed. The results show that the position error of the landmark navigation system based on the new landmark library is about 99 m, and velocity error is about 0.08 m/s. The proposed method can provide high-precision autonomous navigation for spacecraft fast and accurately.

Keywords: satellite autonomous navigation; landmark navigation; landmark library; grayscale field characteristic parameter; maximum correlation coefficient (MCC) algorithm

北京航空航天大学学报

Received: 2017-01-16; **Accepted**: 2017-04-24; **Published online**: 2017-05-16 16:43

URL: kns.cnki.net/kcms/detail/11.2625.V.20170516.1643.002.html

* **Corresponding author**. E-mail: yangbo@buaa.edu.cn

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2016.0972

考虑单元界面关联特性的微定位平台刚度建模

全下文载

宫金良，郑洋洋，张彦斐*
（山东理工大学 机械工程学院，淄博 255049）

摘 要：设计了一种含折叠梁的并联微定位平台，具有大行程、低阻力的特点。采用传递矩阵法求解其刚度，建立各柔性子单元的传递矩阵，利用相邻单元公共结点实现传递性，通过力平衡方程、变形协调方程求解其末端位移与输入力之间的刚度矩阵，并提出了考虑全柔性的弹性折叠梁及弹性移动副刚度的求解方法。将传递矩阵法求解结果与有限元分析结果对比，误差在20.5%以内，在此基础上，考虑到模块化刚度分析方法将各子单元视为独立体，忽略各子单元之间的界面关联特性，提出了一种根据各子单元界面关联特性进行修正的方法，结果表明，该方法使其误差降低到10%以内，更好地满足了实际工程需求。

关 键 词：柔性机构；刚度；传递矩阵；界面关联特性；修正方法
中图分类号：TH112.5
文献标识码：A **文章编号**：1001-5965(2018)01-0010-08

对于柔性机构，目前应用最广泛的为微定位平台[1]，其结构形式有串联、并联、混联等，其中以并联微定位平台为主[2-3]，广泛应用于原子力显微镜、扫描隧道显微镜以及超精机床微进给机构、航空航天等领域[4-6]。

刚度影响机构的动态性能与精度，因此是评价并联微定位平台的重要指标。目前，并联微定位平台刚度的研究方法主要包括伪刚体模型法、柔度矩阵法、有限元法等。于靖军等[7]利用柔度矩阵法建立了三自由度柔性微定位平台刚度，考虑了柔性铰链及长杆变形，将短杆视为刚体，利用坐标变换、位移协调以及力平衡进行求解。Liu等[8]将宏动机构刚度分析理论应用于微动机构，分析了微动机构雅可比矩阵、刚度与机构参数关系以及运动空间等。孙立宁等[9]通过刚度组集以及协调方程建立大行程柔性铰链的6-PSS并联机构的整体系统刚度，并分析得到刚度影响图谱。李育文等[10]利用有限元软件分析微定位平台刚度，以此来检验理论分析的正确性。由以上研究可知，刚度建模过程中，求解繁琐，通用性不好，并且为方便计算，常将机构全部或部分视为刚体[11-12]，因此影响精度。

本文以带折叠梁的大行程并联微定位平台为例，将各单元视为柔性体，建立各单元传递矩阵，根据各相邻单元公共结点的关系建立末端位移与输入力之间的刚度矩阵，将其计算结果与ANSYS分析结果进行对比，分析其误差产生的原因并考虑各子单元之间的界面关联特性，对本文方法进行了修正，误差在10%以内，证明了方法的精确性。

1 传递矩阵法

如图1所示为任意梁单元受力图，由结点力及作用于 p 点的外载荷组成。定义状态向量如下：

$$\boldsymbol{S}=[f_x \quad f_y \quad T \quad u_x \quad u_y \quad \theta]^{\mathrm{T}} \tag{1}$$

收稿日期：2016-12-28；**录用日期**：2017-02-15；**网络出版时间**：2017-03-24 16:04
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170324.1604.003.html
基金项目：国家自然科学基金（61303006）；山东省优秀中青年科学家科研奖励基金（BS2012ZZ009）
*通信作者. E-mail：84374294@qq.com

引用格式：宫金良，郑洋洋，张彦斐. 考虑单元界面关联特性的微定位平台刚度建模[J]. 北京航空航天大学学报，2018，44(1)：10-17. GONG J L，ZHENG Y Y，ZHANG Y F. Stiffness modeling of micro-positioning platform considering unit interface relevance characteristics[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(1)：10-17(in Chinese).

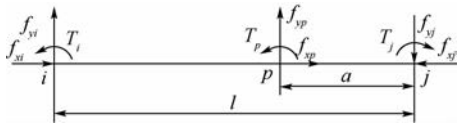


图 1　梁单元结点力及外载作用

Fig. 1　Node force and external load of beam unit

式中：f_x、f_y 和 T 分别为轴向力、剪力和弯矩；u_x、u_y 和 θ 分别为其所对应的轴向位移、横向位移和转角。

左结点 i 与右结点 j 的传递关系为

$$\boldsymbol{S}_j = \boldsymbol{T}\boldsymbol{S}_i + \boldsymbol{S}_p \tag{2}$$

式中：$\boldsymbol{T}_{(6\times6)}$ 为梁单元由左结点 i 到右结点 j 的传递矩阵；$\boldsymbol{S}_{p(6\times1)}$ 为外载荷的状态向量。$\boldsymbol{T}$ 及 $\boldsymbol{S}_p$ 由力平衡及虚功原理求得，文献[13-14]给出了求解过程及表达式。当无外载荷 $\boldsymbol{F}_p$ 时，$\boldsymbol{S}_p$ 为零。

为方便计算，将式(2)写为

$$\begin{bmatrix}\boldsymbol{F}_j\\ \boldsymbol{u}_j\end{bmatrix} = \begin{bmatrix}\boldsymbol{t}_1 & \boldsymbol{t}_2\\ \boldsymbol{t}_3 & \boldsymbol{t}_4\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_i\\ \boldsymbol{u}_i\end{bmatrix} + \begin{bmatrix}\boldsymbol{s}_{p1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{s}_{p2}\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_p\\ \boldsymbol{F}_p\end{bmatrix} \tag{3}$$

式中：

$$\boldsymbol{F}_t = [F_{xt} \quad F_{yt} \quad M_t]^{\mathrm{T}} \qquad t = i,j,p$$

$$\boldsymbol{u}_t = [u_{xt} \quad u_{yt} \quad \theta_t]^{\mathrm{T}} \qquad t = i,j,p$$

$$\boldsymbol{s}_{p1} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & -a & 1\end{bmatrix}$$

$$\boldsymbol{s}_{p2} = \begin{bmatrix}-\dfrac{a}{EA} & 0 & 0\\ 0 & \dfrac{a^3}{6EI} - \dfrac{a}{GA} & -\dfrac{a^2}{2EI}\\ 0 & \dfrac{a^2}{2EI} & -\dfrac{a}{2EI}\end{bmatrix}$$

其中：E 为弹性模量；A 为截面面积；G 为剪切模量；I 为惯性积。在局部坐标系下，传递矩阵传递方向为由左结点到右结点。当梁单元旋转角度 α 时，需要通过坐标变换得到参考坐标系下的传递矩阵。

$$\boldsymbol{T}_0 = \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{T}_0'\boldsymbol{\gamma} \tag{4}$$

式中：$\boldsymbol{T}_0$ 为参考坐标系下的传递矩阵；$\boldsymbol{T}_0'$ 为局部坐标系下的传递矩阵；$\boldsymbol{\gamma}_{(6\times6)} = \begin{bmatrix}\boldsymbol{R} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}\end{bmatrix}$，$\boldsymbol{R}$ 为旋转矩阵，即

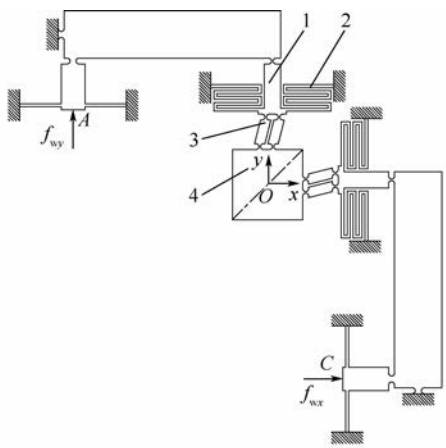
$$\boldsymbol{R} = \begin{bmatrix}\cos\alpha & \sin\alpha & 0\\ -\sin\alpha & \cos\alpha & 0\\ 0 & 0 & 1\end{bmatrix}$$

为了求解机构的刚度，需根据机构结构将其划分单元，考虑各单元变形求解传递矩阵，根据相邻 2 个单元共享的同一结点可将各单元组合。对于并联机构各支链，还需利用变形协调方程求解末端位移。由于本文中公式较多，规定如 $\boldsymbol{t}_1^{\mathrm{hgf}}$ 代表传递矩阵 $\boldsymbol{T}_{\mathrm{h}}$、$\boldsymbol{T}_{\mathrm{g}}$、$\boldsymbol{T}_{\mathrm{f}}$ 相乘后所得到的分块矩阵的第 1 项，以及各子单元的旋转变换不再赘述。

2　大行程并联微定位平台刚度

2.1　机构描述

本文设计了一个大行程并联微定位平台，如图 2 所示，该平台由 2 条支链构成，关于动平台对角线对称。由压电陶瓷分别驱动 A 点及 C 点。每条支链包括含折叠梁的杠杆放大机构及平行四杆机构，依据螺旋理论，每条支链中具有实际约束效果的部分只有折叠梁和平行四杆机构，其中杠杆放大机构的作用是增大折叠梁的实际驱动行程，因此基于其对应的伪刚体模型，动平台的主运动特征为沿 x 方向和 y 方向的平移。



1— 杠杆放大机构；2— 折叠梁；3— 平行四杆机构；4— 动平台。

图 2　大行程并联微定位平台

Fig. 2　Large stroke parallel micro-positioning platform

2.2　子单元刚度

根据结构将大行程并联微定位平台进行单元划分，其中包括直梁、柔性铰链、弹性移动副及弹性折叠梁，直梁及柔性铰链为基本的单元类型，其传递矩阵的表达式参见文献[14]。

2.2.1　弹性移动副

现有的弹性移动副刚度的求解将中间单元视为刚体，利用悬臂梁理论求解[15]。本文将弹性移动副视为全柔性求解，将其根据结构划分如图 3 所示的 4 个单元。

单元 a 为悬臂梁，因此：

$$\boldsymbol{F}_{1\mathrm{p}} = \boldsymbol{k}_{1\mathrm{p}}\boldsymbol{u}_{1\mathrm{p}} \tag{5}$$

式中：$\boldsymbol{k}_{1\mathrm{p}}$ 为悬臂梁的刚度。

单元 b 的传递矩阵关系如下：

北航学报 赠阅

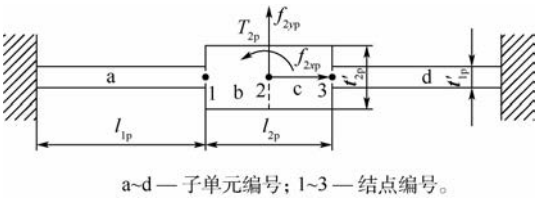


图 3 弹性移动副单元划分

Fig. 3 Compliant prismatic pair unit division

$$\begin{bmatrix} \boldsymbol{F}_{2\mathrm{bp}} \\ \boldsymbol{u}_{2\mathrm{bp}} \end{bmatrix} = \boldsymbol{T}_{\mathrm{bp}} \begin{bmatrix} \boldsymbol{F}_{1\mathrm{p}} \\ \boldsymbol{u}_{1\mathrm{p}} \end{bmatrix} \tag{6}$$

由式(5)、式(6)可以得到

$$\boldsymbol{F}_{2\mathrm{bp}} = \boldsymbol{t}_{1\mathrm{p}}^{\mathrm{b}} \boldsymbol{k}_{1\mathrm{p}} (\boldsymbol{t}_{3\mathrm{p}}^{\mathrm{b}} \boldsymbol{k}_{1\mathrm{p}} + \boldsymbol{t}_{4\mathrm{p}}^{\mathrm{b}})^{-1} \boldsymbol{u}_{2\mathrm{bp}} \tag{7}$$

单元 c 和 d 分别与单元 a 和 b 为对称关系，因此：

$$\boldsymbol{F}_{2\mathrm{cp}} = \boldsymbol{\lambda}^{-1} \boldsymbol{t}_{1\mathrm{p}}^{\mathrm{b}} \boldsymbol{k}_{1\mathrm{p}} (\boldsymbol{t}_{3\mathrm{p}}^{\mathrm{b}} \boldsymbol{k}_{1\mathrm{p}} + \boldsymbol{t}_{4\mathrm{p}}^{\mathrm{b}})^{-1} \boldsymbol{\lambda} \boldsymbol{u}_{2\mathrm{cp}} \tag{8}$$

式中：

$$\boldsymbol{\lambda} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

外力 $\boldsymbol{F}_{2\mathrm{p}}$ 作用于点 2，由力平衡和位移连续性可以得到各单元结点力与外力 $\boldsymbol{F}_{2\mathrm{p}}$ 的关系：

$$\begin{cases} \boldsymbol{F}_{2\mathrm{p}} = -\boldsymbol{F}_{2\mathrm{bp}} - \boldsymbol{F}_{2\mathrm{cp}} \\ \boldsymbol{u}_{2\mathrm{p}} = \boldsymbol{u}_{2\mathrm{bp}} = \boldsymbol{u}_{2\mathrm{cp}} \end{cases} \tag{9}$$

由式(7)～式(9)可以计算出弹性移动副刚度为

$$\boldsymbol{k}_{\mathrm{p}} = -\boldsymbol{t}_{1\mathrm{p}}^{\mathrm{b}} \boldsymbol{k}_{1\mathrm{p}} (\boldsymbol{t}_{3\mathrm{p}}^{\mathrm{b}} \boldsymbol{k}_{1\mathrm{p}} + \boldsymbol{t}_{4\mathrm{p}}^{\mathrm{b}})^{-1} - \boldsymbol{\lambda}^{-1} \boldsymbol{t}_{1\mathrm{p}}^{\mathrm{b}} \boldsymbol{k}_{1\mathrm{p}} (\boldsymbol{t}_{3\mathrm{p}}^{\mathrm{b}} \boldsymbol{k}_{1\mathrm{p}} + \boldsymbol{t}_{4\mathrm{p}}^{\mathrm{b}})^{-1} \boldsymbol{\lambda} \tag{10}$$

2.2.2 弹性折叠梁

弹性折叠梁的结构如图 4 所示，根据其结构形式将其划分为 A～D 共 4 个单元。

对于弹性折叠梁的单元 A，根据结构划分为 a～m 共 9 个单元，如图 5 所示。

单元 a～m 为串联结构，其各单元的传递矩阵如下：

$$\begin{bmatrix} \boldsymbol{F}_{iz} \\ \boldsymbol{u}_{iz} \end{bmatrix} = \boldsymbol{T}_{kz} \begin{bmatrix} \boldsymbol{F}_{jz} \\ \boldsymbol{u}_{jz} \end{bmatrix}$$

$$i = 2,3,\cdots,9;\ j = 1,2,\cdots,8;\ k = \mathrm{b},\mathrm{c},\cdots,\mathrm{h},\mathrm{m} \tag{11}$$

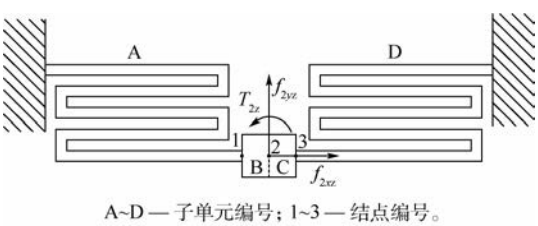


图 4 弹性折叠梁单元划分

Fig. 4 Compliant folded beam unit division

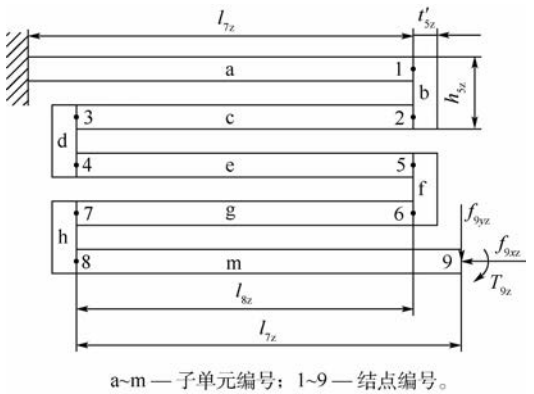


图 5 弹性折叠梁单元 A 划分

Fig. 5 Compliant folded beam unit A division

相邻单元共用的结点力为作用力与反作用力，位移连续。因此：

$$\begin{bmatrix} \boldsymbol{F}_{9z} \\ \boldsymbol{u}_{9z} \end{bmatrix} = \boldsymbol{T}_{\mathrm{mz}} \boldsymbol{T}_{\mathrm{hz}} \boldsymbol{T}_{\mathrm{gz}} \boldsymbol{T}_{\mathrm{fz}} \boldsymbol{T}_{\mathrm{ez}} \boldsymbol{T}_{\mathrm{dz}} \boldsymbol{T}_{\mathrm{cz}} \boldsymbol{T}_{\mathrm{bz}} \begin{bmatrix} \boldsymbol{F}_{1z} \\ \boldsymbol{u}_{1z} \end{bmatrix} \tag{12}$$

单元 a 为悬臂梁，因此：

$$\boldsymbol{F}_{1z} = \boldsymbol{k}_{1z} \boldsymbol{u}_{1z} \tag{13}$$

式中：$\boldsymbol{k}_{1z}$ 为悬臂梁的刚度。

联立式(11)～式(13)，弹性折叠梁中单元 A 的刚度为

$$\boldsymbol{k}_{\mathrm{A}} = \boldsymbol{t}_1^{\mathrm{A}} \boldsymbol{k}_{1z} (\boldsymbol{t}_3^{\mathrm{A}} \boldsymbol{k}_{1z} + \boldsymbol{t}_4^{\mathrm{A}})^{-1} \tag{14}$$

式中：$\begin{bmatrix} \boldsymbol{t}_1^{\mathrm{A}} & \boldsymbol{t}_2^{\mathrm{A}} \\ \boldsymbol{t}_3^{\mathrm{A}} & \boldsymbol{t}_4^{\mathrm{A}} \end{bmatrix} = \boldsymbol{T}_{\mathrm{mz}} \boldsymbol{T}_{\mathrm{hz}} \boldsymbol{T}_{\mathrm{gz}} \boldsymbol{T}_{\mathrm{fz}} \boldsymbol{T}_{\mathrm{ez}} \boldsymbol{T}_{\mathrm{dz}} \boldsymbol{T}_{\mathrm{cz}} \boldsymbol{T}_{\mathrm{bz}}$。

其余同弹性移动副刚度解法，因此弹性折叠梁的刚度为

$$\boldsymbol{k}_z = -\boldsymbol{t}_1^{\mathrm{B}} \boldsymbol{k}_{\mathrm{A}} (\boldsymbol{t}_3^{\mathrm{B}} \boldsymbol{k}_{\mathrm{A}} + \boldsymbol{t}_4^{\mathrm{B}})^{-1} - \boldsymbol{\lambda}^{-1} \boldsymbol{t}_1^{\mathrm{B}} \boldsymbol{k}_{\mathrm{A}} (\boldsymbol{t}_3^{\mathrm{B}} \boldsymbol{k}_{\mathrm{A}} + \boldsymbol{t}_4^{\mathrm{B}})^{-1} \boldsymbol{\lambda} \tag{15}$$

2.3 大行程并联微定位平台刚度求解

2.3.1 杠杆放大机构

根据结构将杠杆放大机构划分为如图 6 所示的 a～h 共 8 个单元。

$\boldsymbol{F}_{7\mathrm{p}'}$ 为平行四杆机构对其产生的作用力。结点 7 为杠杆放大机构、平行四杆机构以及折叠梁的结合处，因此在下文的求解过程中，将结点7 处的力与位移的命名与其单元相联系，如 $\boldsymbol{F}_{7\mathrm{h}}$ 表示

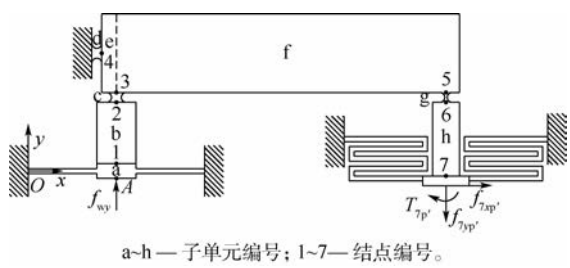


图 6 柔性杠杆放大机构单元划分

Fig. 6 Flexible lever magnifying mechanism unit division

在单元 h 上结点 7 的力。

各子单元传递矩阵关系如下:

$$\begin{bmatrix}\boldsymbol{F}_i\\ \boldsymbol{u}_i\end{bmatrix}=\boldsymbol{T}_k\begin{bmatrix}\boldsymbol{F}_j\\ \boldsymbol{u}_j\end{bmatrix}$$

$$i=2,3,6,7\mathrm{h};j=1,2,5,6;k=\mathrm{b},\mathrm{c},\mathrm{g},\mathrm{h} \tag{16}$$

根据传递矩阵传递性可得

$$\begin{cases}\begin{bmatrix}\boldsymbol{F}_3\\ \boldsymbol{u}_3\end{bmatrix}=\boldsymbol{T}_{\mathrm{c}}\boldsymbol{T}_{\mathrm{b}}\begin{bmatrix}\boldsymbol{F}_1\\ \boldsymbol{u}_1\end{bmatrix}\\ \begin{bmatrix}\boldsymbol{F}_{7\mathrm{h}}\\ \boldsymbol{u}_{7\mathrm{h}}\end{bmatrix}=\boldsymbol{T}_{\mathrm{h}}\boldsymbol{T}_{\mathrm{g}}\begin{bmatrix}\boldsymbol{F}_{5'}\\ u_{5'}\end{bmatrix}\end{cases} \tag{17}$$

单元 f 的受力分析如图 7 所示,结点 3 作用一外载荷 $\boldsymbol{F}_{3'}$。

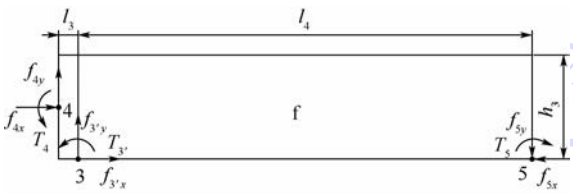


图 7　单元 f 受力分析

Fig. 7　Force analysis of unit f

因此,单元 f 的传递矩阵关系为

$$\begin{bmatrix}\boldsymbol{F}_5\\ \boldsymbol{u}_5\end{bmatrix}=\boldsymbol{T}_{\mathrm{f}}\begin{bmatrix}\boldsymbol{F}_{4'}\\ \boldsymbol{u}_{4'}\end{bmatrix}+\boldsymbol{S}_{\mathrm{f}}\begin{bmatrix}\boldsymbol{F}_{3'}\\ \boldsymbol{F}_{3'}\end{bmatrix} \tag{18}$$

单元 f 弯矩不可忽略。因此,单元 f 的传递矩阵中 $\boldsymbol{t}_1^{\mathrm{f}}$ 由力平衡可得

$$\boldsymbol{F}_5=\boldsymbol{t}_1^{\mathrm{f}}\boldsymbol{F}_4 \tag{19}$$

式中:

$$\boldsymbol{t}_1^{\mathrm{f}}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ -h_3/2 & -l_4-l_3 & 1\end{bmatrix}$$

单元 d 为一端固定的柔性铰链单元,因此:

$$\boldsymbol{F}_4=\boldsymbol{k}_{\mathrm{r4}}\boldsymbol{u}_4 \tag{20}$$

式中:$\boldsymbol{k}_{\mathrm{r4}}$为该单元在结点 4 处的刚度。

由于结点位移的连续性,结点 4 的位移可以通过结点 3 与结点 4 之间的直梁单元 e 由式(3)求得

$$\boldsymbol{u}_3=\boldsymbol{t}_3^{\mathrm{e}}\boldsymbol{F}_{4'}+\boldsymbol{t}_4^{\mathrm{e}}\boldsymbol{u}_{4'} \tag{21}$$

对于单元 a,由力平衡可以得到

$$\boldsymbol{F}_y=\boldsymbol{k}_{1\mathrm{p}}\boldsymbol{u}_1+\boldsymbol{F}_1 \tag{22}$$

式中:$\boldsymbol{F}_y$ 为压电陶瓷作用于单元 a 的外力,可以写成矩阵形式$[0\quad f_{\mathrm{w}y}\quad 0]^{\mathrm{T}}$;$\boldsymbol{k}_{1\mathrm{p}}$为弹性移动副单元 a 的刚度。

由式(17)~式(22)可以得到

$$\begin{cases}\boldsymbol{F}_{7\mathrm{h}}=\boldsymbol{f}_1\boldsymbol{F}_y+\boldsymbol{f}_2\boldsymbol{u}_1\\ \boldsymbol{u}_{7\mathrm{h}}=\boldsymbol{f}_3\boldsymbol{F}_y+\boldsymbol{f}_4\boldsymbol{u}_1\end{cases} \tag{23}$$

式中:

$$\boldsymbol{f}_1=\boldsymbol{t}_1^{\mathrm{hgf}}\boldsymbol{k}_{\mathrm{r4}}(\boldsymbol{t}_3^{\mathrm{e}}\boldsymbol{k}_{\mathrm{r4}}+\boldsymbol{t}_4^{\mathrm{e}})^{-1}\boldsymbol{t}_3^{\mathrm{cb}}+\boldsymbol{t}_1^{\mathrm{hg}}\boldsymbol{s}_1^{\mathrm{f}}\boldsymbol{t}_1^{\mathrm{cb}}$$

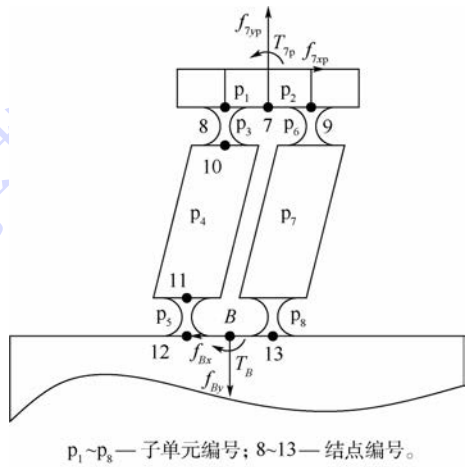
$$\boldsymbol{f}_2=-\boldsymbol{f}_1\boldsymbol{k}_{1\mathrm{p}}+\boldsymbol{t}_1^{\mathrm{hgf}}\boldsymbol{k}_{\mathrm{r4}}(\boldsymbol{t}_3^{\mathrm{e}}\boldsymbol{k}_{\mathrm{r4}}+\boldsymbol{t}_4^{\mathrm{e}})^{-1}\boldsymbol{t}_4^{\mathrm{cb}}$$

$$\boldsymbol{f}_3=(\boldsymbol{t}_3^{\mathrm{hgf}}\boldsymbol{k}_{\mathrm{r4}}+\boldsymbol{t}_4^{\mathrm{hgf}})(\boldsymbol{t}_3^{\mathrm{e}}\boldsymbol{k}_{\mathrm{r4}}+\boldsymbol{t}_4^{\mathrm{e}})^{-1}\boldsymbol{t}_3^{\mathrm{cb}}+(\boldsymbol{t}_3^{\mathrm{hg}}\boldsymbol{s}_1^{\mathrm{f}}+\boldsymbol{t}_4^{\mathrm{hg}}\boldsymbol{s}_2^{\mathrm{f}})\boldsymbol{t}_1^{\mathrm{cb}}$$

$$\boldsymbol{f}_4=-\boldsymbol{f}_3\boldsymbol{k}_{1\mathrm{p}}+(\boldsymbol{t}_3^{\mathrm{hgf}}\boldsymbol{k}_{\mathrm{r4}}+\boldsymbol{t}_4^{\mathrm{hgf}})(\boldsymbol{t}_3^{\mathrm{e}}\boldsymbol{k}_{\mathrm{r4}}+\boldsymbol{t}_4^{\mathrm{e}})^{-1}\boldsymbol{t}_4^{\mathrm{cb}}$$

2.3.2　平行四杆机构

平行四杆机构结构如图 8 所示,根据其结构特点将其划分为 p_1 ~ p_8 共 8 个单元。



p_1~p_8— 子单元编号; 8~13— 结点编号。

图 8　柔性平行四杆机构单元划分

Fig. 8　Flexible parallel four-bar mechanism unit division

平行四杆机构中各子单元传递矩阵关系如下:

$$\begin{bmatrix}\boldsymbol{F}_i\\ \boldsymbol{u}_i\end{bmatrix}=\boldsymbol{T}_k\begin{bmatrix}\boldsymbol{F}_j\\ \boldsymbol{u}_j\end{bmatrix}$$

$$i=8,7\mathrm{p}_2,10,11,12;j=7\mathrm{p}_1,9,8,10,11;$$
$$k=\mathrm{p}_1,\mathrm{p}_2,\mathrm{p}_3,\mathrm{p}_4,\mathrm{p}_5 \tag{24}$$

对于支链 $\mathrm{p}_1/\mathrm{p}_3/\mathrm{p}_4/\mathrm{p}_5$,可以得到以下关系:

$$\begin{bmatrix}\boldsymbol{F}_{12}\\ \boldsymbol{u}_{12}\end{bmatrix}=\boldsymbol{T}_{\mathrm{m}}\begin{bmatrix}\boldsymbol{F}_{7\mathrm{p}_1}\\ \boldsymbol{u}_{7\mathrm{p}_1}\end{bmatrix} \tag{25}$$

式中:$\boldsymbol{T}_{\mathrm{m}}=\boldsymbol{T}_{\mathrm{p}_5}\boldsymbol{T}_{\mathrm{p}_4}\boldsymbol{T}_{\mathrm{p}_3}\boldsymbol{T}_{\mathrm{p}_1}$。

支链 $\mathrm{p}_2/\mathrm{p}_6/\mathrm{p}_7/\mathrm{p}_8$ 与支链 $\mathrm{p}_1/\mathrm{p}_3/\mathrm{p}_4/\mathrm{p}_5$ 相似,因此:

$$\begin{bmatrix}\boldsymbol{F}_{13}\\ \boldsymbol{u}_{13}\end{bmatrix}=\boldsymbol{T}_{\mathrm{n}}\begin{bmatrix}\boldsymbol{F}_{7\mathrm{p}_2}\\ \boldsymbol{u}_{7\mathrm{p}_2}\end{bmatrix} \tag{26}$$

式中:$\boldsymbol{T}_{\mathrm{n}}=\boldsymbol{T}_{\mathrm{p}_8}\boldsymbol{T}_{\mathrm{p}_7}\boldsymbol{T}_{\mathrm{p}_6}\boldsymbol{T}_{\mathrm{p}_2}$。

杠杆放大机构对平行四杆机构产生作用力 $\boldsymbol{F}_{7\mathrm{p}}$,对于结点 7,由力平衡可得

$$\boldsymbol{F}_{7\mathrm{p}}=\boldsymbol{F}_{7\mathrm{p}_1}+\boldsymbol{F}_{7\mathrm{p}_2} \tag{27}$$

由于动平台相比于其他单元结构尺寸大,可将其视为刚体,则 B 点力与位移平衡协调。

$$\boldsymbol{F}_B=\boldsymbol{D}_1\boldsymbol{F}_{12}+\boldsymbol{D}_2\boldsymbol{F}_{13} \tag{28}$$

$$\boldsymbol{u}_B=\frac{1}{2}(\boldsymbol{u}_{12}+\boldsymbol{u}_{13}) \tag{29}$$

北航学报 赠阅

式中：

$$\boldsymbol{D}_1=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & -l_{11}/2 & 1\end{bmatrix}$$

$$\boldsymbol{D}_2=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & l_{11}/2 & 1\end{bmatrix}$$

当动平台为刚体时，与平台相连接的柔性铰链转动角度相等，即 $\theta_{12}=\theta_{13}$。由于其旋转角度微小，因此：

$$\begin{cases}\sin\theta_i=\dfrac{u_{13y}-u_{12y}}{l_{11}}=\theta_i\\ \cos\theta_i=\dfrac{u_{13x}+l-u_{12x}}{l_{11}}=1\end{cases}\quad i=12,13 \tag{30}$$

由式(30)得到结点 13 与结点 12 位移之间的关系：

$$\boldsymbol{u}_{13}=\boldsymbol{D}_3\boldsymbol{u}_{12} \tag{31}$$

式中：

$$\boldsymbol{D}_3=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & l_{11}\\ 0 & 0 & 1\end{bmatrix}$$

由式(25)～式(29)、式(31)得

$$\begin{bmatrix}\boldsymbol{F}_B\\ \boldsymbol{u}_B\end{bmatrix}=\begin{bmatrix}\boldsymbol{t}_1^{\mathrm{p}} & \boldsymbol{t}_2^{\mathrm{p}}\\ \boldsymbol{t}_3^{\mathrm{p}} & \boldsymbol{t}_4^{\mathrm{p}}\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_{7\mathrm{p}}\\ \boldsymbol{u}_{7\mathrm{p}}\end{bmatrix} \tag{32}$$

式中：$\boldsymbol{t}_1^{\mathrm{p}}=\boldsymbol{f}_5\boldsymbol{f}_7^{-1}$；$\boldsymbol{t}_2^{\mathrm{p}}=\boldsymbol{f}_5\boldsymbol{f}_7^{-1}\boldsymbol{f}_8-\boldsymbol{f}_6$；$\boldsymbol{t}_3^{\mathrm{p}}=\boldsymbol{f}_7^{-1}$；$\boldsymbol{t}_4^{\mathrm{p}}=\boldsymbol{f}_7^{-1}\boldsymbol{f}_8$；$\boldsymbol{f}_5=2\boldsymbol{D}_1\boldsymbol{t}_1^{\mathrm{m}}(\boldsymbol{t}_3^{\mathrm{m}})^{-1}(\boldsymbol{D}_3+\boldsymbol{E})^{-1}+2\boldsymbol{D}_2\boldsymbol{t}_1^{\mathrm{n}}(\boldsymbol{t}_3^{\mathrm{n}})^{-1}\cdot(\boldsymbol{D}_3^{-1}+\boldsymbol{E})^{-1}$；$\boldsymbol{f}_6=\boldsymbol{D}_1\boldsymbol{t}_1^{\mathrm{m}}(\boldsymbol{t}_3^{\mathrm{m}})^{-1}\boldsymbol{t}_4^{\mathrm{m}}+\boldsymbol{D}_2\boldsymbol{t}_1^{\mathrm{n}}(\boldsymbol{t}_3^{\mathrm{n}})^{-1}\boldsymbol{t}_4^{\mathrm{n}}$；$\boldsymbol{f}_7=2(\boldsymbol{t}_3^{\mathrm{m}})^{-1}(\boldsymbol{D}_3+\boldsymbol{E})^{-1}+2(\boldsymbol{t}_3^{\mathrm{n}})^{-1}(\boldsymbol{D}_3^{-1}+\boldsymbol{E})^{-1}$；$\boldsymbol{f}_8=(\boldsymbol{t}_3^{\mathrm{m}})^{-1}\boldsymbol{t}_4^{\mathrm{m}}+(\boldsymbol{t}_3^{\mathrm{n}})^{-1}\boldsymbol{t}_4^{\mathrm{n}}$。

2.3.3 计算整体刚度

结点 7 同时受到弹性折叠梁、杠杆放大机构、平行四杆机构的作用力，因此由力平衡可以得到

$$\boldsymbol{k}_z\boldsymbol{u}_7=\boldsymbol{F}_{7\mathrm{p}}-\boldsymbol{F}_{7\mathrm{h}} \tag{33}$$

由式(23)、式(32)、式(33)得

$$\boldsymbol{F}_B=\boldsymbol{k}_1\boldsymbol{u}_B+\boldsymbol{k}_2\boldsymbol{F}_y \tag{34}$$

式中：$\boldsymbol{k}_1=\boldsymbol{f}_{12}\boldsymbol{f}_{10}^{-1}$；$\boldsymbol{k}_2=\boldsymbol{f}_{11}-\boldsymbol{f}_{12}\boldsymbol{f}_{10}^{-1}\boldsymbol{f}_9$；$\boldsymbol{f}_9=(\boldsymbol{t}_3^{\mathrm{p}}\boldsymbol{k}_z+\boldsymbol{t}_4^{\mathrm{p}})\cdot\boldsymbol{f}_3+\boldsymbol{t}_3^{\mathrm{p}}\boldsymbol{f}_1$；$\boldsymbol{f}_{10}=(\boldsymbol{t}_3^{\mathrm{p}}\boldsymbol{k}_z+\boldsymbol{t}_4^{\mathrm{p}})\boldsymbol{f}_4+\boldsymbol{t}_3^{\mathrm{p}}\boldsymbol{f}_2$；$\boldsymbol{f}_{11}=(\boldsymbol{t}_1^{\mathrm{p}}\boldsymbol{k}_z+\boldsymbol{t}_2^{\mathrm{p}})\cdot\boldsymbol{f}_3+\boldsymbol{t}_1^{\mathrm{p}}\boldsymbol{f}_1$；$\boldsymbol{f}_{12}=(\boldsymbol{t}_1^{\mathrm{p}}\boldsymbol{k}_z+\boldsymbol{t}_2^{\mathrm{p}})\boldsymbol{f}_4+\boldsymbol{t}_1^{\mathrm{p}}\boldsymbol{f}_2$。

水平支链与竖直支链关于动平台对角线对称，因此进行坐标变换即可得到竖直支链末端力。

$$\boldsymbol{\beta}\boldsymbol{F}_D=\boldsymbol{k}_1\boldsymbol{\beta}\boldsymbol{u}_D+\boldsymbol{k}_2\boldsymbol{\beta}\boldsymbol{F}_x \tag{35}$$

式中：$\boldsymbol{F}_x$ 为$[f_{wx}\quad 0\quad 0]^{\mathrm{T}}$；$\boldsymbol{\beta}$ 为坐标变换矩阵，即

$$\boldsymbol{\beta}=\begin{bmatrix}0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & -1\end{bmatrix}$$

2 条支链对动平台作用力及运动示意图如图 9所示，由动平台力平衡得

$$\boldsymbol{F}_B+\boldsymbol{D}_4\boldsymbol{F}_D=\boldsymbol{0} \tag{36}$$

式中：

$$\boldsymbol{D}_4=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ l_{\mathrm{d}}/2 & l_{\mathrm{d}}/2 & 1\end{bmatrix}$$

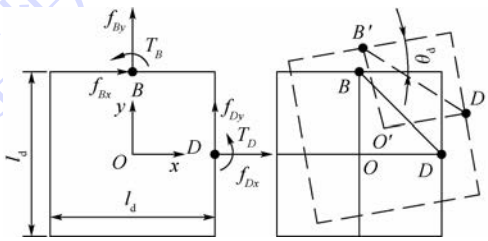


图 9 动平台作用力及运动示意图

Fig. 9 Moving platform force and motion diagram

由图 9 可以计算出动平台旋转角度 θ_{d} 时：

$$\begin{cases}\boldsymbol{u}_O=\boldsymbol{u}_D+[\,l_{\mathrm{d}}(1-\cos\theta_{\mathrm{d}})/2\quad -l_{\mathrm{d}}\sin\theta_{\mathrm{d}}/2\quad 0\,]^{\mathrm{T}}\\ \boldsymbol{u}_O=\boldsymbol{u}_B+[\,l_{\mathrm{d}}\sin\theta_{\mathrm{d}}/2\quad l_{\mathrm{d}}(1-\cos\theta_{\mathrm{d}})/2\quad 0\,]^{\mathrm{T}}\end{cases} \tag{37}$$

θ_{d} 值接近零，因此 $\sin\theta_{\mathrm{d}}\approx\theta_{\mathrm{d}}$；$\cos\theta_{\mathrm{d}}\approx 1$。式(37)可简化为

$$\begin{cases}\boldsymbol{u}_O=\boldsymbol{D}_5\boldsymbol{u}_D\\ \boldsymbol{u}_O=\boldsymbol{D}_6\boldsymbol{u}_B\end{cases} \tag{38}$$

式中：

$$\boldsymbol{D}_5=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & -l_{\mathrm{d}}/2\\ 0 & 0 & 1\end{bmatrix}$$

$$\boldsymbol{D}_6=\begin{bmatrix}1 & 0 & l_{\mathrm{d}}/2\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

为方便计算，将平台所受到的 x 方向与 y 方向输入力转换为外力矢 $\boldsymbol{F}=[F_x\quad F_y\quad 0]^{\mathrm{T}}$。

$$\begin{cases}\boldsymbol{F}_{\mathrm{dx}}=\boldsymbol{D}_7\boldsymbol{F}\\ \boldsymbol{F}_{\mathrm{dy}}=\boldsymbol{D}_8\boldsymbol{F}\end{cases} \tag{39}$$

式中：

$$\boldsymbol{D}_7=\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 1\end{bmatrix}$$

$$\boldsymbol{D}_8=\begin{bmatrix}0 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

联合式(34)～式(39)，即可得到输入力与输

北航学报 赠阅

出位移之间的关系：

$$\boldsymbol{F}=\boldsymbol{K}\boldsymbol{u}_O \tag{40}$$

式中：

$$\boldsymbol{K}=-(\boldsymbol{D}_4\boldsymbol{\beta}^{-1}\boldsymbol{k}_2\boldsymbol{\beta}\boldsymbol{D}_7+\boldsymbol{k}_2\boldsymbol{D}_8)^{-1}\cdot(\boldsymbol{D}_4\boldsymbol{\beta}^{-1}\boldsymbol{k}_1\boldsymbol{\beta}\boldsymbol{D}_5^{-1}+\boldsymbol{k}_1\boldsymbol{D}_6^{-1})$$

3　算例分析

3.1　传递矩阵法与有限元法结果对比

分析比较本文方法与有限元分析之间的误差。运用 ANSYS14.0 建立大行程并联微定位平台模型进行刚度研究，单元类型选择三维 20 结点实体单元 Solid95，设定柔性杠杆放大机构的材料为 60Si2Mn，基本材料参数为：$E=206$ GPa；$\mu=0.27$；$G=79$ GPa。

大行程并联微定位平台模型参数如图 10 所示，取其结构参数为：厚度 $b=8$ mm，$l_1=14$ mm，$t_1'=1$ mm，$l_2=8$ mm，$h_1=15.5$ mm，$R_1=1$ mm，$h_2=13.5$ mm，$t_2'=2$ mm，$R_2=1$ mm，$t_3'=2$ mm，$l_3=3$ mm，$l_4=67.25$ mm，$l_5=73$ mm，$h_3=16$ mm，$R_3=1$ mm，$t_4'=0.5$ mm，$l_6=5.5$ mm，$h_4=15$ mm，$l_7=16$ mm，$l_8=14$ mm，$t_5'=1$ mm，$h_5=3$ mm，$l_9=9.5$ mm，$h_6=2$ mm，$t_6'=0.5$ mm，$l_{10}=8.25$ mm，$h_7=3.4$ mm，$R_4=1$ mm，$l_{11}=4.5$ mm，$\theta=76°$，$l_d=23.75$ mm。

图 11 为大行程并联微定位平台有限元模型变形云图。其分析过程为：分别约束 2 条支链的弹性移动副两端、弹性折叠梁两端，以及柔性铰链连接底座的一端。如图 2 所示，在 C 点处施加 f_{wx}，在 A 点处施加 f_{wy}。在分析过程中，令 $f_{wx}=100$ N，f_{wy} 为变量。求解后提取动平台中心点的位移量将其与本文求解结果对比，如表 1 所示。

由表 1 中修正前误差数据可以看出，传递矩阵法与有限元分析值误差在20.5% 以内，其误差产生的原因为：传递矩阵法基于梁理论，在计算过程中是建立在一定假设条件下，如忽略轴向变形等；将微定位平台划分子单元求解，忽略了单元之间的联系及单元之间的接触变形。

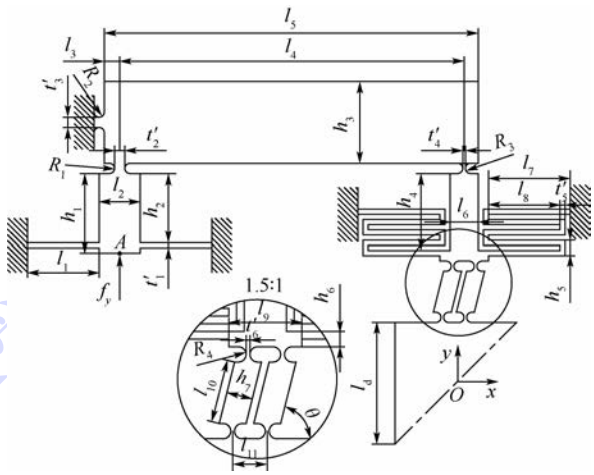


图 10　大行程并联微定位平台参数模型

Fig. 10　Large stroke parallel micro-positioning platform parameter model

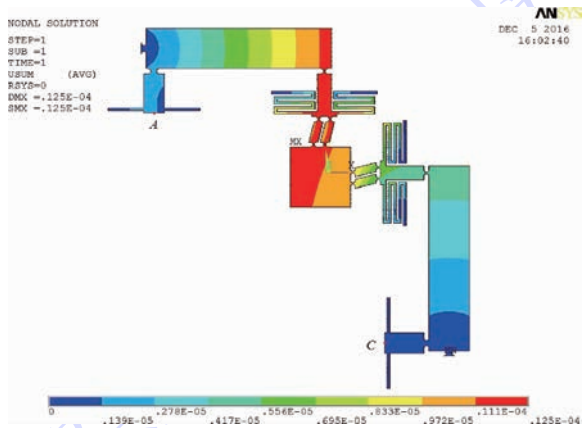


图 11　大行程并联微定位平台有限元模型

Fig. 11　Finite element model of large stroke parallel micro-positioning platform

表 1　传递矩阵法与有限元分析值对比

Table 1　Comparison between transfer matrix method and finite element analysis

f_{wy}/N	0	50	100	150	200
u_x 有限元法/μm	−2.03063	1.55645	5.14353	8.73061	12.31770
u_x 修正前/μm	−2.09430	1.30869	4.71168	8.11466	9.81616
u_x 修正后/μm	−2.01307	1.41088	4.83482	8.25877	11.68270
u_x 修正前误差/%	3.135	15.918	8.396	7.055	20.309
u_x 修正后误差/%	0.865	9.353	6.002	5.404	5.155
u_y 有限元法/μm	7.18439	6.17025	5.15611	4.14196	3.12782
u_y 修正前/μm	6.80597	5.75883	4.71168	4.18810	3.14065
u_y 修正后/μm	6.84790	5.84136	4.83482	3.82829	2.82175
u_y 修正前误差/%	5.267	6.668	8.619	1.114	0.410
u_y 修正后误差/%	4.684	5.330	6.231	7.573	9.785

北航学报 赠阅

3.2 考虑界面关联特性的修正方法

根据传递矩阵法误差产生的原因提出考虑单元之间联系性的修正方法。以单元 a 与单元 b 求解过程为例。单元 a、b 结构图如图 12 所示。

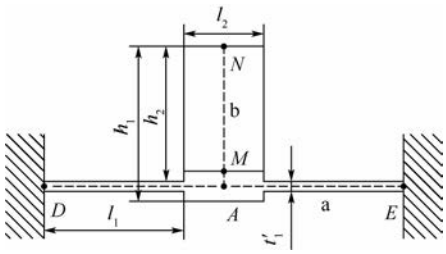


图 12 单元 a、b 结构图

Fig. 12 Unit a, b structure chart

单元 a、b 在求解过程中将其简化为梁单元，单元 a 为变截面横向梁单元 DE，最终求解的为结点 A 的刚度。单元 b 为竖向梁单元 MN，其 2 个结点之间的传递矩阵关系如式(41)所示。在以上求解过程中，视结点 M 与结点 A 力为作用力与反作用力位移连续。而实际中结点 M 与结点 A 不重合，忽略了结点 M 与结点 A 之间的关系。而结点 M 与结点 A 之间的关系可由竖向梁单元 MA 求解(见式(42))。

$$\begin{bmatrix} \boldsymbol{F}_N \\ \boldsymbol{u}_N \end{bmatrix} = \boldsymbol{T}_{NM} \begin{bmatrix} \boldsymbol{F}_M \\ \boldsymbol{u}_M \end{bmatrix} \tag{41}$$

$$\begin{bmatrix} \boldsymbol{F}_M \\ \boldsymbol{u}_M \end{bmatrix} = \boldsymbol{T}_{MA} \begin{bmatrix} \boldsymbol{F}_A \\ \boldsymbol{u}_A \end{bmatrix} \tag{42}$$

因此，结点 N 与结点 A 之间的传递矩阵关系为

$$\begin{bmatrix} \boldsymbol{F}_N \\ \boldsymbol{u}_N \end{bmatrix} = \boldsymbol{T}_{NM}\boldsymbol{T}_{MA} \begin{bmatrix} \boldsymbol{F}_A \\ \boldsymbol{u}_A \end{bmatrix} = \boldsymbol{T}_{NA} \begin{bmatrix} \boldsymbol{F}_A \\ \boldsymbol{u}_A \end{bmatrix} \tag{43}$$

所以为求解精确，在求解单元 b 传递矩阵时，单元 b 的长度由 MN 修正为 NA。

利用此方法考虑各相邻单元之间的联系对大行程并联微定位平台弹性折叠梁及单元 h 的长度进行修正。其计算结果如表 1 中修正后误差数据所示，传递矩阵法修正后的结果与有限元分析值误差在 10% 以内，精度得到提高，证明了此修正方法的有效性。

4 结 论

1）本文设计了一种压电陶瓷驱动的大行程并联微定位平台，采用含折叠梁的杠杆放大机构，提高其位移输出倍率，减少传动链的阻力，实现大行程位移输出。

2）利用传递矩阵法求解并联微定位平台，将平台按结构划分单元，分别求解各单元传递矩阵，根据各相邻单元公共结点实现其传递性；再由变形协调方程、力平衡方程求解整体刚度；并求解了弹性移动副及弹性折叠梁考虑全柔性的刚度公式。

3）将本文求解结果与有限元分析结果对比，误差在 20.5% 以内，分析其误差产生的原因并提出了考虑各子单元界面关联特性的修正方法，修正后误差降低到 10% 以内，证明了修正方法的有效性和传递矩阵的准确性。

参考文献（References）

[1] 于靖军，郝广波，陈贵敏，等. 柔性机构及其应用研究进[J]. 机械工程学报，2015，51(13)：53-68.
YU J J，HAO G B，CHEN G M，et al. State-of-art of compliant mechanisms and their applications[J]. Chinese Journal of Mechanical Engineering，2015，51(13)：53-68(in Chinese).

[2] 杨启志，郭宗和，马履中，等. 全柔性机构在并联微动机器人中的应用[J]. 机械设计与研究，2005，21(5)：45-48.
YANG Q Z，GUO Z H，MA L Z，et al. Application of the fullly compliant mechanism in the parallel micro-motion robots[J]. Machine Design and Research，2005，21(5)：45-48(in Chinese).

[3] HAO G B，KONG X W. A novel large-range XY compliant parallel manipulator with enhanced out-of-plane stiffness[J]. Journal of Mechanical Design，2012，134(6)：061009.

[4] 艾青林，黄伟锋，张洪涛. 并联机器人刚度与静力学研究现状与进展[J]. 力学进展，2012，42(5)：583-592.
AI Q L，HUANG W F，ZHANG H T. Review of stiffness and statics analysis of parallel robot[J]. Advances in Mechanics，2012，42(5)：583-592(in Chinese).

[5] STANFORD B，BERAN P. Conceptual design of compliant mechanisms for flapping wing with topology optimization[J]. AIAA Journal，2011，49(4)：855-867.

[6] WISSA A，TUMMALA Y，HUBBARD J E，et al. Passively morphing ornithopter wings using a novel compliant spine：Design and testing[J]. Smart Materials and Structures，2012，21(9)：094028.

[7] 于靖军，毕树生，宗光华. 空间全柔性机构位置分析的刚度矩阵法[J]. 北京航空航天大学学报，2002，28(3)：323-326.
YU J J，BI S S，ZONG G H. Stiffness matrix method for displacement analysis of fully spatial compliant mechanisms[J]. Journal of Beijing University of Aeronautics and Astronautics，2002，28(3)：323-326(in Chinese).

[8] LIU X J，WANG J，GAO F，et al. On the design of 6-DOF parallel micro-motion manipulators[C]// Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway：IEEE Press，2001：343-348.

[9] 孙立宁，董为，杜志江. 基于大行程柔性铰链的并联机器人刚度分析[J]. 机械工程学报，2005，41(8)：90-95.
SUN L N，DONG W，DU Z J. Stiffness analysis on a wide-range flexure hinge-based parallel manipulator[J]. Chinese Journal of Mechanical Engineering，2005，41(8)：90-95(in Chinese).

北航学报 赠阅

[10] 李育文，张华，杨建新，等. 6-UPS 并联机床静刚度的有限元分析和实验研究[J]. 中国机械工程，2004，15(2)：112-115.
LI Y W，ZHANG H，YANG J X，et al. Finite element analysis and experimental study for the stiffness of a 6-UPS parallel kinematics machine[J]. China Mechanical Engineering，2004，15(2)：112-115(in Chinese).

[11] YU Y Q，FENG Z L，XU Q P. A pseudo-rigid-body 2R model of flexural beam in compliant mechanisms[J]. Mechanism and Machine Theory，2012，55(9)：19-33.

[12] 邱丽芳，霍明磊，李威. 六杆柔顺机构的伪刚体模型[J]. 北京科技大学学报，2013，35(5)：682-686.
QIU L F，HUO M L，LI W. Pseudo-rigid-body model of a six-bar full-compliant mechanism[J]. Journal of University of Science and Technology Beijing，2013，35(5)：682-686(in Chinese).

[13] 李青宁. 变截面杆元传递矩阵法[J]. 西安建筑科技大学学报，2001，33(1)：18-23.
LI Q N. The transfer matrix method of bar elements with variable cross-section[J]. Journal of Xi'an University of Architecture & Technology，2001，33(1)：18-23(in Chinese).

[14] 郑洋洋，宫金良，张彦斐. 基于传递矩阵法的柔性杠杆放大机构刚度分析[J]. 北京航空航天大学学报，2017，43(4)：849-856.
ZHENG Y Y，GONG J L，ZHANG Y F. Rigidity analysis of a flexible lever magnifying mechanism based on transfer matrix method[J]. Journal of Beijing University of Aeronautics and Astronautics，2017，43(4)：849-856(in Chinese).

[15] 杨春辉. 平行板型柔性移动副的刚度计算及分析[J]. 现代制造工程，2013(12)：30-33.
YANG C H. The stiffness design calculation and analysis of parallel plate flexible prismatic pair[J]. Modern Manufacturing Engineering，2013(12)：30-33(in Chinese).

作者简介：

宫金良　男，博士，副教授。主要研究方向：并联机器人分析与设计理论。

郑洋洋　女，硕士研究生。主要研究方向：并联机器人分析与设计理论。

张彦斐　女，博士，副教授。主要研究方向：并联机器人分析与设计理论。

Stiffness modeling of micro-positioning platform considering unit interface relevance characteristics

GONG Jinliang，ZHENG Yangyang，ZHANG Yanfei*

(School of Mechanical Engineering，Shandong University of Technology，Zibo 255049，China)

Abstract：A parallel micro-positioning platform with folded beams was designed，which has characteristics of large stroke and low resistance. Stiffness was solved by transfer matrix method. First，transfer matrix of flexible subunit was established. The transitivity characteristics became available by taking advantage of the common node belonging to adjacent elements. Finally，the stiffness matrix between input force and output displacement of flexible mechanism was solved according to the force balance equation and compatibility equation of deformation. A method to solve the stiffness of compliant folded beam and compliant prismatic pair considering full flexibility was put forward. The result of deformation error is less than 20.5% compared with finite element analysis. On this basis，this method ignored the correlation between the subunits because the modular stiffness analysis method regarded each subunit as independent. A modified method considering each subunit dependency was put forward，which reduced the error to less than 10% and made the results better meet the actual engineering needs.

Keywords：flexible mechanism；stiffness；transfer matrix；interface relevance characteristics；revision method

Received：2016-12-28；**Accepted**：2017-02-15；**Published online**：2017-03-24 16：04

URL：kns.cnki.net/kcms/detail/11.2625.V.20170324.1604.003.html

Foundation items：National Natural Science Foundation of China (61303006)；the Research Award Fund for Outstanding Young Scholars of Shandong Province (BS2012ZZ009)

* **Corresponding author**. E-mail：84374294@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0010

基于 TVP-VAR 模型的多模式交通需求耦合分析

马晓磊，孙硕，丁川*，王云鹏

（北京航空航天大学 交通科学与工程学院，北京 100083）

摘 要：城市交通出行选择多模式条件下多种出行方式之间的出行需求关系密切，为了探讨城市居民对地铁、公交车、私家车3种出行方式出行需求之间的耦合关系及时变特征，采用时变参数向量自回归(TVP-VAR)模型，使用道路拥堵指数近似代表私家车出行量，对2个小区一个月内工作日时间地铁出行量、公交车出行量和道路拥堵指数三者的平均值进行分析。研究结果表明：地铁出行量、公交车出行量、私家车出行量三者相互作用关系针对不同的用地类型，整体影响趋势不会变化太大，影响大小在时间上存在差异；城市交通小区内居民对地铁、公交车、私家车3种出行方式的出行需求之间全天存在相互作用关系，地铁出行需求的增加会减少公交车出行需求，公交车出行需求的增加会增加地铁出行需求，而私家车出行需求的增加会增加地铁出行需求并减少公交车出行需求。研究结果有助于深入认识现阶段中国城市多模式交通出行方式之间的耦合关系，以更好地应对城市交通拥堵问题。

关 键 词：出行需求；多模式出行方式；时变分析；耦合影响；TVP-VAR 模型
中图分类号：U121
文献标识码：A **文章编号**：1001-5965(2018)01-0018-09

随着国家经济的发展和城市人口的增长，2015 年中国汽车保有量达 1.72 亿辆，其中私家车 1.24 亿，占比达 72.1%。与此同时，居民的出行方式日趋多样化，公交、地铁已经成为许多大、中城市运载出行者的主要工具。然而许多城市仍面临严重的交通拥堵问题，事实证明，随着私家车数量的增加，原有的拓宽道路和增设路线已经不能适应城市的发展，城市多模式出行交通网络需要深入研究。

近年来，国内外许多学者通过搭建模型，制定了大量的城市多模式出行交通网络下的最优平衡配流模型，同时证明了时变的交通需求分析要比时不变的更有效[1-9]。Larry 和 Keyvan[3] 通过对比分析了不同算法在交通需求的弹性分配问题方面的应用，结果发现埃文斯的模型在收敛性和鲁棒性方面更优。王力等[5]提出了一种新的模糊动态交通分配(FDTA)模型，描述了动态旅行时间并找出了模糊最短路径集合，最后将 FDTA 模型与经典随机动态交通分配模型比较，证明在正常情况下 FDTA 模型更加符合实际交通情况。四兵锋等[6]通过分析城市多模式出行交通系统的结构特征，提出了城市多模式出行交通网络平衡配流的变分不等式模型。Zhou 等[10]通过分析指出动态时变的交通配流更利于解决复杂拥挤的多模式城市交通网络，同时也可以更好地捕捉居民出行信息。然而这些研究的结果多是基于严格的假设或者仿真，没有深入探究多模式出行交通网络下不同出行方式出行需求之间的相互作用关系，因此在实际情况下可能会存在些许偏差。李靖睿[11]从经济发展的角度分析了公交车与私家车

收稿日期：2017-01-12；**录用日期**：2017-03-06；**网络出版时间**：2017-04-14 16:54
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170414.1654.001.html
基金项目：国家自然科学基金（51408019，71503018，U1564212）
* **通信作者**. E-mail：cding@buaa.edu.cn

引用格式：*马晓磊，孙硕，丁川，等. 基于TVP-VAR 模型的多模式交通需求耦合分析[J]. 北京航空航天大学学报，2018，44(1)：18-26. MA X L，SUN S，DING C，et al. A coupling analysis of multimodal transportation demands based on TVP-VAR models[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(1)：18-26 (in Chinese).*

北航学报 赠阅

之间的相互作用关系。郝忠娜和何玉宏[12]探究了地铁与常规公交之间的竞争、制约以及资源浪费等问题。罗焕然[13]从地铁定价的角度分析了出行者对于私家车与地铁两者之间的需求作用关系。然而这些探究普遍存在以下几点问题:①交通网络是整体,研究中只考虑了 2 种出行方式出行需求之间的相关性;②城市交通不同出行方式的交通需求是全天时变的,研究中忽略了出行需求的时变特征;③基于交通调查数据或是纯粹理论分析,没有利用稳定可靠的数据分析问题。

因此总体来看,从时变角度出发,城市多模式交通网络中多种出行方式之间的时变相互作用关系有待深入研究。本文尝试运用时变计量经济分析方法,针对城市多模式交通网络中不同出行方式出行需求的动态相互作用关系展开研究,并对变量之间的时变作用关系大小量化描述,分析时变影响程度。本文研究成果有助于深化认识城市多模式交通网络中不同出行方式出行需求之间的相互作用关系,为政策制定者对交通网络提供多方面信息。

1　研究方法

1.1　时变向量自回归模型

为了解决大型联立方程模型的弊端,Sims[14]提出了向量自回归(VAR)模型,模型中不再区分变量类型以及每个方程的具体含义,VAR 模型现已成为分析宏观实证研究的主流方法。然而传统的 VAR 模型系数和方差-协方差矩阵是不变的,变量之间相关性是按照固定的规律分析,变量之间的传导机制是固定的,因此不利于分析复杂系统中变量之间的相关性。国外学者基于 VAR 模型拓展到结构向量自回归(SVAR)模型。近年来,Primiceri[15]基于 SVAR 模型提出了带随机波动的时变参数向量自回归(TVP-VAR)模型,允许了系数和方差-协方差矩阵的变化,因此可以更好地捕捉复杂系统中多变量之间的时变相关性。基于 Primiceri[15]模型,Nakajima[16]提出了对 TVP-VAR 模型的估计算法。

SVAR 模型标准形式为

$$\boldsymbol{A}\boldsymbol{y}_t=\boldsymbol{B}_0+\boldsymbol{B}_1\boldsymbol{y}_{t-1}+\boldsymbol{B}_2\boldsymbol{y}_{t-2}+\cdots+\boldsymbol{B}_p\boldsymbol{y}_{t-p}+\boldsymbol{u}_t \quad (1)$$

式中:$t=p+1,\cdots,n$,p 表示滞后阶数;$\boldsymbol{y}_t,\boldsymbol{y}_{t-1},\cdots,\boldsymbol{y}_{t-p}$为 $m\times1$ 阶的 m 个变量的向量;$\boldsymbol{A},\boldsymbol{B}_0,\boldsymbol{B}_1,\cdots,\boldsymbol{B}_p$ 为 $m\times m$ 的时变系数矩阵,$\boldsymbol{A}$ 代表同期相关性,$\boldsymbol{B}_0,\boldsymbol{B}_1,\cdots,\boldsymbol{B}_p$ 代表跨时期的相关性;$\boldsymbol{u}_t$ 为 t 时刻的扰动项,同时假定 $\boldsymbol{u}_t\sim N(0,\boldsymbol{\Sigma}_u\boldsymbol{\Sigma}_u')$,$\boldsymbol{\Sigma}_u$ 为对角矩阵。

假定 $\boldsymbol{A}$ 为下三角矩阵,这样可以显著地减少运算的复杂性。

$$\boldsymbol{A}=\begin{bmatrix}1&0&\cdots&0\\a_{2,1}&\ddots&\ddots&\vdots\\\vdots&\ddots&\ddots&0\\a_{n,1}&\cdots&a_{n,n-1}&1\end{bmatrix}$$

式(1)左右同乘 $\boldsymbol{A}^{-1}$,则得到

$$\begin{aligned}\boldsymbol{y}_t=&\boldsymbol{F}_0+\boldsymbol{F}_1\boldsymbol{y}_{t-1}+\boldsymbol{F}_2\boldsymbol{y}_{t-2}+\cdots+\\&\boldsymbol{F}_p\boldsymbol{y}_{t-p}+\boldsymbol{A}^{-1}\boldsymbol{\Sigma}_u\boldsymbol{\Sigma}_u'\boldsymbol{\varepsilon}_t\\&\boldsymbol{\varepsilon}_t\sim N(0,\boldsymbol{I}_n)\end{aligned} \quad (2)$$

$$\boldsymbol{\Sigma}_u=\begin{bmatrix}\sigma_{u1}&0&\cdots&0\\0&\sigma_{u2}&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&\sigma_{un}\end{bmatrix}$$

式中:$\boldsymbol{F}_i=\boldsymbol{A}^{-1}\boldsymbol{B}_i$,$i=0,1,\cdots,p$;$\boldsymbol{I}_n$为 n 阶单位矩阵。

定义 $\boldsymbol{X}_t=\boldsymbol{I}_n\otimes(1,\boldsymbol{y}_{t-1}',\cdots,\boldsymbol{y}_{t-p}')$,$\boldsymbol{\beta}=(\boldsymbol{F}_0,\boldsymbol{F}_1,\cdots,\boldsymbol{F}_p)'$,$\otimes$表示克罗内克积,因此式(2)可简化为

$$\boldsymbol{y}_t=\boldsymbol{X}_t\boldsymbol{\beta}+\boldsymbol{A}^{-1}\boldsymbol{\Sigma}_u\boldsymbol{\Sigma}_u'\boldsymbol{\varepsilon}_t \quad (3)$$

基于式(3),允许所有参数变化,因此:

$$\boldsymbol{y}_t=\boldsymbol{X}_t\boldsymbol{\beta}_t+\boldsymbol{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\Sigma}_t'\boldsymbol{\varepsilon}_t \quad (4)$$

式中:$\boldsymbol{\beta}_t$、$\boldsymbol{A}_t$、$\boldsymbol{\Sigma}_t$ 都是时变的。将 $\boldsymbol{A}_t$ 中主对角线及以下的 $n\times(n-1)/2$ 个元素排列成向量 $\boldsymbol{a}_t$,即 $\boldsymbol{a}_t=(a_{21,t},a_{31,t},a_{32,t},\cdots,a_{n(n-1),t})'$,假定 $\boldsymbol{h}_{jt}=\ln\sigma_{jt}^2$,$j=1,2,\cdots,n$,$t=p+1,p+2,\cdots,n$,即 $\boldsymbol{h}_t=(\ln\sigma_{1t}^2,\ln\sigma_{2t}^2,\cdots,\ln\sigma_{nt}^2)$,定义 $\boldsymbol{\Sigma}_t$ 中的元素为向量 $\boldsymbol{\sigma}_t$。基于式(4),假设 $\boldsymbol{\beta}_t$、$\boldsymbol{a}_t$、$\boldsymbol{h}_t$ 中的元素为一阶随机游走过程,即 $\boldsymbol{\beta}_{t+1}=\boldsymbol{\beta}_t+\boldsymbol{v}_{t+1}$,$\boldsymbol{a}_{t+1}=\boldsymbol{a}_t+\boldsymbol{\varsigma}_{t+1}$,$\boldsymbol{h}_{t+1}=\boldsymbol{h}_t+\boldsymbol{\xi}_{t+1}$,$\boldsymbol{v}_{t+1}$、$\boldsymbol{\varsigma}_{t+1}$、$\boldsymbol{\xi}_{t+1}$表示 $\boldsymbol{\beta}_t$、$\boldsymbol{a}_t$、$\boldsymbol{h}_t$ 的扰动项,$\boldsymbol{\beta}_{p+1}\sim N(\boldsymbol{v}_0,\boldsymbol{\Omega}_\beta)$,$\boldsymbol{a}_{p+1}\sim N(\boldsymbol{\varsigma}_0,\boldsymbol{\Omega}_a)$,$\boldsymbol{h}_{p+1}\sim N(\boldsymbol{\xi}_0,\boldsymbol{\Omega}_h)$,$\boldsymbol{\Omega}_\beta$、$\boldsymbol{\Omega}_a$、$\boldsymbol{\Omega}_h$ 分别表示 $\boldsymbol{\beta}_{p+1}$、$\boldsymbol{a}_{p+1}$、$\boldsymbol{h}_{p+1}$的方差-协方差矩阵。以上演算过程假设参数服从一阶随机游走,这与实际情况可能会有所差别,但由于模型的估计计算是在一段时间内进行的,因此结果不会产生大的偏差,并且 Primiceri[15]将此估计下的结果与更普通的 VR 过程对比,证明这个问题影响不大,而且结果更有利于捕捉参数变化,反映交通出行变化的细微特征,同时极大程度地降低了计算难度。同时为了方便计算,假定不同参数之间不相关,则合并的正态分布 $\boldsymbol{V}$ 如下:

$$\boldsymbol{V}=\mathrm{Var}\left(\begin{bmatrix}\boldsymbol{\varepsilon}_t\\\boldsymbol{v}_t\\\boldsymbol{\varsigma}_t\\\boldsymbol{\xi}_t\end{bmatrix}\right)=\begin{bmatrix}\boldsymbol{I}_n&0&0&0\\0&\boldsymbol{\Omega}_\beta&0&0\\0&0&\boldsymbol{\Omega}_a&0\\0&0&0&\boldsymbol{\Omega}_h\end{bmatrix}$$

对于参数估计部分,由于参数太多,因此需要先设定初值,即使用最小二乘算法先验训练,得出初值: $\boldsymbol{\beta}_0 \sim N(\boldsymbol{\beta}_{\text{ols}}, 4V(\boldsymbol{\beta}_{\text{ols}}))$, $\boldsymbol{a}_0 \sim N(\boldsymbol{a}_{\text{ols}}, 4V(\boldsymbol{a}_{\text{ols}}))$, $\boldsymbol{h}_0 \sim N(\boldsymbol{h}_{\text{ols}}, 4\boldsymbol{I}_n)$, $\boldsymbol{\beta}_{\text{ols}}$、$\boldsymbol{a}_{\text{ols}}$、$\boldsymbol{h}_{\text{ols}}$ 分别为 $\boldsymbol{\beta}_0$、$\boldsymbol{a}_0$、$\boldsymbol{h}_0$ 通过最小二乘算法估计得到的初值, $V(\boldsymbol{\beta}_{\text{ols}})$、$V(\boldsymbol{a}_{\text{ols}})$、$\boldsymbol{I}_n$ 分别为 $\boldsymbol{\beta}_0$、$\boldsymbol{a}_0$、$\boldsymbol{h}_0$ 通过最小二乘算法估计得到的方差。将方差部分乘以 4,是为了表示更大的不确定性,防止先验信息约束过紧,再根据蒙特卡罗马尔可夫链(MCMC)算法估计参数值,对于得出的结果,需要使用脉冲响应量化表示。

1.2 脉冲响应

由于模型中同期以及跨期相关性的存在,任何一个解释变量的变化都是触发其他解释变量的变化。在 VAR 模型中,脉冲响应表示,当解释变量的扰动项增加一个单位或一个标准差而其他变量扰动项保持不变的情况下,所有变量的当期和未来的变化即为脉冲响应。脉冲响应很好地解决了存在同期相关和变量间滞后动态影响在解释上的问题,即可以了解到公交车出行量、地铁出行量和私家车出行量之间影响程度的持续时间和大小。

2 实证分析

2.1 数据选取及描述

根据北京市实际情况,居民主流出行方式主要有公交车、地铁、私家车、骑行(自行车、电动车和摩托车等)、步行,而相对而言,骑行和步行的出行数据难以全天完整地采集,因此本文选择北京市公交车、地铁和私家车 3 种出行方式的出行量进行分析。

针对地铁和公交车的出行需求,需要借助城市居民的出行量表示,而城市公交 IC 卡很好地记录了城市居民的出行信息,本文即采用公交 IC 卡数据统计城市地铁和公交的上车量。

而对于私家车,出行量难以确定,然而路网中私家车的路权使用占比达 90%[17],再根据高德手机的私家车驾车导航数据可以计算小区道路拥堵程度,进而在一定程度上可以使用道路拥挤程度代替私家车出行量的变化趋势。对于道路拥堵程度,需基于 Taylor 和 Richardson 等[18-19]提出的道路拥堵指数定义一条道路的拥堵程度: $\text{CI} = (T - T_0)/T_0$。其中,CI 为道路拥堵指数, T 为车辆通过该道路的实际行驶时间, T_0 为道路通畅时车辆通过该道路所需的时间。假设路网有 n 条道路,根据道路 i 的车辆数占区域车辆总数的比例设定该道路对于整体路网的权重 ω_i,整合即得出整体区域的道路拥堵指数: $S = \sum_{i=1}^{n} (\text{CI} \times \omega_i)$, $i = 1, 2, \cdots, n$,即代表私家车出行量。

交通小区作为交通规划的基础单位,北京市交通委员会综合考虑城市规划的各个分区以及人口、就业等信息,将北京市划分为 1 911 个交通小区。本文针对北京市交通小区,选取 2015 年 6 月 1 日到 2015 年 6 月 30 日之间工作日的地铁进站 IC 卡信息和公交上车 IC 卡信息,并对结果取平均值。需要注意的是,本文是针对该交通小区平均数据一段时间内的不同出行方式之间相关性的研究,因此结果可以仅表征该月小区的整体状况,针对不同情况和需求,可以将数据的时间段随意调整。对于城市道路信息,则使用北京市交通小区内高德手机的用户 GPS 信息和路网实际道路长度信息,得到交通小区内用户通过该路段的使用时间、路段长度和路段上的用户数,根据道路拥堵指数公式即得出道路拥堵指数。

根据实际情况,选取北京 2 个不同土地利用类型的交通小区:小区 a 和小区 b,如图 1 和图 2 中黑色线框内的区域。其中,小区 a 位于三里屯,用地类型为商务办公场所,小区b内部是芳群



图 1 小区 a 地理信息图

Fig. 1 Geographic information map of Zone a



图 2 小区 b 地理信息图

Fig. 2 Geographic information map of Zone b

公寓群,用地类型是居民住宅区。2 个小区的道路拥堵指数、地铁出行量、公交车出行量时间变化趋势折线图如图 3 和图 4 所示。

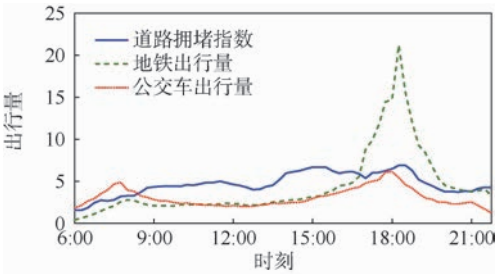


图 3　小区 a 三变量的变化趋势

Fig. 3　Variation tendency of three variables for Zone a

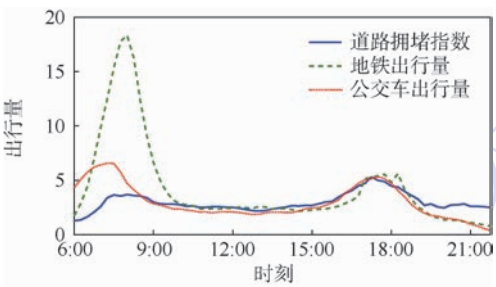


图 4　小区 b 三变量的变化趋势

Fig. 4　Variation tendency of three variables for Zone b

2.2　数据平稳性及协整关系检验

在建立 TVP-VAR 模型之前,需要确定道路拥堵指数、地铁出行量、公交车出行量的数据平稳性和变量之间的长期相关性。

2.2.1　平稳性检验

采用单位根检验法中的增广迪基-福勒检验法(ADF)对道路拥堵指数、地铁出行量、公交车出行量 3 个时间序列进行平稳性检验,结果如表 1和表 2 所示。

结果表明,在1% 的显著水平下,小区a和小区 b 原始水平的地铁出行量、公交车出行量和道路拥堵指数的 ADF 检验值均大于临界值,表明 3 个时间序列都是不平稳的,同时 3 组序列的一阶差分序列也是如此。而小区 a 和小区 b 3 组时间序列的二阶差分均小于 1% 显著水平下临界值,表明其二阶差分序列为平稳序列。因此,小区 a 和小区 b 的地铁出行量、公交车出行量和道路拥堵指数都是二阶单整序列,可以进行协整关系检验。

表 1　小区 a 的单位根检验

Table 1　Unit root test of Zone a

差分阶数	1% 显著水平下的临界值	ADF 检验值		
		地铁出行量	公交车出行量	道路拥堵指数
原始水平	-3.540 198	-1.744 661	-1.922 983	-2.712 431
一阶差分	-3.544 063	-3.468 943	-2.889 990	-3.305 034
二阶差分	-3.546 099	-6.302 398	-5.330 916	-6.655 348

表 2　小区 b 的单位根检验

Table 2　Unit root test of Zone b

差分阶数	1% 显著水平下的临界值	ADF 检验值		
		地铁出行量	公交车出行量	道路拥堵指数
原始水平	-3.541 098	-3.300 278	-3.174 048	-3.350 398
一阶差分	-3.544 210	-3.075 129	-2.977 441	-2.704 181
二阶差分	-3.542 079	-6.862 239	-6.738 689	-13.140 510

2.2.2　协整关系检验

协整关系检验的目的是探究变量之间是否存在长期稳定的相关性,结果如表 3 和表 4 所示。迹统计量和最大特征值统计量均在至多 2 个条件下小于 5% 显著水平下的临界值。因此,表明小区 a 和小区 b 的地铁出行量、公交车出行量和道路拥堵指数之间存在协整关系,即长期稳定的相关性。

表 3　小区 a 的协整关系检验

Table 3　Co-integration test of Zone a

协整关系方程个数	特征根	迹统计量			最大特征值统计量		
		检验值	5% 显著水平下的临界值	P 值	检验值	5% 显著水平下的临界值	P 值
无	0.320 949	34.334 680	29.797 070	0.014 0	24.384 740	21.131 620	0.016 8
至多 1 个	0.107 859	9.949 939	15.494 710	0.284 6	7.190 264	14.264 600	0.466 9
至多 2 个	0.042 859	2.759 675	3.841 466	0.096 7	2.759 675	3.841 466	0.096 7

表 4　小区 b 的协整关系检验

Table 4　Co-integration test of Zone b

协整关系方程个数	特征根	迹统计量			最大特征值统计量		
		检验值	5% 显著水平下的临界值	P 值	检验值	5% 显著水平下的临界值	P 值
无	0.775 583	117.285 700	29.797 070	0.000 0	94.137 620	21.131 620	0.000 0
至多 1 个	0.285 834	23.148 040	15.494 710	0.002 9	21.208 330	14.264 600	0.003 4
至多 2 个	0.030 320	1.939 713	3.841 466	0.163 7	1.939 713	3.841 466	0.163 7

北航学报 赠阅

2.3 结果与分析

针对第 1 节的算法,将地铁出行量、公交车出行量、道路拥堵指数作为 3×1 的向量 $\boldsymbol{y}$ 代入 TVP-VAR 模型中,通过赤池信息准则(AIC)确定滞后阶数 $p=1$,对于脉冲响应过程中,由于数据以15 min为单位时间间隔,确定了滞后阶数为 1 后,即选择在 15、30、60 和 90 min 滞后时间点分别计算脉冲响应,不同滞后时间反映变量间的相互作用关系在该滞后时间点的大小,选择多个滞后时间也是为了最大程度地捕捉出行方式之间的影响。先在时不变的模型中进行 3 000 次最小二乘迭代作为先验抽样,再基于先验的基础进行 30 000 万次蒙特卡罗马尔可夫链迭代,最终得出结果。

2.3.1 小 区 a

以每 15 min 为 1 个单位,图 5 展示了小区 a 在 15、30、60、90 min 滞后情况下地铁出行量、公交车出行量和私家车出行量三者之间相互的时变脉冲响应。图中,横轴表示时间标识,纵轴表示脉冲响应大小,脉冲响应越大,则表示影响越显著。以地铁对公交车(见图 5(b))为例,地铁对公交车表示地铁出行量的增加对公交车出行量的影响。值得注意的是:①小区 a 是工作地,除了私家车出行量对自身的影响较为平稳,其余各子图都会在 17:30 时存在一个凸型峰值;②地铁出行量对私家车出行量影响(见图 5(c))和私家车出行量对公交车出行量影响(见图 5(h))的峰值在曲线中更接近 0 的点,表明在晚高峰时间,地铁出行量的增加不会对私家车出行量造成大的影响,同时由于公交车出行量的饱和,私家车出行量的增加也不会对公交车出行量造成大的影响。而变量间的其他相互关系则均在晚高峰时作用更明显,对于非高峰期,各出行方式之间的相互作用关系较为平稳。

地铁出行量和公交车出行量对私家车的影响(见图 5(c)、(f))很小,表明车主的出行不会受到公交车和地铁出行量的影响,主要原因是公交车和地铁班次数量是全天固定的,不会因为乘客出行量的增加而增加班次,因此道路的拥堵程度不会产生变化,即私家车出行量不会受到大的影响。对于地铁和公交车,地铁对公交车的脉冲响应(见图 5(b))是负的,表明地铁出行量的增加会减少公交车出行量,但是影响程度较弱,而公交车对地铁是正的(见图 5(d)),说明公交车出行量的增加会增加地铁出行量,即公交车和地铁两者的相互作用关系不是对称的。从而可以看出,选择地铁的乘客选择稳定性更强,同时当路网拥堵时,乘客更愿意选择地铁,相应的,如果地铁也拥堵,公交车的出行量也会减少。对于私家车对地铁和公交车的影响(见图 5(g)、(h)),私家车出行量的增加,即路网越拥堵,则地铁出行量会越多,相应的公交车出行量会越少,这与上面地铁和公交车的关系也是相对应的,即表明路网越拥堵,地铁出行量会越多,而公交车出行量会越少。

2.3.2 小 区 b

图 6 展示了小区 b 在 15、30、60、90 min 滞后情况下地铁出行量、公交车出行量和私家车出行量三者之间相互的时变脉冲响应。不同于小区 a,小区 b 是居民住宅区,因此该小区地铁出行量、公交车出行量和私家车出行量之间的相互作用关系大小在 6:30 存在峰值,针对此峰值,地铁出行量对公交车出行量的影响(见图 6(b))大小为凹型峰值,这是因为如果地铁出行量增加了,则表明更多的人选择了地铁,选择公交车出行的人即会变少,其余的峰值均为凸型。同样的,地铁出行量对私家车出行量影响(见图 6(c))和私家车出行量对公交车出行量影响(见图 6(h))的峰值为曲线中更接近 0 的点,说明在该小区,早高峰时间地铁出行量的增加不会对私家车出行量造成大的影响,私家车出行量的增加也不会对公交车出行量造成大的影响,同时,小区 b 在 16:00 地铁出行量对公交车出行量、私家车出行量的影响和公交车出行量对私家车出行量的影响(见图 6(b)、(c)、(f))存在凸型峰值,并且该峰值更接近 0,表明在该时间点地铁出行量对公交车出行量和私家车出行量的影响较小,公交车出行量对私家车出行量的影响也较弱,而非峰值时段,不同出行方式之间的相互作用关系较为平稳。

与小区 a 相似,在小区 b 中地铁对公交车的脉冲响应(见图 6(b))是负的,公交车对地铁的脉冲响应(见图 6(d))正的,地铁出行量的增加对降低公交车出行量的影响较弱,而公交车出行量的增加会极大地增加地铁出行量,同样说明了选择地铁出行的选择稳定性要更强,而选择公交车出行的会受拥挤程度影响较大。公交车对私家车的脉冲响应(见图 6(f))是负的,即表明公交车出行量的增加会减少车主选择私家车出行,因为小区 b 是住宅区,说明针对不同的区域,车主针对开车出行的态度会有所不同。私家车出行量对地铁出行量和公交车出行量的影响(见图 6(g)、(h))也和小区 a 相似,私家车出行量的增加会增加地铁出行量而减少公交车出行量,而在时间变化趋势上差异较大,小区a的影响程度会在下午

北航学报 赠阅

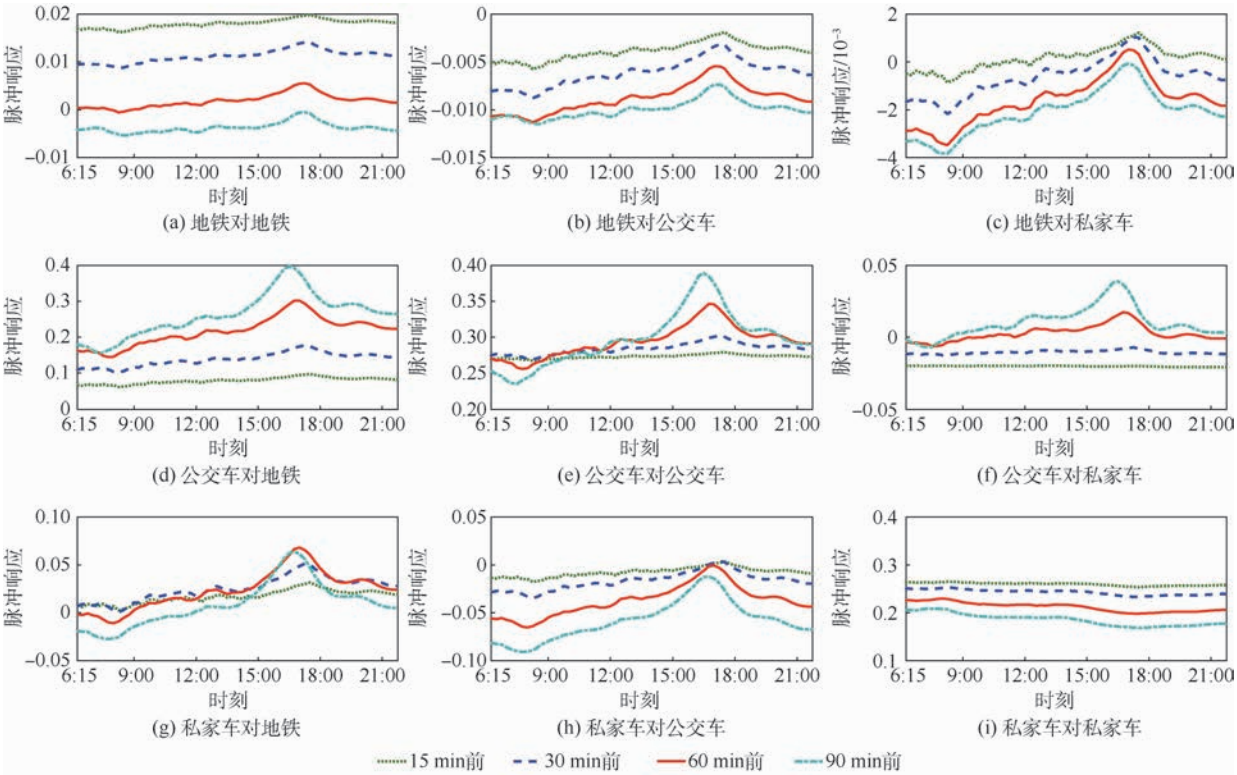


图 5　小区 a 不同滞后时间下的时变脉冲响应

Fig. 5　Time-varying impulse response of Zone a for different lag time

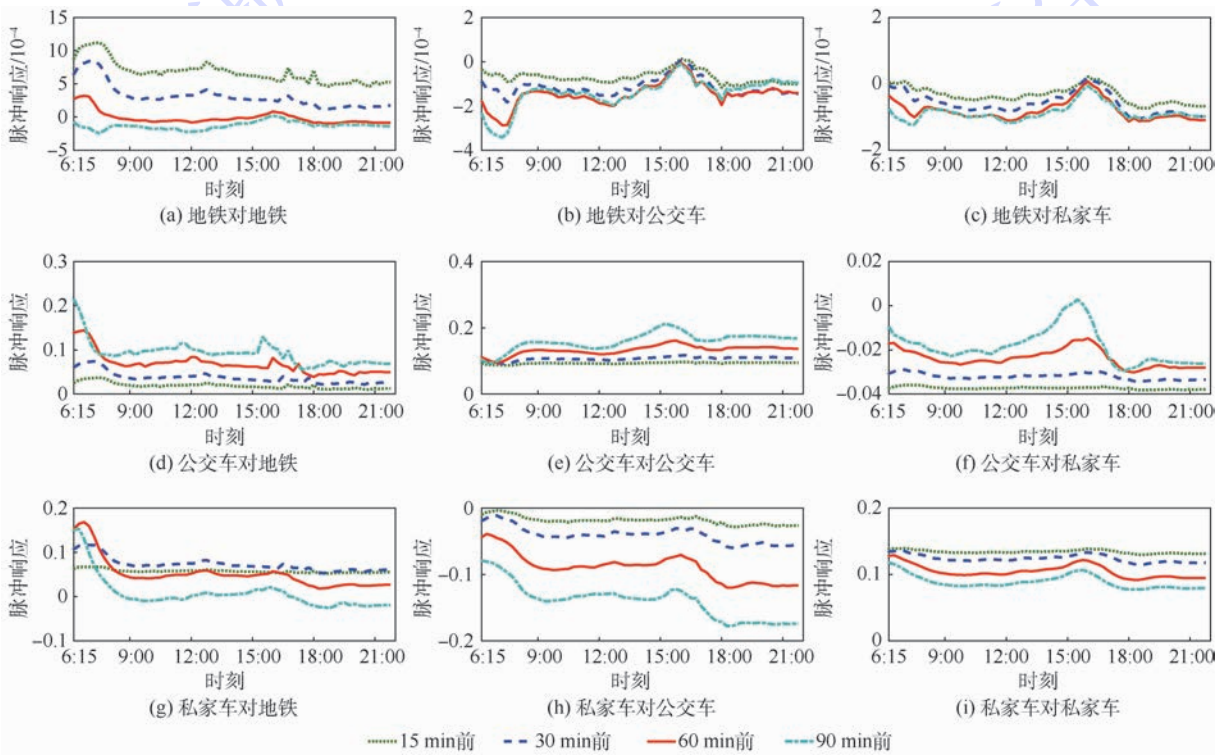


图 6　小区 b 不同滞后时间下的时变脉冲响应

Fig. 6　Time-varying impulse response of Zone b for different lag time

达到峰值,而小区 b 会在早晨达到峰值,即说明对于早高峰拥堵情况的区域,居民出行更倾向于选择地铁,而拥堵的道路环境会减少居民选择公交车出行。

北航学报 赠阅

3 结 论

本文通过 TVP-VAR 模型探究城市交通小区地铁、公交车和私家车 3 种出行方式之间的时变相互作用关系,能更好地认识了中国现阶段多模式出行方式的时变相互作用关系,为政策制定者提供深层次的城市道路相互影响信息,以针对不同时间点做出相应的对策,研究对于城市建设和解决道路拥堵问题具有正向意义。具体地,本文根据北京市 2 个交通小区一个月工作日的全天分时段平均地铁出行量、公交车出行量和道路拥堵指数的时间序列数据,利用 TVP-VAR 模型分析三者之间的时变相互作用关系,结论表明,交通小区内地铁出行量、公交车出行量和私家车出行量三者之间存在一种长期的相互作用关系,具体结论如下:

1) 针对不同的用地类型,地铁出行量、公交车出行量和私家车出行量三者之间相互作用关系的大小整体不会有所不同,但是在时间演变趋势上会有差异。本文研究选择的小区 a 和小区 b 中,乘客对于地铁、公交车和私家车的态度几乎一致,区别在于:小区 a 中在 17:30 达到三者相互作用大小的高峰,小区 b 在 6:30 达到三者相互作用大小的高峰,同时在高峰时间点,地铁对私家车的影响和私家车对公交车的影响会减弱,而非高峰时段各出行方式间的相互作用关系则较为平稳。

2) 针对本文中 2 个小区,地铁出行量和公交车出行量之间的相互作用关系是非对称的,针对不同的用地类型,地铁出行量的增加会微弱地减少选择公交车出行的人数,而公交车出行量的增加会极大地增加选择地铁出行的人数。即居民选择地铁出行的稳定性要更大,而选择公交车出行的更容易受到影响而选择其他出行方式,这可能与公交车和地铁的行驶方式乘客等待环境相关,地铁因在地下行驶,到站准确率更高,而公交车则会受路网拥堵情况影响,时间准确性上会有所偏差,同时地铁是进站刷卡,然后再等待车辆到来,即乘客不会再因拥堵而选择其他出行方式,而公交车是上车刷卡,即乘客会根据等待的人数和道路环境而确定是否要选择公交车出行。

3) 针对本文中 2 个小区,地铁出行量和公交车出行量的大小变化不会影响私家车出行量。地铁和公交车出行量的增加不会导致道路班次的增加,因而私家车车主不会感受到这 2 个变量的影响,只会受到私家车出行量自身的影响。而私家车出行量的增加会增加地铁出行量并减少公交车出行量,即路网拥堵的情况下,出行者倾向于选择地铁,而选择公交车出行的乘客也会相应的选择其他出行方式,地铁的到站准确率为其吸引了大量的乘客,而选择公交车出行的乘客受到环境的影响选择其他方式的可能性会更大。

此外,本文也有一些局限性。本文只考虑了 3 种出行方式并使用道路拥堵指数代替私家车出行量,计算过程中没有考虑城市道路中的其他出行方式,虽然其他出行方式不会对本文中出行方式两两分析造成很大的影响,但结果会缺少因考虑其他出行方式而发生的变化,这与实际情况有一定不符。在今后的研究中也会努力完善。

参考文献 (References)

[1] 郑雪琳,干宏程. 居民交通方式选择行为影响因素分析[J]. 上海理工大学学报,2013,35(6):563-566.
ZHENG X L,GAN H C. Analysis on influencial factors of travel mode choice behavior of residents[J]. Journal of University of Shanghai for Science and Technology,2013,35(6):563-566(in Chinese).

[2] 熊丽音,陆锋,陈传彬. 城市多模式交通网络特征连通关系表达模型[J]. 武汉大学学报(信息科学版),2008,33(4):393-396.
XIONG L Y,LU F,CHEN C B. Connectivity model for multi-model urban transportation networks[J]. Geomatics & Information Science of Wuhan University,2008,33(4):393-396(in Chinese).

[3] LARRY J,KEYVAN F. Efficient algorithms for solving elastic demand traffic assignment problems and mode split-assignment problems[J]. Transportation Science,1981,15(4):306-317.

[4] 韩印,袁鹏程. 多用户多方式混合随机交通平衡分配模型[J]. 交通运输工程学报,2008,8(1):97-101.
HAN Y,YUAN P C. Multi-user and multi-mode assignment model of mixed stochastic traffic balance[J]. Journal of Traffic & Transportation Engineering,2008,8(1):97-101(in Chinese).

[5] 王力,王川久,沈晓蓉,等. 基于模糊旅行时间的动态交通分配模型[J]. 北京航空航天大学学报,2005,31(10):1149-1152.
WANG L,WANG C J,SHEN X R,et al. Dynamic traffic assignment model based on fuzzy travel time[J]. Journal of Beijing University of Aeronautics and Astronautics,2005,31(10):1149-1152(in Chinese).

[6] 四兵锋,杨小宝,高亮,等. 基于出行需求的城市多模式交通配流模型[J]. 中国公路学报,2010,23(6):85-91.
SI B F,YANG X B,GAO L,et al. Urban multimodal traffic assignment model based on travel demand[J]. China Journal of Highway & Transport,2010,23(6):85-91(in Chinese).

[7] 雷磊. 城市与区域一体化的出行需求分析理论与方法研究[D]. 成都:西南交通大学,2010.
LEI L. A study on the theory and method of travel demand analysis of urban and regional integration[D]. Chengdu:Southwest

北航学报 赠阅

Jiaotong University,2010(in Chinese).

[8] 黄健. 含拼车的多模式交通网络配流模型及系统优化[D]. 南京:南京大学,2016:18-24.
HUANG J. Multi-mode traffic network assignment model and sys tem optimization for carpooling[D]. Nanjing:Nanjing University,2016:18-24(in Chinese).

[9] 刘京京. 考虑距离因素的城市多模式交通配流模型[D]. 北京:北京交通大学,2016:13-20.
LIU J J. Multi-mode traffic assignment model considering distance factors[D]. Beijing:Beijing Jiaotong University,2016:13-20(in Chinese).

[10] ZHOU X,HANI S,ZHANG M. Dynamic micro-assignment modeling approach for integrated multimodal urban corridor management[J]. Transportation Research Part C:Emerging Technologies,2008,16(2):167-186.

[11] 李靖睿. 调节公交车与私家车的关系对经济发展的影响[J]. 现代经济信息,2012(2):318.
LI J R. Adjust the relationship between bus and private car on economic development[J]. Modern Economic Information,2012(2):318(in Chinese).

[12] 郝忠娜,何玉宏. "新地铁时代"常规公交的扬长避短与创新优化[J]. 上海城市管理,2013(2):59-62.
HAO Z N,HE Y H. Pros and cons of regular traffic and transportation and innovative improvement in new metro age[J]. Shanghai Urban Management,2013(2):59-62(in Chinese).

[13] 罗焕然. 北京地铁定价的思考:基于对私家车的替代角度[J]. 北京市经济管理干部学院学报,2014(2):13-18.
LUO H R. A study on the metro ticket pricing in beijing-Based on the substitute for private cars[J]. Journal of Beijing Institute of Economics and Management,2014(2):13-18(in Chinese).

[14] SIMS C A. Macroeconomics and reality[J]. Econometrica,1980,48(1):1-48.

[15] PRIMICERI E. Time varying structural vector auto-regressions and monetary policy[J]. The Review of Economic Studies,2005,72(3):821-852.

[16] NAKAJIMA J. Time-varying parameter VAR model with stochastic volatility:An overview of methodology and empirical applications[J]. Jouchi Nakajima,2011,29:107-142.

[17] 王俊秀. 中国汽车社会发展报告:汽车社会与规则[M]:北京:社会科学文献出版社,2013:78-90.
WANG J X. China automotive social development report:Automotive society and rules[M]. Beijing:Social Science Academic Press,2013:78-90(in Chinese).

[18] MICHEAL A,TAYLOR E,WOOLLEY Z. Integration of the global positioning system and geographical information systems for traffic congestion studies[J]. Transportation Research Part C Emerging Technologies,2000,8(1-6):257-285.

[19] RICHARDSON A,TAYLOR M. Travel time variability on commuter journeys[J]. High Speed Ground Transportation Journal,1978,12(1):77-99.

作者简介:

马晓磊　男,博士,副教授,博士生导师。主要研究方向:交通大数据、公共交通。

孙硕　男,硕士研究生。主要研究方向:多模式交通系统分析。

丁川　男,博士,讲师。主要研究方向:交通行为分析。

王云鹏　男,博士,教授,博士生导师。主要研究方向:车联网及车路协同。

北京航空航天大学学报

北航学报 赠阅

A coupling analysis of multimodal transportation demands based on TVP-VAR models

MA Xiaolei, SUN Shuo, DING Chuan*, WANG Yunpeng

(School of Transportation Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: Under the condition of multimodal travel choices in large cities, there is a close relationship of traffic demands among subway, buses and private cars. In order to explore the coupling relationship and time-varying feature among multimodal traffic demands, with road congestion index representing the traffic demands of private cars, this paper adopts the time-varying parameter vector autoregressive (TVP-VAR) model to analyze one-month travel demands of subway, buses and road congestion index on weekdays. The empirical results of two traffic zones demonstrate that: Depending on different land use types, the relationship among subway demand, bus demand and private car demand does not significantly change in trend, but varies in time scale; In urban traffic zones, there is an interactive relationship among traffic demands of subway, buses and private cars. As subway demand increases, less passengers will switch to buses. When bus demand increases, subway demand will increase correspondingly. The increase of private car travel demand will increase the subway travel demand but reduce the bus travel demand. The research is helpful to understand the coupling relationship among different travel modes in our country at the present stage, and thus to better cope with the issue of urban traffic congestion.

Keywords: traffic demands; multimodal travel choices; time-varying analysis; coupling influence; TVP-VAR model

北京航空航天大学学报

Received: 2017-01-12; **Accepted**: 2017-03-06; **Published online**: 2017-04-14 16:54

URL: kns.cnki.net/kcms/detail/11.2625.V.20170414.1654.001.html

Foundation items: National Natural Science Foundation of China (51408019, 71503018, U1564212)

* **Corresponding author**. E-mail: cding@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2016.0979

面向高分辨率遥感影像的地形辐射校正方法

丁一帆[1,2,*], 尤红建[1], 张浩[3], 陈双军[4], 许斌[1,2], 孙韬[1]

(1. 中国科学院电子学研究所, 北京 100190; 2. 中国科学院大学, 北京 100049;
3. 中国科学院遥感与数字地球研究所, 北京 100094; 4. 信息工程大学 地理空间信息学院, 郑州 450052)

摘 要: 地形辐射校正对获取准确的地表定量遥感精度意义重大。针对传统地形辐射校正模型不适用于高分辨率遥感影像的问题,提出了一种基于辐射传输模型,同时严控误差源的地形辐射校正方法,以资源三号 01 星高分辨率全色及多光谱遥感影像为例进行相关实验,实现对高分辨率遥感影像的地形辐射校正,并进行了主客观分析与评价。分析结果表明:本文提出的地形辐射校正模型和方法,能有效解决全色遥感影像在绝对辐射定标系数缺失情况下校正效果差以及如何保持高分辨率遥感影像细节等难点,较传统方法更适用于高分辨率遥感影像。

关 键 词: 高分辨率遥感影像; 地形辐射校正; 辐射传输; 绝对定标系数; 数字表面模型 (DSM)

中图分类号: P237

文献标识码: A **文章编号**: 1001-5965(2018)01-0027-09

在复杂地形区域,地表起伏对地物及目标的辐射特性有深刻影响。①因地形遮挡而产生的阴影造成阴影区地物辐射减弱;②由于倾斜地表接收太阳直射、大气散射、周围背景反射辐射与平坦地表具有明显差异。上述因素导致地物光谱严重失真,不考虑这些因素的定量化处理难以达到应用所需的定量遥感精度。地形辐射校正的实质是将所有像元的辐射量度变换到某一参考平面上(一般取水平面),对因地形不规则而产生的遥感影像辐射亮度或反射率畸变进行修正,主要包括多光谱影像光谱失真修正问题和全色影像的阴影信息恢复问题。传统的地形辐射校正模型有 3 类:①经验模型,例如波段比模型[1]等,模型没有物理意义,无法实现真实的定量化反演。②半物理半经验模型,目前应用最多的是 C 校正模型,该模型是基于样本统计来建立回归方程的校正,半经验参数 C 没有确切物理含义,且其获取依赖于具体样本影像,模型不具有普遍适用性[2]。③物理模型,Sandmeier 校正模型是其中经典模型之一,其对于地物反射辐射部分计算较为简单,该模型只是简单地通过计算周围像元的平均反射率来获取周围地形的反射辐射[3-4],这对于山脊或山谷像元的地形辐射校正会存在问题[2]。ATCOR 地形辐射校正模型是当前应用较为广泛的商用算法,也是比较严格的物理模型,该模型能同时进行大气校正和地形辐射校正,其中 ATCOR3 模型对卫星观测角、太阳高度角以及高程都有约束条件(规定卫星观测角应小于 ±8°,且太阳高度角保持不变。除此之外,该模型还规定研究区的高程必须小于 1.5 km),因此该模型适合用于中低海拔坡度中等区域[5]。ATCOR4 模型引入了地表二向性分布函数(BRDF)模型,但

收稿日期: 2016-12-29; **录用日期**: 2017-02-06; **网络出版时间**: 2017-03-20 10:35
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170320.1035.001.html
基金项目: 国家自然科学基金 (61672076)
* **通信作者**. E-mail: dingyifan14@mails.ucas.ac.cn

引用格式: *丁一帆, 尤红建, 张浩, 等. 面向高分辨率遥感影像的地形辐射校正方法[J]. 北京航空航天大学学报, 2018, 44(1): 27-35. DING Y F, YOU H J, ZHANG H, et al. Topographic correction method for high-resolution remote sensing images [J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(1): 27-35 (in Chinese).*

北航学报 赠阅

参数的确定依赖于样本影像,因此计算结果精度有差异。可以看出,传统模型存在各自的优缺点。

综上所述,利用数字高程模型(DEM)或者数字表面模型(DSM)结合卫星传感器所接收的实际辐射观测值来获取地表真实参数是地形辐射校正的一般思路[6]。但是,目前几乎所有的模型及实验都主要针对于中低分辨率的影像,例如 MODIS、TM 等影像,对于高分辨率处理相对较少。目前,随着高分辨率遥感卫星发射日益增多,应用需求也越来越多,而地形效应在高分辨率影像的表现更加明显,对辐射信号影响更大。传统的地形辐射校正方法因对分辨率提高所带来的误差源控制以及模型建模等因素考虑不足,对于高分辨率影像适用性存在一定问题。究其原因,除了高精度、高分辨率的 DEM(或 DSM)不易获取是主要限制外,针对高分辨率影像的地形辐射校正模型和方法自身也存在一定问题。与此同时,相当数量遥感卫星针对全色影像并不提供绝对辐射定标系数,因此如何在没有定标系数的条件下实现对高分辨率全色影像进行地形辐射校正也是亟待解决的问题。

针对上述问题,本文重点利用辐射传输模型和更高分辨率(10 m)的 DSM,建立了一种面向高分辨率遥感影像的地形辐射校正方法。通过利用北京北部山区的资源三号 01 星高分辨率多光谱和全色遥感影像进行地形辐射校正实验。实验结果表明,本文方法优于传统方法,更适合于高分辨率影像地形辐射校正。

1 地形辐射校正误差源分析

众多模型中,基于物理模型的地形辐射校正的实质是地形因子和大气参数同步解算,在建模过程中需综合考虑各种影响因素,考虑的影响因素越多,输入的参数(例如绝对辐射定标系数、大气参数等)越真实,中间计算的过程(例如 DEM 与影像分辨率越接近,两者配准越精确)越合理,最终的校正结果就越准确。因此,影响高分辨率影像地形辐射校正精度的主要因素可概括为 2 个:①校正模型;②输入数据及参数的误差。

1.1 校正模型分析

对于复杂地形而言,地面一点所接收的有效光照可分解为 3 项:①太阳的直射辐照度 E_d;②天空漫散射辐照度 E_f;③周围邻近山坡辐照度 E_a。复杂地形地表总入射太阳辐照度可表示为

$$E = E_d + E_f + E_a \tag{1}$$

如果模型中只考虑前一或两项辐射,计算较为简单,但效果并不好。原因在于当目标位于深谷或阴影区时,周围邻近山坡辐射项 E_a 的辐射量可能会在一点所接收的总辐射量中占有较大的比重,而模型却没有如实反映。除此之外,研究发现,当地表有雪覆盖或是近红外波段有植被覆盖时,邻近地形的反射辐射也不可忽略[7]。这些都充分说明地形辐射校正模型中仅考虑太阳直接辐射是不够的[8]。

本文在建模上全面考虑太阳直接辐射、天空漫散射辐射和周围邻近山坡辐射等所有辐射项,以真实反映辐射能量分布的实际情况。

1.2 输入数据及参数的误差分析

根据物理模型,输入的误差源主要表现为 DEM(DSM)、大气辐射传输计算和地表特性3 类。具体如下:

1)计算地形因子的 DEM 或 DSM

对于高分辨率遥感影像而言,DSM 包含地物高程信息,比 DEM 更具备适应性,而且 DSM(或 DEM)应与高分辨率影像相当或稍低[9],如果相差较大,则 DSM 无法准确描述真实地形起伏情况。同时,无论用什么方法,DSM 与高分辨率影像的配准精度也非常重要,这些将决定后续计算地形因子的精度。

2)大气辐射传输计算

大气参数估算是地形辐射校正中非常重要的一步,如果有条件最好能够针对实验区域同步(或准同步)实测大气参数[10],减少估算误差,最大限度提高地形辐射校正的精度。

3)地表特性

地表特性的假设是地形辐射校正中的一项重要误差源。很多地形辐射校正模型是基于朗伯体假设,把地表理想化为均匀、各向同性的表面,而实际地表是具有明显方向性的非朗伯体。可采用常数 K 简化描述地表二向性分布函数[11]。

以上 3 个误差源的分析表明,DSM 的选取及配准精度影响最大;其次是地表特性描述准确程度;最后是大气参数精度。研究表明:大气参数误差对地形辐射校正影响相对较小,特别是天空漫散射和程辐射,并且在不同大气模式下,大气参数误差引起地表反射率的反演误差相差不大[12]。因此,针对高分辨率影像的地形辐射校正,需要从上述源头控制好误差,才能提升高分辨率影像的地形辐射校正精度。本文通过控制这些误差,研究了一种适用于高分辨率影像的地形辐射校正方法。

2　地形辐射校正方法

2.1　光照分解模型

大部分地物都是由 3 部分辐射综合作用经地表目标的方向反射进入传感器而被观测。根据 Sandmeier 和 Itten[13] 的模型提出,在太阳有效入射角 i_s 和天顶角 θ_s、地表目标所处地形坡度 S 和坡向 A 下,山区地表目标接收的总入射辐射照度 E_{inTotal} 为

$$E_{\text{inTotal}} = E_{\text{d}}^{\text{h}} \frac{\cos i_s}{\cos \theta_s} b(x,y) + E_{\text{f}}^{\text{h}}(1-K)V_{\text{f}} + E_{\text{f}}^{\text{h}} K \frac{\cos i_s}{\cos \theta_s} b(x,y) + E_{\text{a}} \tag{2}$$

式中:$b(x,y)$ 为二进制阴影遮蔽因子,用来判断像元是否接收到直射辐射,当处于非阴影区时 $b=1$,处于阴影区时 $b=0$;V_{f} 为天空开阔度因子,定义为从坡面像元观察的天空区域与在无遮挡的水平面上观察的天空区域之比,取值为 0 ~ 1 之间;等式右边第 1 项表示太阳直射辐照度,E_{d}^{h} 为平坦地表接收的太阳直射辐照度;等式右边第 2 项和第 3 项分别表示环日各向异性散射辐照度与天空各向同性散射辐照度,两者综合作用为天空漫散射辐照度 E_{f},E_{f}^{h} 为平坦地表接收的天空漫散射辐照度,K 为各向异性指数,表示环日各向异性散射占天空漫散射的权重,用地面法线方向接收的太阳直射辐照度与大气层顶太阳辐照度之比计算;等式右边第 4 项表示周围地形的反射辐射 E_{a},针对高分辨率遥感影像,此项辐射不可忽略。

2.2　多光谱遥感影像地形辐射校正模型

根据光照分解模型,推导出多光谱遥感影像地形辐射校正模型如下:

$$\rho(x,y) = \pi[(c_0 + c_1 \text{DN}(x,y)) - \text{DN}_{\text{ba}}] / (\tau_{\text{v}}\{b(x,y)E_{\text{d}}\tau_{\text{s}}\cos i_s(x,y) + E_{\text{dif}}[b(x,y)\tau_{\text{s}}\cos i_s(x,y)/\cos\theta_s + (1-b(x,y)\tau_{\text{s}})V_{\text{f}}] + E_{\text{iter}}\}) \tag{3}$$

式中:$\rho(x,y)$ 为地物反射率;DN 为卫星影像灰度值;DN_{ba} 为程辐射对应的图像 DN 值,在无法准确估算的情况下,可以采用图像有效像元区域最小值来表示;c_0 和 c_1 为影像绝对辐射定标系数;E_{iter} 为临近像元的多次反射辐照度;模型中与大气相关的因子 τ_{v}、τ_{s}、E_{d} 和 E_{dif} 分别为大气上行透射率、下行透射率、地面直射辐照度和地面散射辐照度,可直接由 6SV 软件计算得到。

模型中与地形相关的因子有遮蔽因子 $b(x,y)$、天空开阔度因子 V_{f}、太阳有效入射角 i_s 以及邻近地物反射的天空观测因子 V_{sky},所有这些地形相关因子均可由 DSM 计算的坡度、坡向等地形参数获取。其中太阳有效入射角 i_s 公式为

$$\cos i_s = \cos\theta_s\cos S + \sin S\sin\theta_s\cos(\phi_s - A) \tag{4}$$

式中:ϕ_s 为太阳方位角。坡度 S、坡向 A 是通过 DEM 或者 DSM 计算出来的。邻近像元的多次反射辐照度 E_{iter} 的计算公式如下:

$$E_{\text{iter}}^{(i)} = E_{\text{g}} \frac{\rho_{\text{terrain}}^{(i-1)}(1 - V_{\text{sky}}(x,y))}{1 - \rho_{\text{terrain}}^{(i-1)}(1 - V_{\text{sky}}(x,y))} \tag{5}$$

式中:E_{g} 为水平表面接收的太阳的直射辐照度与漫散射辐照度之和;$V_{\text{sky}} = 0.5 + 0.5\cos S$;$\rho_{\text{terrain}}^{(i-1)}$ 表示像元周边一定范围内(一般 0.5 km)的平均反射率,属于中间结果。多次反射辐照度 E_{iter} 的计算通常迭代 2 ~ 3 次即可。

2.3　全色遥感影像地形辐射校正模型

全色遥感影像的地形辐射校正方法基于严格的物理模型,综合了辐射传输模型与经验系数校正的优点。如果全色遥感影像在有绝对辐射定标系数前提下,上述多光谱遥感影像地形辐射校正模型也适用于全色。但实际上很多型号卫星并不提供全色遥感影像的定标系数,因而无法得到绝对地表反射率。例如 HJ-1[14] 和 CBERS-04 等卫星,以及 GF-2 卫星的 PSM2 载荷等。因此,对于这些定标系数缺失的全色遥感影像,本文提出了如下校正方法计算相对反射率实现地形辐射校正:

$$\rho_{\text{t}}(x,y) = \pi(\text{DN}(x,y) - \text{DN}_{\text{ba}}) / (\tau_{\text{v}}\{b(x,y)E_{\text{d}}\tau_{\text{s}}\cos i_s(x,y) + E_{\text{dif}}[b(x,y)\tau_{\text{s}}\cos i_s(x,y)/\cos\theta_s + (1-b(x,y)\tau_{\text{s}})V_{\text{f}}] + E_{\text{iter}}\}) \tag{6}$$

式中:DN_{ba} 可选择图像有效像元区域最小值代表该数值;V_{f} 由坡度计算得到。

令 $\xi = (1 - V_{\text{f}})/10$,由于无法得到绝对地表反射率,难以进行多次迭代,可采用式(7)简化计算:

$$E_{\text{iter}} = (E_{\text{d}} + E_{\text{dif}})\xi^2/(1-\xi) \tag{7}$$

本文针对全色遥感影像的地形辐射校正采用了 3 种方法,即绝对辐射定标系数法(即多光谱遥感影像地形辐射校正模型)、缺少定标系数的相对反射率法(即本节模型)以及相对反射率后带掩模优化法。其中,相对反射率后带掩模优化法主要考虑到影像分辨率与所采用 DEM 分辨率差异大,通过加掩模有利于补偿校正效果。

2.4　处理步骤及流程

步骤 1　重投影、正射校正与配准。

按照影像的投影方式对 DSM 进行重新投影,确保两者投影一致。利用有理函数模型(RPC)参数及 DSM 数据对一级影像进行正射校正,获取正

北航学报 赠阅

射影像。通过正射校正,确保影像和 DSM 数据之间配准精度。

步骤 2 数据预处理。

1)裁剪。对 DSM 和影像的重叠区域进行裁剪,获取对应相同范围的 DSM 数据和影像数据。

2)辐射定标。将绝对辐射定标系数应用到多光谱及全色遥感影像中,得到辐亮度影像(单位为 $W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$)。

步骤 3 地形辐射校正。

地形辐射校正主要分 3 步,可采用图 1 所示流程图表示。

1)地形相关因子计算。通过 DSM 完成坡度、坡向、天空观测因子、遮蔽因子计算。

2)大气参数计算。输入实测气象数据,通过 6S 模型计算大气辐射传输量输出项,分别是:到达地面直射辐照度、到达地面漫射辐射辐照度、日地平均距离处太阳辐照度、程辐射、上行透射率、半球反照率。

3)地表反射率计算。基于辐射传输参量计算校正后的地表反射率。

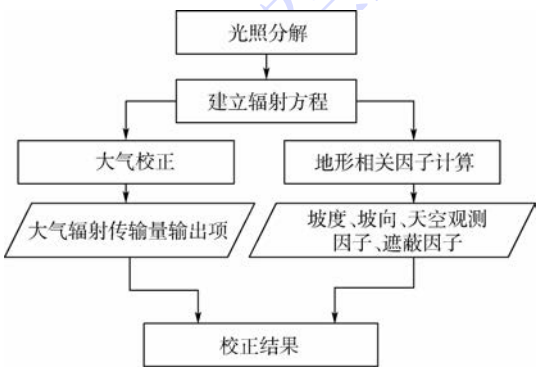


图 1 地形辐射校正流程图

Fig. 1 Topographic correction flowchart

3 地形辐射校正实验

3.1 实验数据

3.1.1 图像数据

实验区域位于北京北部以十三陵水库为中心的山区,该区域以山地为主,海拔 800 ~ 1 000 m。实验数据为 2016 年 5 月 16 日成像的资源三号 01 星多光谱遥感影像及全色遥感影像共 2 景,以 DSM 数据的实际分布情况作为特征区域的选择依据,本实验在多光谱、全色 2 景遥感影像中分别选取 10 个特征区域作为样本影像。其中,多光谱遥感影像分辨率为 5.8 m,全色遥感影像分辨率为 2.1 m。影像中心经度为 116.89°,中心纬度为 40.03°。数据格式为 Geotiff,影像产品级别为1 级加上配套 RPC 参数文件,可自行处理生成 2 级或正射影像。配套的地形数据是资源三号 01 星立体像对数据通过摄影测量方法生成的 10 m 分辨率的 DSM 数据(见图 2)。

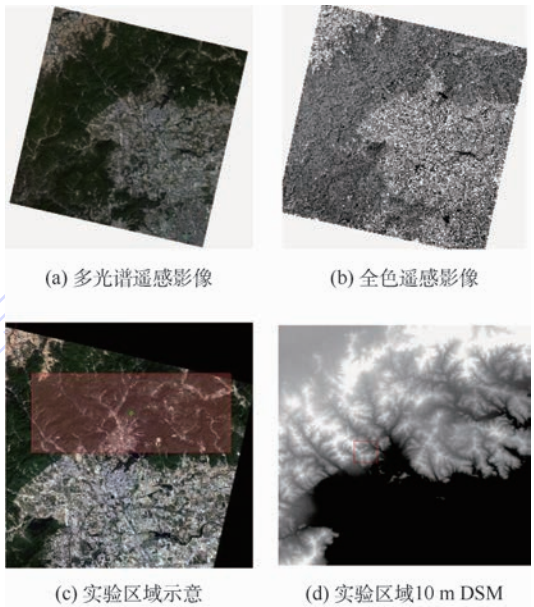


图 2 实验图像示例

Fig. 2 Example of experimental images

3.1.2 辅助数据

1)绝对辐射定标系数

多光谱遥感影像的绝对辐射定标系数来自于资源卫星中心网站,由于影像数据获取时间为 2016 年 5 月 16 日 10:58:43,采用最新 2015 年提供的绝对辐射定标系数;全色遥感影像 2015 年没有绝对辐射定标系数,但 2014 年提供了绝对辐射定标系数。全色与多光谱遥感影像的绝对辐射定标系数见表 1。

利用绝对辐射定标系数将卫星影像 DN 值转换为辐亮度图像的公式为

$$L_e(\lambda_e) = G \cdot DN + Offset \tag{8}$$

式中:$L_e(\lambda_e)$ 为转换后辐亮度,$W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$;$G$ 为定标斜率;Offset为绝对辐射定标系数偏移量,空缺值为 0。

表 1 资源三号 01 星遥感影像绝对辐射定标系数

Table 1 Absolute radiometric calibration coefficients for ZY3-01 satellite images

载荷	波段	光谱范围/μm	定标系数增益项	定标系数偏移量
多光谱	Band 1	0.45 ~ 0.52	0.233 0	0
	Band 2	0.52 ~ 0.59	0.216 2	0
	Band 3	0.63 ~ 0.69	0.178 9	0
	Band 4	0.77 ~ 0.89	0.194 9	0
全色	Pan	0.50 ~ 0.80	0.170 8	0

北航学报 赠阅

2）光谱响应函数

在资源中心网站查询多光谱波段及全色数据的光谱响应函数。

3）大气数据

影像区域含有 2 个 AERONET 观测站，分别为 Beijing 站（图 3 中绿点所示），经度 116.381°，纬度 39.977°，以及 Beijing_CAMS 站（图 3 中红点所示），经度 116.381°，纬度 39.933°。

取成像时刻前后 15 min 平均值作为图像大气数据。根据 Angstrom 公式，分别计算 2 个站点 550 nm 气溶胶光学厚度，然后求均值得到 AOT550 = 0.156，水汽含量为 0.912 g/cm^2，可以发现影像获取时刻区域内能见度较高，气溶胶影响相对较小。



图 3　观测站位置示意图

Fig. 3　Schematic of location of observation station

3.2　处理结果及评价

3.2.1　处理结果对比

1）多光谱遥感影像处理前后辐射对比（见图 4）。

2）多光谱遥感影像处理前后细节对比（见图 5）。

3）全色遥感影像处理前后辐射对比（见图 6 ~ 图 8）。

3.2.2　结果评价与分析

针对地形辐射校正的结果评价一般分为主观评价和客观评价。主观评价以目视评价为主；客观评价以校正前后客观指标值变化进行量化评判，主要包含 2 部分，一是统计特征评价，例如直方图、标准差等，二是光谱一致性评价，例如相关系数、变异系数等。

1）目视评价

从处理前后不同区域的辐射对比图（见图 4 ~ 图 8）中可见，校正后影像的立体感变弱，阴坡区域辐亮度提升，说明本文方法较好地消除了地形起伏带来的像元辐亮度畸变或地物反射率畸变。其中对于高分辨率多光谱遥感影像，位于阴、阳坡的同一类地物光谱差异明显变小，同时较好地保留了影像细节。对于全色遥感影像，本文分别采

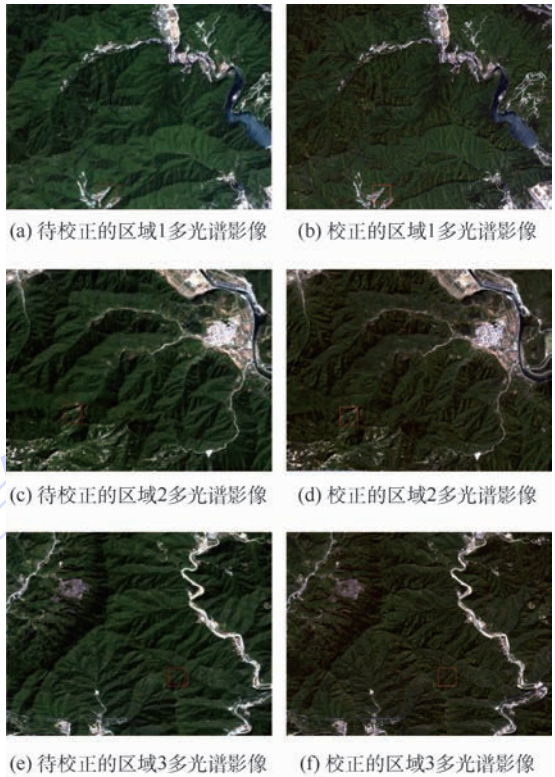


图 4　多光谱遥感影像处理结果辐射对比

Fig. 4　Comparison of radiation of multispectral image experimental process results

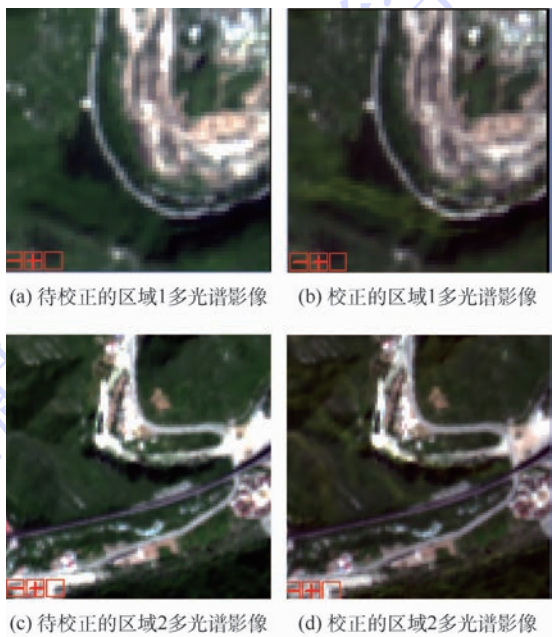
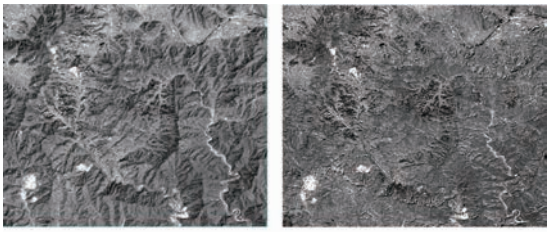


图 5　多光谱遥感影像处理细节对比

Fig. 5　Comparison of process results for details of multispectral images

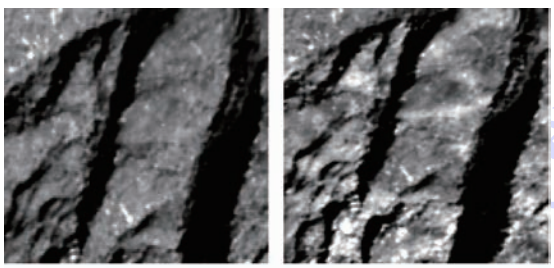
北航学报 赠阅



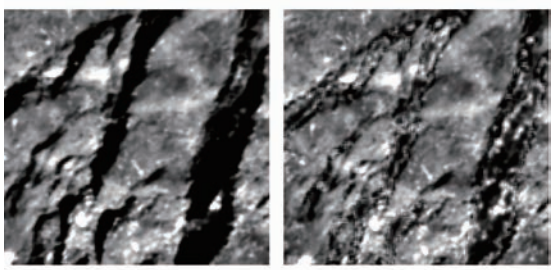
(a) 全色处理前 (b) 本文方法全色处理后

图 6 全色遥感影像处理结果辐射对比

Fig. 6 Comparison of radiation of panchromatic image process results



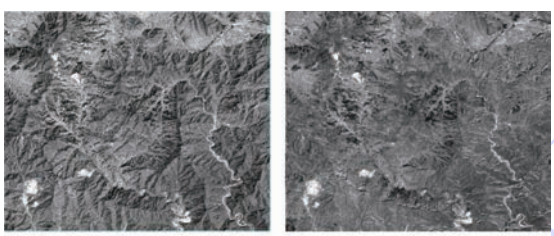
(a) 原始影像 (b) 绝对辐射定标系数法



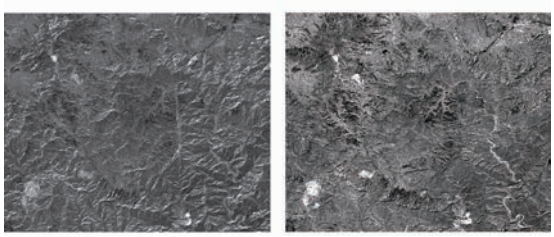
(c) 相对反射率法 (d) 相对反射率后带掩模优化法

图 7 针对全色遥感影像的不同方法结果对比

Fig. 7 Comparison of different methods for panchromatic images



(a) 原始影像 (b) C校正法



(c) ATCOR (d) 本文方法

图 8 不同方法的全色遥感影像校正结果对比

Fig. 8 Comparison of correction results for panchromatic images among different methods

用 3 种方法处理,绝对辐射定标系数法采用的是 2014 年度提供的定标系数。通过 3 种方法校正结果对比,绝对辐射定标系数法在某些阴影区域信息并未恢复出来,原因可能是 2014 年的绝对辐射定标系数不能适用于全色遥感影像。相对反射率法部分恢复阴影信息,相对反射率后带掩模优化法则较好地恢复了阴影区域信息。

实验中同时还采用了传统的 C 校正和商业软件 ATCOR 针对相同的数据源进行处理并对比。其中 C 校正结果稍差,体现为细节缺失;ATCOR 算法与本文方法结果基本相当,但是 ATCOR 算法不能处理缺少绝对辐射定标系数的图像,且对于图像分辨率远高于 DSM 分辨率的情况处理效果不甚理想。本文方法具有一定的普适性。

2)统计特征评价

进行统计特征评价前,首先需要统一对比基准,本实验采用地表未经地形辐射校正的反射率影像作为处理前影像,与经地形辐射校正后的反射率结果进行对比,统计特征的实验区域如图 9 所示。对比的统计特征值分别为直方图、均值、标准差、变异系数以及 X、Y 方向剖面图。

由图 10 可以看出,校正前存在不止一个波峰,说明存在坡度、坡向影响,校正后呈高斯分布,说明与自然界地物随机性一致。

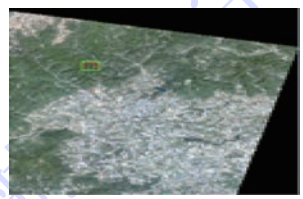
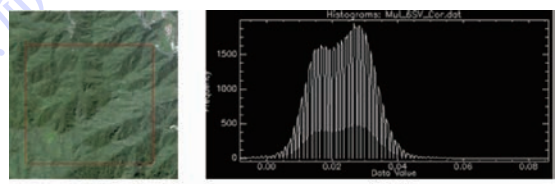
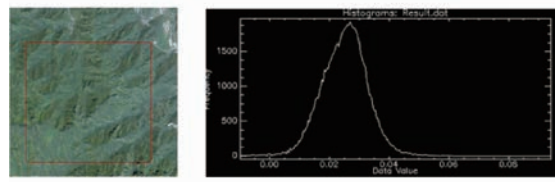


图 9 统计特征区域示意图

Fig. 9 Schematic of statistical characteristic area



(a) 校正前统计区域 (b) 校正前统计区域的直方图



(c) 校正后统计区域 (d) 校正后统计区域的直方图

图 10 多光谱遥感影像处理前后直方图统计对比

Fig. 10 Statistic comparison of histogram for multispectral image process results

北航学报 赠阅

由表 2 可以看出,校正后标准差较校正前有所降低,整体均值没有明显变化且略有提升,说明校正了地形效应影响,同时较好保持了影像细节信息。这里同时引用变异系数[15],定义为标准差与均值的比值,一般如果处理后有效果,该值应该变小,从统计结果也可看到,处理后每个波段的变异系数均有所降低。

由表 3 可以看出,与多光谱遥感影像类似,全色遥感影像经过不同方法处理后,标准差及变异系数都有所降低,整体均值都略有提升,符合地形效应影响减弱或消除的特征,说明不同方法针对全色遥感影像都有效。其中从统计特征中的变异系数项看,本文方法结果稍优于 C 校正和 ATCOR 算法。需要说明的是,全色遥感影像对比的统一基准是相对反射率影像。

表 2　多光谱遥感影像校正前后统计特征值对比

Table 2　Comparison of statistical eigenvalues before and after correction for multispectral images

波　段	校正前			校正后		
	均值/10^{-2}	标准差/10^{-2}	变异系数	均值/10^{-2}	标准差/10^{-2}	变异系数
1	0.407	0.420	1.032	0.505	0.420	0.832
2	2.339	0.832	0.356	2.529	0.757	0.299
3	2.400	0.655	0.273	2.630	0.579	0.220
4	29.049	5.722	0.197	31.469	5.107	0.162

表 3　全色遥感影像校正前后统计特征值对比

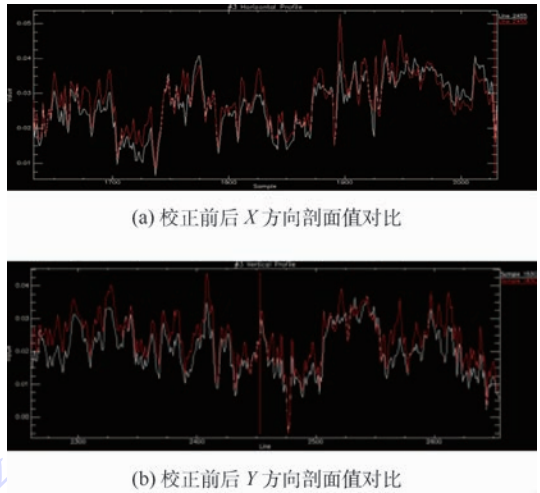
Table 3　Comparison of statistical eigenvalues before and after correction for panchromatic images

统 计 量	原始图像	C 校正法	ATCOR 算法	本文方法
均值/10^{-2}	7.59	10.07	8.28	9.55
标准差/10^{-2}	4.36	2.98	2.03	2.10
变异系数	0.574	0.296	0.245	0.220

分析结果如图 11 所示,图中白色曲线为校正前剖面曲线,红色曲线为校正后剖面曲线。由 X 和 Y 方向剖面线数值变化可看出,经过地形辐射校正后的区域数值幅度明显高于校正前数值,这说明本文方法的有效性。

3) 光谱一致性评价

对于高分辨率多光谱遥感影像,可采用相似度即相关系数指标对实验区影像进行统计,以分析影像波谱响应的相似程度。具体方法是:校正前后影像中,先选取若干个(5 ~ 10 个)阴坡、阳坡区域,各自构成区域对(见图 12(a)),分别计算校正前后相邻阴坡、阳坡区域对中各自点集光谱曲线的相关系数。除此之外,也对处理前后相同位置的阴阳坡点对的光谱线离散程度进行对比。

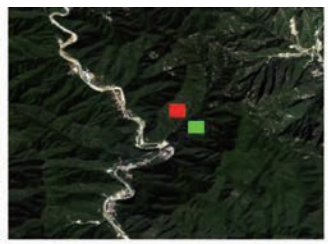


(a) 校正前后 X 方向剖面值对比

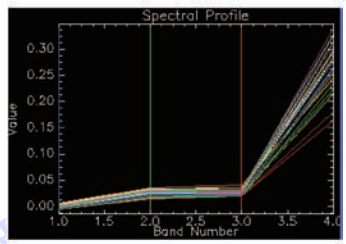
(b) 校正前后 Y 方向剖面值对比

图 11　多光谱遥感影像处理前后 X 与 Y 方向剖面对比

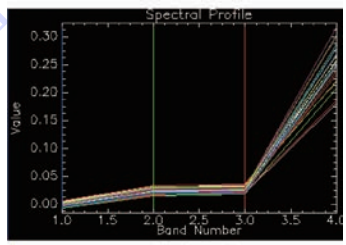
Fig. 11　Comparison of X and Y direction profiles before and after correction for multispectral image



(a) 实验区域点位



(b) 校正前阴阳坡光谱曲线图



(c) 校正后阴阳坡光谱曲线图

图 12　光谱一致性评价图

Fig. 12　Evaluation chart of spectral consistency

从图 12(b)、(c)的阴阳坡光谱曲线图可看到,校正后的曲线离散度明显小于校正前。从平均相关系数来看,尽管校正前后都接近于 1,但校正后仍有 0.3% 的提升,说明地形辐射校正减弱了地形对地物信息遮挡的效应,从整体上改善了图像质量。

北航学报 赠阅

4 结 论

本文利用资源三号 01 星全色及多光谱遥感影像,通过相应 DSM 数据、实时观测大气数据以及绝对辐射定标系数等配套输入数据,反演地表反射率或相对反射率,探讨针对高分辨率影像的地形辐射校正方法。

通过对实验结果评价分析可看出,本文方法实现了高分辨率影像的地形辐射精确校正,影像的纹理细节保真效果较好。同时,本文也针对缺少绝对辐射定标系数的高分辨率全色影像的地形辐射校正提出了新的方法,并进行了不同方法效果的对比,实验结果表明该方法有效。

参考文献(References)

[1] COLBY J D. Topographic normalization in rugged terrain[J]. Photogrammetric Engineering and Remote Sensing, 1991, 57: 531-537.

[2] 陈志明. 遥感影像地形辐射校正方法研究与系统实现[D]. 福州:福建师范大学,2009:4.
CHEN Z M. Research on topographic correction methods for remote sensing image and system implementation[D]. Fuzhou: Fujian Normal University, 2009:4(in Chinese).

[3] 武瑞东. 卫星遥感影像数据的地形影响校正[J]. 遥感信息,2005(4):31-34.
WU R D. Topographic correction of satellite remote sensing image data[J]. Remote Sensing Information, 2005(4):31-34(in Chinese).

[4] SHEPHERD J D, DYMOND J R. Correcting satellite imagery for the variance of reflectance and illumination with topography [J]. International Journal of Remote Sensing, 2003, 24(17): 3503-3514.

[5] RICHTER R. Correction of atmospheric and topographic effects for high spatial resolution satellite imagery[J]. International Journal of Remote Sensing, 1997, 18(5): 1099-1111.

[6] 段四波,阎广建. 山区遥感图像地形辐射校正模型研究综述[J]. 北京师范大学学报(自然科学版),2007,43(3):362-366.
DUAN S B, YAN G J. A review of models for topographic correction of remotely sensed images in mountainous area[J]. Journal of Beijing Normal University(Natural Science), 2007, 43(3):362-366(in Chinese).

[7] PROY C, TANRE D, DESCHAMPS P Y. Evaluation of topographic effects in remotely sensed data[J]. Remote Sensing of Environment, 1989, 30(1):21-32.

[8] 高永年,张万昌. 遥感影像地形辐射校正研究进展及其比较试验[J]. 地理研究,2008,27(2):467-468.
GAO Y N, ZHANG W C. Comparison test and research progress of topographic correction on remotely sensed data[J]. Geographical Research, 2008, 27(2):467-468(in Chinese).

[9] ZHANG Y, YAN G, BAI Y. Sensitivity of topographic correction to the DEM spatial scale[J]. IEEE Geoscience Remote Sensing Letters, 2015, 12(1):53-57.

[10] SOLA I, GONZÁLEZ-AUDÍCANA M, ÁLVAREZ-MOZOS J. Validation of a simplified model to generate multispectral synthetic images[J]. Remote Sensing, 2015, 7(3):2942-2951.

[11] VANONCKELEN S, LHERMITTE S, VAN ROMPAEY A. The effect of atmospheric and topographic correction on pixel-based image composites: Improved forest cover detection in mountain environments[J]. International Journal of Applied Earth Observation and Geoinformaiton, 2015, 35B: 320-328.

[12] 段四波,阎广建,穆西晗,等. 基于 DEM 的山区遥感图像地形辐射校正方法[J]. 地理与地理信息科学,2007,23(6):23-24.
DUAN S B, YAN G J, MU X H, et al. DEM based remotely sensed imagery topographic correction method in mountainous areas[J]. Geography and Geo-Information Science, 2007, 23(6):23-24(in Chinese).

[13] SANDMEIER S, ITTEN K I. A physically-based model to correct atmospheric and illumination effects in optical satellite data of rugged terrain[J]. IEEE Transactions on Geoscience and Remote Sensing, 1997, 35(3):708-717.

[14] JIANG B, LIANG S, TOWNSHEND J R, et al. Assessment of the radiometric performance of Chinese HJ-1 satellite CCD instruments[J]. IEEE Journal of Selected Topics in Applied Earth Observation & Remote Sensing, 2013, 6(2):840-850.

[15] 罗良清,魏和清. 统计学[M]. 北京:中国财政经济出版社,2011:49.
LUO L Q, WEI H Q. Statistics[M]. Beijing: China Financial and Economic Publishing House, 2011:49(in Chinese).

作者简介:

丁一帆 男,博士研究生,高级工程师。主要研究方向:遥感卫星数据处理与应用。

尤红建 男,研究员,博士生导师。主要研究方向:遥感处理与应用。

北航学报 赠阅

Topographic correction method for high-resolution remote sensing images

DING Yifan[1,2,*], YOU Hongjian[1], ZHANG Hao[3], CHEN Shuangjun[4], XU Bin[1,2], SUN Tao[1]

(1. Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China;
3. Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China;
4. School of Geo-spatial Information, Information Engineering University, Zhengzhou 450052, China)

Abstract: Topographic correction is critical to the accuracy of the earth's surface quantitative remote sensing. Traditional topographic corrective models are not so suitable for high-resolution remote sensing images. In order to address the problem, this paper proposes a topographic correction method based on radiation transfer model and strict control over error sources. High-resolution panchromatic and multispectral images of ZY3-01 satellite were taken as examples to conduct relevant experiments, and the topographic correction of high-resolution remote sensing images was realized with subjective and objective analysis and evaluation. The analysis results show that the problems of poor correction effect without absolute radiometric calibration coefficient in topographic correction of panchromatic remote sensing images can be effectively solved and the detail of high-resolution remote sensing images can also be maintained by the proposed model and method. So it is more suitable for high-resolution remote sensing images than traditional methods.

Keywords: high-resolution remote sensing image; topographic correction; radiation transfer; absolute radiometric calibration coefficient; digital surface model (DSM)

北京航空航天大学学报

Received: 2016-12-29; **Accepted**: 2017-02-06; **Published online**: 2017-03-20 10:35
URL: kns.cnki.net/kcms/detail/11.2625.V.20170320.1035.001.html
Foundation item: National Natural Science Foundation of China (61672076)
* **Corresponding author**. E-mail: dingyifan14@mails.ucas.ac.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2016.0964

原子自旋陀螺气室加热电磁噪声抑制实验研究

周斌权[1,2,*], 郝杰鹏[1,2], 梁晓阳[1,2], 全伟[1,2], 刘刚[1,2]

(1. 北京航空航天大学 惯性技术国家级重点实验室, 北京 100083;
2. 北京航空航天大学 仪器科学与光电工程学院, 北京 100083)

摘 要: 原子自旋陀螺仪作为目前最新一类陀螺仪,具有超高的理论精度。碱金属气室是原子自旋陀螺仪承载原子自旋的敏感表头。通过电加热使碱金属达到饱和蒸气压,但是电加热过程中会引入电磁干扰等噪声,进而影响原子自旋陀螺仪的精度和灵敏度。为减小碱金属气室加热的电磁噪声对原子自旋陀螺仪的影响,从加热器结构与加热驱动信号2个方面进行了电磁噪声抑制实验研究。设计了具有磁场噪声抑制作用的异形加热膜,使高频正弦波作为加热驱动信号,构建了碱金属气室集成化无磁电加热单元。通过实验验证,系统的等效磁场噪声优于17 fT/Hz$^{1/2}$,气室内部的温度稳定度优于±0.006 ℃,为原子自旋陀螺仪的性能提升提供了可靠保障。

关 键 词: 原子自旋陀螺仪; 无磁加热; 噪声抑制; 加热膜; 高频正弦波
中图分类号: V241.62^{+}2
文献标识码: A **文章编号**: 1001-5965(2018)01-0036-07

陀螺仪广泛应用于航空航天等导航与制导领域,其精度决定了导航的精度。随着量子技术和微加工技术的发展,基于原子自旋的一些量子器件发展迅速,极大地提升了仪器的精密测量能力[1-2]。原子自旋陀螺是原子自旋器件的一个重要研究领域,其凭借超高理论精度和小体积,有望成为未来主流的一类陀螺[3-5]。

碱金属气室是原子自旋陀螺仪敏感表头的核心部件。为了提高陀螺仪的测量灵敏度,敏感表头内必须有较高密度的碱金属原子。原子的密度正比于碱金属温度,因此需要对气室进行高温加热[3]。根据敏感介质的不同,气室加热的温度不同,通常为70~200 ℃之间[4-5]。原子自旋陀螺仪的精度较高,磁场对其影响较大,因此需要气室加热的电磁噪声尽量小,最理想的情况是没有电磁噪声[6]。此外,原子弛豫时间会随着气室的温度变化,因此,必须实现气室温度的高稳定性控制。

目前,原子自旋陀螺仪气室加热方式有4种:热气流加热、间断电加热、高频电加热和激光加热[7]。北京航空航天大学早期采用热气流加热方式,这种方式虽然不会引入电磁噪声,但是气流扰动较大,温度稳定性不高;间断电加热方式实现比较简单,但是原子自旋陀螺仪需要连续测量,因此这种方式会影响陀螺仪的测量带宽;国外有研究采用激光加热方式,虽没有电磁噪声,但是目前采用的方案加热功率较小;目前关于高频电加热方式的研究还比较少,这种方式实现起来相对复杂,但是温控精度和温度稳定性高[8]。

综上所述,本文采用高频电加热方式,从加热器结构与加热驱动信号2个方向研究电磁噪声对系统的影响。实验表明,系统的等效磁场噪声优于17 fT/Hz$^{1/2}$,气室内部温度稳定度优于±0.006 ℃,为原子自旋陀螺仪的性能提升提供了可靠保障。

收稿日期: 2016-12-22; **录用日期**: 2017-01-13; **网络出版时间**: 2017-03-08 09:29
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170308.0929.001.html
基金项目: 国家"863"计划(2014AA123401); 国家自然科学基金(61227902,61374210)
* **通信作者**. E-mail: bqzhou@buaa.edu.cn

引用格式: *周斌权, 郝杰鹏, 梁晓阳, 等. 原子自旋陀螺气室加热电磁噪声抑制实验研究[J]. 北京航空航天大学学报, 2018, 44(1): 36-42. ZHOU B Q, HAO J P, LIANG X Y, et al. Experimental study on electromagnetic noise suppression of atomic spin gyroscope heating chamber[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(1): 36-42 (in Chinese).*

1　磁场抑制方法及加热器结构设计

碱金属气室是通过电流驱动电加热丝产生焦耳热来实现加热的功能。因此，在电流通过电加热丝时会产生磁场，从而影响原子自旋陀螺仪的灵敏度。为使由加热引入的磁场噪声得到抑制，采用电磁学的理论计算和有限元分析的方法对加热器结构进行优化设计。

为了抑制加热器引入的磁场噪声，将电加热丝布置为平行邻近形式，并且相邻的电加热丝内电流大小相等、方向相反，因此产生的磁场也是大小相等、方向相反。磁场相消示意图如图 1 所示，2 个邻近的方形电加热丝线圈边长分别为 a、b，距离为 d。

$$B_P = B_{\text{out}} - B_{\text{in}} = \frac{4\mu_0 Ia^2}{\pi(a^2+2x^2)\sqrt{2a^2+4x^2}} - \frac{4\mu_0 Ib^2}{\pi(b^2+2x^2)\sqrt{2b^2+4x^2}} \propto d \tag{1}$$

式中：I 为电流强度；μ_0 为真空磁导率。

根据毕奥-萨伐尔定律可以得出，内外 2 个方形线圈在中心轴线距离线圈为 x 的 P 处的磁场计算公式(1)，其与 2 个线圈的距离 d 成正比[9]。因此，为了实现最好的磁场噪声抑制效果，在设计加热器时应尽可能小得减小电加热丝的间距。

设计的加热器结构示意图如图 2(a)所示，电加热丝彼此互相平行，加工到耐高温的材料基底上。相互平行的电加热丝产生大小相等、方向相反的磁场，从而实现磁场噪声的抑制。

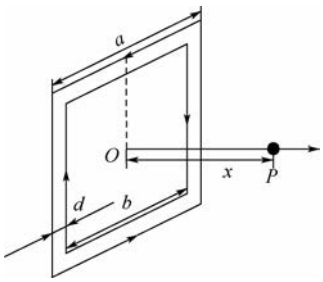


图 1　磁场相消示意图

Fig. 1　Schematic of magnetic field cancellation

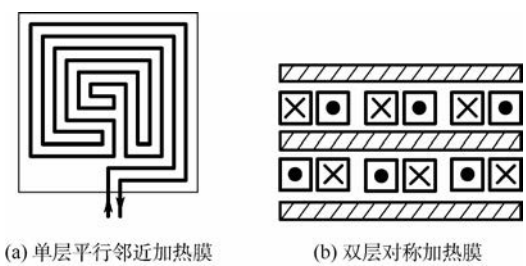


图 2　电加热膜原理示意图

Fig. 2　Schematic of principle of electric heating film

平行邻近的电加热丝在一定程度上使磁场噪声得到抑制，然而电加热丝间距还在 mm 量级，为了更好的抑制磁场噪声，可以将电加热丝在上下层也形成对称结构，电加热丝间距可达 μm 量级，结构示意图如图 2(b)所示。由于加热膜层间距离很小，因此干扰磁场相互抵消效果明显[10]。

电加热膜结构如图 3 所示，基底材料是耐高温的，电加热丝是无磁的镍铬合金，其间距为 0.2 mm，双层间距为 50 μm。

利用电磁场分析软件 ANSYS MAXWELL 对设计的电加热膜进行了磁场分析[11]。实际应用中，电加热膜用于边长为 12 mm 的方形碱金属气室，因此分析了电加热膜中心轴线上 0 ~ 12 mm 的区域。此外，电加热膜的工作电流为 200 mA。单层加热膜与双层加热膜的磁场分析结果如图 4 所示。

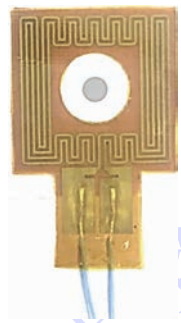


图 3　双层对称结构电加热膜实物图

Fig. 3　Picture of electric heating film with double-layer symmetrical structure

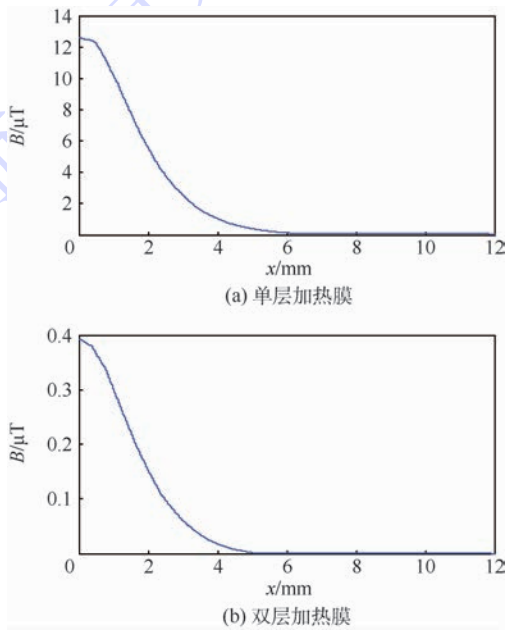


图 4　电加热膜磁场仿真图

Fig. 4　Diagram of electric heating film magnetic field simulation

单层加热膜的最大磁场为 12.616 0 μT,双层加热膜的最大磁场为 0.392 9 μT,单层加热膜产生的最大磁场是双层的 32 倍,表明采用双层结构的加热膜磁场噪声抑制效果较单层加热膜有极大提升。

2 加热驱动信号电磁噪声抑制

常用于加热的驱动信号有直流电和交流电 2 种。用直流电加热引入的噪声主要是白噪声和 1/f 噪声,并且主要为低频段的 1/f 噪声[12]。由于原子自旋陀螺仪的动态范围以及碱金属原子、惰性气体核子的进动频率等都处于低频段,因此,用直流电加热引入的低频噪声必定会影响到原子自旋陀螺仪的精度。

直流电加热噪声功率密度可以表示为

$$e_n^2 = e_{nw}^2\left(\frac{f_{ce}}{f} + 1\right) \tag{2}$$

式中:e_n 为电加热噪声电平;e_{nw} 为白噪声电平;f_{ce} 为转角频率;f 为电流信号的频率。由式(2)可以看出,随着频率的增大,1/f 噪声逐渐减弱,最终趋近于白噪声。因此,为了避免电加热引入磁场噪声,利用调制的技术把电流信号调制到高频。

由于方波信号或者 PWM 信号的频率成分复杂,含有丰富的高次谐波[13],会对原子自旋陀螺仪的检测信号产生串扰,因此,选用频率成分单一的高频正弦波信号作为加热信号。

3 集成化无磁电加热系统设计

集成化无磁电加热系统的结构框图如图 5 所示,主要包括温度传感器、温度测量与 A/D 转换电路、微控制器、高频正弦波信号发生电路、加热膜驱动电路以及加热膜等部分。

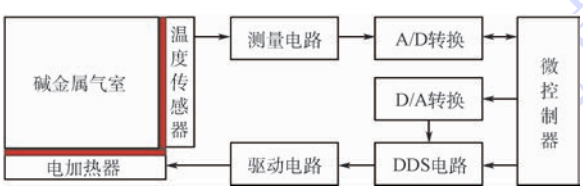


图 5 无磁电加热系统原理框图

Fig. 5 Functional block diagram of non-magnetic electric heating system

3.1 温度测量原理

3.1.1 温度传感器及其测温方式

使用铂热电阻作为温度传感器,具有测温精度高、长期稳定性好、测量范围大的优点[14]。铂热电阻在工作时也会引入磁场噪声。为了避免铂热电阻自身剩磁引入磁场噪声,设计了无磁性的薄膜铂热电阻 Pt1000 作为碱金属气室温度传感器。

铂热电阻传感器主要采用恒流源法的激励方式,因为恒流源法采集到的电信号与电阻呈线性关系,测温精度高;利用恒流激励传感器产生的恒定磁场可以采用主动磁补偿技术将其抵消。此外,采用了四线制引线方式,能有效消除导线电阻和接触电阻带来的测量误差。

采用比例式测量方法,即传感器信号(铂电阻两端的电压)和 A/D 转换器的参考电压由同一个恒流源激励产生。这种方法可以消除温度漂移或噪声对传感器信号和参考电压产生的共同误差[14]。

3.1.2 A/D 转换及其数字滤波

采用比例式测量方法的温度传感器阻值测量公式如下:

$$R_{Pt} = \frac{DR_{ref}}{2^N} \tag{3}$$

式中:D 为 A/D 转换器的输出值;R_{ref} 为参考电阻的阻值;N 为 A/D 转换器的位数。参考电阻采用低温漂电阻,阻值为 5 kΩ;A/D 转换器采用 TI 公司型号为 ADS1248 的 24 位 A/D 转换器。因此,测量电阻的理论分辨率为 0.3 mΩ,转换为测量温度的分辨率为 7.7×10^{-5} ℃。

测温数据设计了一种限幅滤波与均值滤波相结合的复合数字滤波算法,滤波过程如图 6 所示,t_n 为本次采集的测温数据,t_{n-1} 为上次采集的测

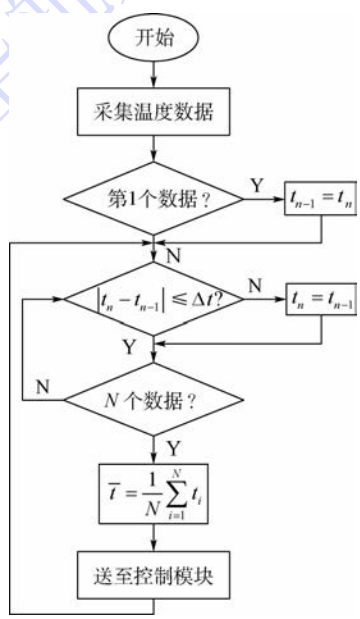


图 6 测温数据复合数字滤波流程图

Fig. 6 Flowchart of temperature measurement data composite digital filter

温数据。首先，判断本次采集的数据是否为第 1 个数据，若是，则将本次采集的测温数据赋给上次采集的测温数据，若否，则进行限幅滤波，即比较本次采集数据与上次采集数据差值的绝对值是否不大于某个设定值 Δt，若是则本次数据保留，若否，则舍弃；然后，进行均值滤波，即对 N 个测温数据取平均值；最后，将结果送至控制模块。

实验发现，经过复合数字滤波算法处理后，测温数据的标准差由 1.159 3 ℃降低为 0.001 7 ℃，说明对测温数据采用复合数字滤波处理可以得到更高的测量稳定度。

3.2　温度控制方法与实现

3.2.1　温度控制方法

为了缩短原子自旋陀螺仪的启动时间，需要碱金属气室加热温控系统提高快速性，因此可以采用比例控制，增大比例系数，然而增大比例系数会产生过大的超调，影响系统稳定，并且存在静态误差。如果加入积分控制，可以消除静态误差，然而又会降低调节速度。所以，单一的控制方法已经不能满足控制要求，因此，为了使碱金属气室温控系统同时满足快速性和稳定性的要求，设计了分段式控制算法，即根据偏差的大小确定采取何种控制算法，分段式控制表示如下：

$$u=\begin{cases}C_1 & e\geqslant E_1\\ \mathrm{Ke} & E_2\leqslant e<E_1\\ [\mathrm{P}\cdot\mathrm{I}\cdot\mathrm{D}]e & E_3\leqslant e<E_2\\ C_2 & e<E_3\end{cases}\tag{4}$$

式中：u 为控制模块的输出量；e 为温度设定值与实际值的偏差；E_1、E_2、E_3 为划分的控制范围边界值；C_1、C_2 为常数值；Ke 表示比例控制方式；$[\mathrm{P}\cdot\mathrm{I}\cdot\mathrm{D}]e$ 表示 PID 控制算法。

当 $e\geqslant E_1$ 时，实际温度与设定温度有很大偏差，控制模块输出最大驱动电压进行加热，快速缩小与设定温度的差值；当 $E_2\leqslant e<E_1$ 时，实际温度接近设定温度，采用比例控制算法，既可保证控制的快速性，又能防止超调和振荡的产生；当 $E_3\leqslant e<E_2$ 时，实际温度在设定温度附近，采用 PID 控制算法；当 $e<E_3$ 时，加热温度超出上限，立即停止加热，防止温度过高造成设备损坏。

控制参数的确定采用了 Ziegler-Nichols 闭环整定法，无需准确辨识系统的模型，可以快速确定 PID 控制参数[15]。具体方法为：使控制系统在比例控制下工作，从零开始增大比例系数，观察由上位机软件实时采集到的测温数据绘制的温度变化曲线，直到系统出现等幅振荡，此时的比例系数即为系统临界系数 K_{Pcrit}，等幅振荡的周期即为系统临界振荡周期 T_{crit}，然后根据表 1 确定 PID 控制器的控制参数，最后对参数进行微调达到最佳效果。

表 1　Ziegler-Nichols 闭环整定法参数

Table 1　Ziegler-Nichols closed loop setting method's parameters

参　数	K_{P}	K_{I}	K_{D}
取　值	$0.6K_{\mathrm{Pcrit}}$	$2K_{\mathrm{P}}/T_{\mathrm{crit}}$	$0.12K_{\mathrm{P}}T_{\mathrm{crit}}$

3.2.2　加热驱动信号产生

高频正弦波信号发生电路如图 7 所示，输出正弦波的频率为 80 ~ 110 kHz，正弦波的幅度由 D/A 转换芯片 AD5620 输出的直流电压信号对其进行调制。

原子自旋陀螺仪碱金属气室的加热温度取决于碱金属原子源，核磁共振陀螺仪使用^{87}Rb 碱金属，工作时气室的温度约为 100 ℃时，正弦波的峰峰值为 80 V。信号发生电路输出的正弦波峰峰值不超过 1 V，远低于加热膜将碱金属气室加热到所需温度要求的电压。为了满足高电压、高增益带宽和高压摆率的要求，采用 APEX 公司放大芯片 PA96，其输出电压高达 300 V，增益带宽可达 175 MHz，压摆率可达 250 V/μs，满足要求。

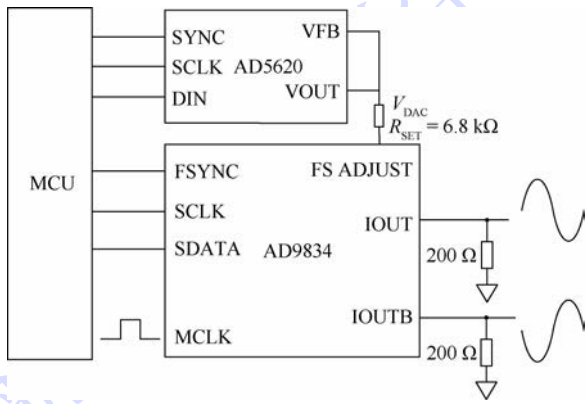


图 7　高频正弦波信号发生电路原理框图

Fig. 7　Functional block diagram of high-frequency sine wave signal generation circuit

4　测试实验与结果

4.1　电加热膜磁场噪声抑制效果测试实验

本文设计的双层对称结构电加热膜的磁场噪声抑制效果测试实验在 4 层磁场屏蔽装置内进行（见图 8），测试仪器采用中国计量科学研究院研制的型号为 CTM-6W 的磁通门磁强计，其磁场测量精度为 ± 1 nT，分辨率小于 1 nT。分别对单层加热膜和双层加热膜施加逐渐增大的直流驱动电流，电流值为 0 ~ 0.1 A。实验中，测试点位于距离

北航学报 赠阅

电加热膜 5 ~ 6 mm 附近的位置。

图 9(a)为根据单层加热膜测试数据拟合的一次函数曲线,可以计算出磁场变化率为 1 332.8 nT/A;图 9(b)为根据双层加热膜测试数据拟合的一次函数曲线,可以计算出磁场变化率为 32.4 nT/A,二者相差约 40 倍,因此采用双层结构的电加热膜的磁场噪声抑制效果较好。

采用双层对称结构的电加热膜,且加热信号被调制到高频,因此加热带来的磁场噪声比较微弱。为了更好的测试系统的噪声水平,采用间接方法来评估其噪声水平,在无自旋交换弛豫(Spin-Exchange Relaxation Free,SERF)原子自旋磁场测量装置进行测试,通过开关电源的方式进行对比。经过测试评估,该方案的电加热系统引入磁场噪声优于 17 fT/Hz$^{1/2}$,该指标的实现标志着本方案的技术水平已经达到了实验室采用仪器搭建的系统,为加热系统的工程化奠定了基础。



图 8 磁场屏蔽装置

Fig. 8 Magnetic field shielding device

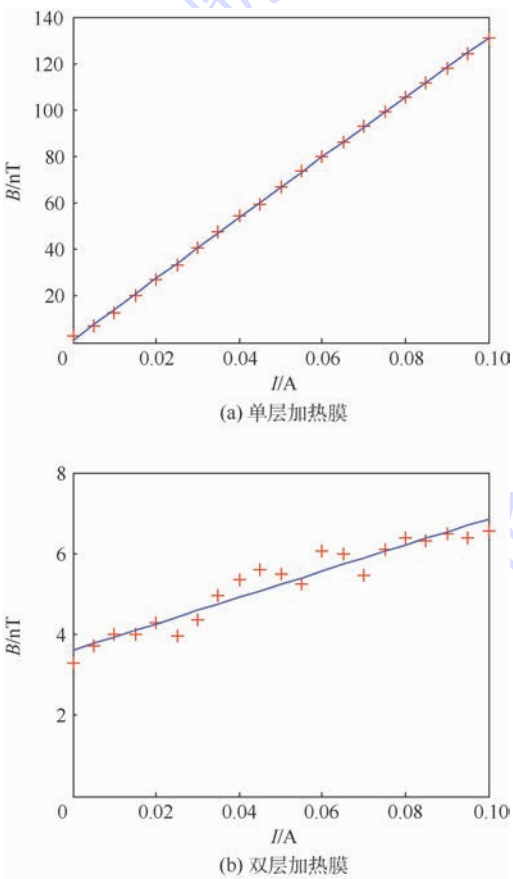


图 9 电加热膜磁场噪声抑制效果测试

Fig. 9 Test on magnetic field noise suppression performance of electric heating film

4.2 驱动系统性能测试实验

分别用高频正弦波和方波将碱金属气室加热到相同的温度。对加热驱动信号做 FFT 频谱分析,结果如图 10(a)所示,高频正弦波加热驱动信号明显只有主频信号,而方波加热驱动信号除了主频信号外,还有大量的奇次谐波信号以及倍频信号。对原子自旋陀螺仪的检测信号做 FFT 频谱分析,结果如图 10(b)所示,使用高频正弦波加热驱动信号相较于方波加热驱动信号,检测信号具有更低的噪声底线,因此可以达到更高的信噪比。

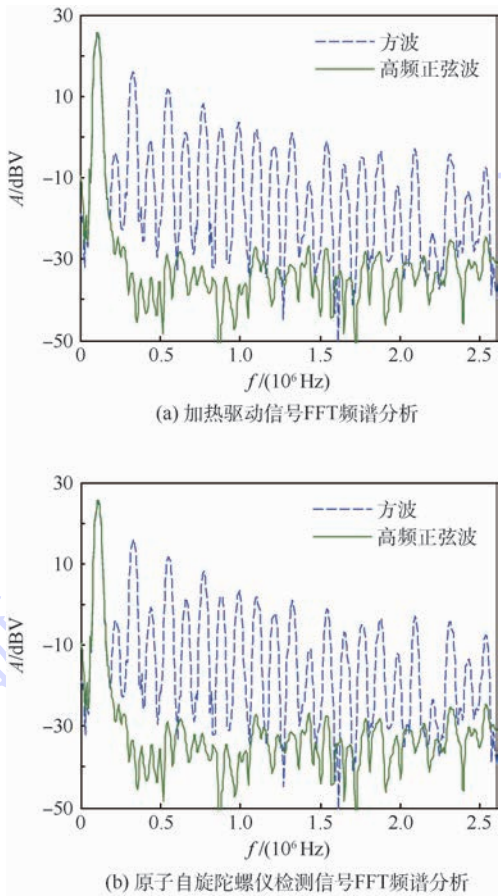


图 10 高频正弦波和方波加热驱动信号对比

Fig. 10 Comparison of heating driving signal between high-frequency sine wave and square wave

4.3 集成化无磁电加热系统温控性能测试实验

分别对核磁共振陀螺仪和 SERF 原子自旋陀螺仪进行测试。如图 11 所示,2 种陀螺仪均可以实现 ± 0.006 ℃ 的稳定度,这说明采用本方案基本可以达到仪器搭建的系统,满足高精度原子自

旋陀螺仪对加热技术小型化的要求，为下一步实现芯片化技术奠定了技术基础。

表 2 为无磁电加热膜在核磁共振陀螺仪和 SERF 原子自旋陀螺仪上进行的测试。可以看出，在 100 ℃和 200 ℃时的标准差都在 0.003 ℃以内，可以满足实验的需求。

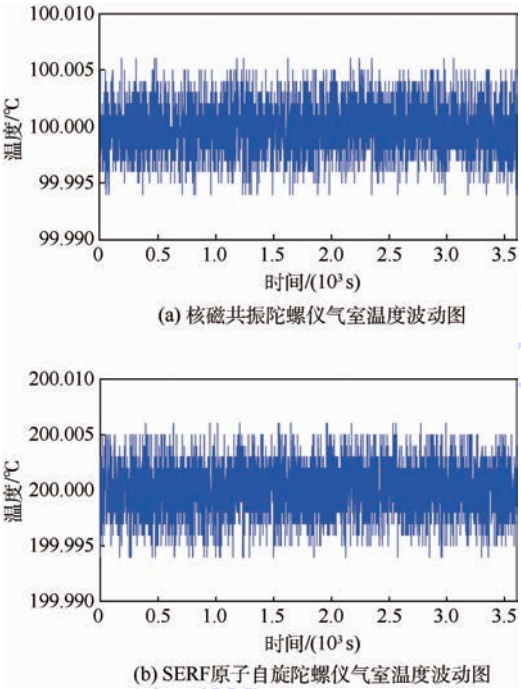


图 11　温控性能测试实验结果

Fig. 11　Experimental results of temperature control performance test

表 2　温控性能测试结果数据分析

Table 2　Data analysis of temperature control performance test results　℃

陀螺类型	设定值	平均值	偏差	标准差
核磁共振陀螺仪	100	99.999 887	$1.130\,6\times10^{-4}$	0.002 3
SERF 原子自旋陀螺仪	200	200.000 067	$-6.666\,7\times10^{-5}$	0.002 5

5　结　论

1）本文设计制作了具有磁场噪声抑制作用的双层对称结构电加热膜，与单层加热膜相比，磁场噪声抑制效果可提升 40 倍。

2）选用高频正弦波作为加热驱动信号，与普遍采用的方波加热驱动信号相比，显著降低了原子自旋陀螺仪检测信号的噪声底线，提高了检测信号的信噪比。

3）设计并搭建了碱金属气室集成化无磁电加热系统，测温数据采用复合数字滤波处理，温度控制采用分段式控制算法，实现加热引入的等效磁场噪声优于 17 $fT/Hz^{1/2}$，气室温度稳定度优于 ±0.006 ℃，为原子自旋陀螺仪的性能提升提供了可靠保障，满足高精度原子自旋陀螺仪对加热技术小型化的要求，为下一步实现芯片化技术奠定了技术基础。

参考文献（References）

[1] ZOU S, ZHANG H, CHEN X Y, et al. A novel calibration method research of the scale factor for the all-optical atomic spin inertial measurement device[J]. Journal of the Optical Society of Korea, 2015, 19(4): 415-420.

[2] FANG J C, CHEN Y, ZOU S, et al. Low frequency magnetic field suppression in an atomic spin co-magnetometer with a large electron magnetic field [J]. Journal of Physics B: Atomic, Molecular and Optical Physics, 2016, 49(6): 65006-1-65006-8.

[3] 万双爱. 基于 SERF 的原子自旋陀螺仪自旋耦合控制方法及实验研究[D]. 北京：北京航空航天大学，2014：3-5.
WAN S A. Study on the spin coupling control methods for atomic spin gyroscope based on SERF[D]. Beijing: Beihang University, 2014: 3-5 (in Chinese).

[4] DANG H B, MALOOF A C, ROMALIS M V. Ultrahigh sensitivity magnetic field and magnetization measurements with an atomic magnetometer[J]. Applied Physics Letters, 2010, 97(15): 151110-1-151110-3.

[5] LEDBETTER M P, SAVUKOV I M, ACOSTA V M, et al. Spin-exchange-relaxation-free magnetometry with Cs vapor[J]. Physical Review A, 2008, 77(3): 0334083-1-0334083-7.

[6] LU J X, QIAN Z, FANG J C, et al. Suppression of vapor cell temperature error for spin-exchange-relaxation-free magnetometer[J]. Review of Scientific Instruments, 2015, 86(8): 0831038-1-0831038-4.

[7] LIU G B, LI X F, SUN X P, et al. Ultralow field NMR spectrometer with an atomic magnetometer near room temperature[J]. Journal of Magnetic Resonance, 2013, 237(7): 158-163.

[8] 陆吉玺. SERF 原子磁强计噪声抑制方法与实验研究[D]. 北京：北京航空航天大学，2016：10-11.
LU J X. Methods and experimental study on noise suppression for atomic magnetometer in SERF regime[D]. Beijing: Beihang University, 2016: 10-11 (in Chinese).

[9] 袁起航，林贵平，李广超，等. 电脉冲除冰系统电磁脉冲力仿真分析[J]. 北京航空航天大学学报，2016，42(3)：633-638.
YUAN Q H, LIN G P, LI G C, et al. Simulation and analysis on electromagnetic impulse force of electro-impulse de-icing system [J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(3): 633-638 (in Chinese).

[10] BULATOWICZ M. Electrical resistive heaters for magnetically sensitive instruments[C]// 45th Annual Meeting of the APS Division of Atomic, Molecular and Optical Physics. Denver: American Physical Society Press, 2014, 59(8): 93.

[11] 陈鹏，葛红娟，倪一洋，等. 电脉冲除冰系统非线性等效电路分析[J]. 北京航空航天大学学报，2015，41(8)：1539-1545.

CHEN P, GE H J, NI Y Y, et al. Nonlinear equivalent circuit analysis of electro-impulse de-icing system[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(8): 1539-1545(in Chinese).

[12] 孙建辉,刘军涛,徐声伟,等.高精度微弱脑电检测数模混合控制芯片系统[J].仪器仪表学报,2016,37(5):1025-1032.
SUN J H, LIU J T, XU S W, et al. Weak brain-EEG signal high resolution detection based on hybrid analog/digital signal control ASIC system[J]. Chinese Journal of Scientific Instrument, 2016, 37(5): 1025-1032(in Chinese).

[13] 罗毅,施云波,渠立亮,等.谐波分析及交流比较的高空温度探测方法研究[J].仪器仪表学报,2014,35(4):721-728.
LUO Y, SHI Y B, QU L L, et al. Research on atmosphere aloft temperature probing based on harmonic analysis and AC comparison methods[J]. Chinese Journal of Scientific Instrument, 2014, 35(4): 721-728(in Chinese).

[14] 陶蓓,朱静,刘玉涛.高精度电阻测量方法及其应用[J].计量与测试技术,2011,38(10):35-36.
TAO B, ZHU J, LIU Y T. The method and application of the higher accuracy resistance measurement[J]. Metrology & Measurement Technique, 2011, 38(10): 35-36(in Chinese).

[15] 王亚刚,许晓鸣,邵惠鹤.基于 Ziegler-Nichols 频率响应方法的自适应 PID 控制[J].控制工程,2012,19(4):607-613.
WANG Y G, XU X M, SHAO H H. Adaptive PID controller based on Ziegler-Nichols frequency method[J]. Control Engineering of China, 2012, 19(4): 607-613(in Chinese).

作者简介:

周斌权 男,博士,讲师。主要研究方向:原子传感器技术。

Experimental study on electromagnetic noise suppression of atomic spin gyroscope heating chamber

ZHOU Binquan[1,2,*], HAO Jiepeng[1,2], LIANG Xiaoyang[1,2], QUAN Wei[1,2], LIU Gang[1,2]

(1. National Key Laboratory of Inertial Technology, Beijing University of Aeronautics and Astronautics, Beijing 100083, China;
2. School of Instrumentation Science and Opto-electronics Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: Atomic spin gyroscope is the latest type of gyroscope, which has ultra-high theoretical precision. Alkali vapor cell is the sensing element of atomic spin gyroscope, which carries the atomic spin effect. Electric heating makes alkali attain saturated vapor pressure, which will introduce electromagnetic interference and other noises, thereby affecting the accuracy and sensitivity of atomic spin gyroscope. To reduce the influence of heating chamber electromagnetic noise on the atomic spin gyroscope, the electromagnetic noise suppression experiment was studied from two aspects of heater structure and heating driving signal. A special shaped heating film with magnetic noise suppression was designed. A high frequency sine wave was designed as the heating driving signal. In addition, a non-magnetic heating system of alkali vapor cell was constructed. The test results show that the equivalent magnetic noise is within 17 $fT/Hz^{1/2}$, and the temperature stability of the alkali vapor cells is within ± 0.006℃, which provides a reliable guarantee for the performance improvement of atomic spin gyroscope.

Keywords: atomic spin gyroscope; non-magnetic heating; noise suppression; heating film; high frequency sine wave

Received: 2016-12-22; **Accepted**: 2017-01-13; **Published online**: 2017-03-08 09:29

URL: kns.cnki.net/kcms/detail/11.2625.V.20170308.0929.001.html

Foundation items: National High-tech Research and Development Program of China (2014AA123401); National Natural Science Foundation of China (61227902, 61374210)

* **Corresponding author**. E-mail: bqzhou@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2016.0980

实时单核和谐周期分区系统时间窗口分配算法

尚利宏[1,*], 谭特[1], 周密[2]
(1. 北京航空航天大学 计算机学院, 北京 100083; 2. 北京中航瑞博航空电子技术有限公司, 北京 100192)

摘 要: 目前航空电子系统正快速朝着综合模块化方向发展。为了防止同一计算平台上的应用相互干扰,IMA 软件普遍采用分区机制。由于时间分区的引入,传统的实时周期任务可调度性分析已经不再适用。为此研究了一类特殊的分区系统——和谐周期分区系统在单处理器下的可调度性。给出了和谐周期分区系统的形式化定义以及系统中任务可调度性的充分必要条件,并基于此提出了一种分区时间窗口分配算法。该算法为每个分区在主时间帧内分配多个时间窗口,并且保证只要和谐周期分区系统在理论上可调度,该算法就一定能生成一个可行的调度表,使得当全局调度器按照此调度表周期地调度分区时,各个分区中的任务不会超时。本文提出的算法可以运用在实际的工程中。

关 键 词: 分区; 实时; 和谐周期; 可调度性; 时间窗口
中图分类号: V247; TP316.2
文献标识码: A **文章编号**: 1001-5965(2018)01-0043-07

目前,航空电子系统正飞速从联合架构(federated avionics)向综合模块航电(Integrate Modular Avionics,IMA)架构转换[1]。IMA 架构在一台共享的计算平台上部署拥有多个航电子系统功能的高集成度分区系统,即各个分系统共享软件(操作系统)和硬件资源。这种架构已经广泛应用在实际飞机航电系统中,例如波音 777、空客 A380 等机型[2]。

由于在 IMA 架构下,多个应用共享计算资源,因此一个应用的错误可能会影响到其他应用。为了提高系统可靠性和安全性,IMA 系统软件必须提供应用之间的隔离机制[3]。

ARINC 653[4] 是一个分区操作系统标准,其通过引入分区来提供应用在时间和空间上的隔离。在空间上,任何应用只能访问预先分配给它的物理内存,而不能访问分配给其他分区的物理内存;在时间上,任何应用只能在预先分配给它的时间窗口内执行,而不能占用其他应用的时间窗口。

ARINC 653 调度策略包括 2 级。分区间采用时间片轮转调度算法。用户需定义在一个主时间帧内各个分区的调度表,操作系统以主时间帧为周期,根据调度表中为每个分区分配的时间窗口调度分区,应当注意分区没有优先级[5]。分区内的实时任务可以选用固定优先级抢占调度算法。

由于时间分区的引入,传统的响应时间可调度性分析方法[6] 不再适用。目前,已经有很多文献基于传统的方法,对分区系统的可调度性进行了研究。文献[7]对分区进行了建模,提出了关键分区(critical partition)的概念,并给出了多时间窗口分区中任务可调度的充分条件。文献[8]给出了多时间窗口分区任务可调度的充分必要条件,但算法复杂度要高得多。除了理论计算,还可以通过建模工具进行可调度性分析。例如,文献[9]给出了一种基于时序 Petri 网的分析方法,不但可以分析可调度性,而且可以分析时序约束的满足性。文献[10]综述了分区系统的可调度

收稿日期: 2016-12-29; **录用日期**: 2017-03-24; **网络出版时间**: 2017-04-27 16:30
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170427.1630.002.html
* **通信作者**. E-mail: shanglh@buaa.edu.cn

引用格式: *尚利宏, 谭特, 周密. 实时单核和谐周期分区系统时间窗口分配算法[J]. 北京航空航天大学学报, 2018, 44(1): 43-49.*
SHANG L H, TAN T, ZHOU M. Time windows distribution algorithm for real-time harmonic-period partition system on uniprocessor platform[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(1): 43-49 (in Chinese).

北航学报 赠阅

性分析方法。

除了可调度性研究,分区时间窗口分配也是当前研究热点。文献[11]研究了当各个分区采用 RM 静态调度算法,且各个分区在主时间帧内只被分配一个时间窗口时,分区任务的最大利用率 U 与分区时间窗口和主时间帧的比值 π_Ω(称为分区参数)之间的关系式,并以此来设计分区参数。文献[12]给出了当各个分区采用任意优先级调度算法时,分区参数的设计方法。这些方法的主要思想是将其他分区视为本分区的一个最高优先级的任务。

文献[13]将一个分区的周期调度抽象为服务,并引入处理器降速因子来模拟分区中的时间流逝,在此基础上提出了一种分区参数的设计方法。文献[14]研究了服务模型下任务可调度的充分条件。

上述文献提出的算法只为分区在主时间帧中分配一个时间窗口。事实上,为满足某些系统的可调度性,必须要在主时间帧内为某些分区分配多个时间窗口。目前还没有太多文献给出多时间窗口的分配算法。本文研究在单处理器平台下满足特定条件(和谐周期[15]条件)的分区系统的多时间窗口分配算法。与已有算法相比,本文提出的算法能够为任何在理论上可调度的和谐周期分区系统生成可行的分区调度表,因此更具有实用性。

1 调度模型定义

本文研究分区系统的调度。在分区系统中,分区间采用时间片轮转的调度算法,而分区内部的任务采用可抢占的固定优先级调度算法。全局调度器周期地按照预先定义好的调度表调度分区占用计算资源。当分区被调度运行后,首先执行优先级最高的任务,并且就绪的高优先级任务可以抢占正在执行的低优先级任务。另外,规定分区中的所有任务都是周期任务。

定义 1 系统 S 定义为分区 P_k 的集合,即 $S=\{P_k\}$。

定义 2 分区 P_k 定义为周期任务 τ 的集合,即 $P_k=\{\tau\}$。

定义 3 周期任务 τ 定义为四元组,即 $\tau=(p,d,e,\mathrm{pri})$,$p$ 为任务的周期,d 为任务的截止时间,e 为任务的最坏执行时间,pri 为任务的优先级。

值得注意的是,任务参数 e 是最坏执行时间。文献[16]从执行时间概率的角度提出了一种优化任务响应时间的分区调度算法。

定义 4 工作实例 J 表示任务在某个特定周期内的实际执行。

在下文中,用 $\tau.x$ 表示元组 τ 中的分量 x。另外,用时间片表示任务执行时间的单位,例如 $\tau.e=1$ 表示任务 τ 的一次执行需要 1 个时间片。规定任务 τ 的最坏执行时间 $\tau.e$ 是整数。

2 和谐周期分区系统及其可调度性

定义 5(和谐周期[15]任务集) 假设集合 T 是任务的集合,如果 T 中的任何任务 $\tau\in T$ 的截止时间 $\tau.d$ 和周期 $\tau.p$ 相等,并且对于任何 $\tau_i,\tau_j\in T$,如果 $\tau_i.p<\tau_j.p$,则有 $\tau_j.p$ 是 $\tau_i.p$ 的倍数,那么称 T 是和谐周期任务集。

定义 6(和谐周期分区系统) 假设系统 $S=\{P_k\}$,其所有分区 $P_k=\{\tau\}$ 中的所有任务构成集合 $\mathrm{ST}=\bigcup_{P_k\in S}P_k$。如果 ST 是和谐周期任务集,则称系统 S 是和谐周期分区系统。

下面讨论和谐周期分区系统各个分区中任务的可调度性。假设和谐周期分区系统 $S=\{P_k\}$,其所有分区中的所有任务构成的任务集为 $\mathrm{ST}=\bigcup_{P_k\in S}P_k$。用集合 period 表示 ST 中所有任务的所有可能的周期的集合,即

$$\mathrm{period}=\{p_i\mid\exists\tau(\tau\in\mathrm{ST}\wedge\tau.p=p_i)\}$$

period 中的周期按照从小到大的顺序排序,其中 p_1 为最小周期,$p_{\max}$ 为最大周期,p_i 表示第 i 小的周期。

由于 $p_{\max}$ 是所有 ST 中所有任务的周期的最小公倍数,因此,可将主时间帧的长度设置为 $p_{\max}$[8]。

定义 7 集合 A_{ki} 表示分区 P_k 中所有周期小于等于 p_i 的任务的集合,即 $A_{ki}=\{\tau\in P_k\mid\tau.p\leqslant p_i\}$。

定义 8 集合 B_{ki} 形式化地定义为:$B_{ki}=\{\tau\in P_k\mid\tau.p>p_i\wedge\exists\tau'(\tau'\in A_{ki}\wedge\tau'.\mathrm{pri}<\tau.\mathrm{pri}\}$,即表示 P_k 中所有周期大于 p_i 且优先级高于 A_{ki} 中任务最低优先级的任务的集合。

引理 1 对于任何分区 P_k,周期 $p_i\in\mathrm{period}$ 以及 $l=0,1,\cdots,\frac{p_{\max}}{p_i}-1$,如果任务 $\tau\in B_{ki}$ 的周期 $\tau.p$ 满足 $lp_i\bmod\tau.p\neq0$,即 τ 中没有任何工作实例在 lp_i 时刻释放,则任务 $\tau\in B_{ki}$ 的任何工作实例不在 $[lp_i,(l+1)p_i)$ 间隔内执行。

证明 由于任务 τ 的周期大于 p_i,所以在 $[lp_i,(l+1)p_i)$ 间隔内最多只可能执行 τ 的一个

北航学报 赠阅

工作实例 J,并且 J 在 lp_i 之前释放,截止时间在 lp_i 时刻之后(见图 1)。假设 J 的释放时间为 $m\tau.p$,由于 $\tau.p$ 是 p_i 的整倍数,所以 $m\tau.p$ 一定可以写成 $l'p_i, l'<l$。由于在时间间隔$[l'p_i,(l'+1)p_i)$内,在 $l'p_i$ 时刻释放的 A_{ki} 中的任务全部执行完成。而 τ 的任务的优先级比 A_{ki} 中任务的最低优先级高,因此,在 $l'p_i$ 时刻释放的 A_{ki} 中的任务全部执行完成之前,工作实例 J 一定执行完成,即 J 在$(l'+1)p_i$ 时刻之前完成。由于$(l'+1)p_i \leqslant lp_i$,所以 J 一定在$[lp_i,(l+1)p_i)$开始前完成,并且 τ 不会在$(l+1)p_i$ 时刻前产生新的工作实例。因此,τ 不会有任何工作实例在区间$[lp_i,(l+1)p_i)$内执行。

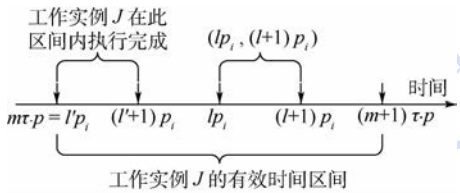


图 1　引理 1 证明示意图

Fig. 1　Schematic for proving Lemma 1

现在考虑在满足可调度性的前提下,任何一个分区 P_k 在时间区间$[lp_i,(l+1)p_i)$内应当被分配到的时间片数。在该区间内,任何周期不大于 p_i 的任务 $\tau \in A_{ki}$,都会产生 $p_i/\tau.p$ 个工作实例,并且第 1 个工作实例的释放时刻为 lp_i,最后 1 个工作实例的截止时刻为$(l+1)p_i$。因此,在该区间内分区 P_k 得到的时间片应满足这些工作实例的执行时间需求。其次,对于周期大于 p_i 的任务 $\tau \in B_{ki}$,如果其某个工作实例在时刻 lp_i 释放,由于其优先级高于 A_{ki} 中任务的最低优先级,那么,当 A_{ki} 中的任务产生的工作实例完成之前,此高优先级任务的实例一定被执行完成。因此,分区 P_k 在$[lp_i,(l+1)p_i)$间隔内得到的时间片还要满足在 lp_i 时刻释放的 B_{ki} 中任务的工作实例的执行时间需求。而对于 B_{ki} 中不在 lp_i 时刻释放的任务,根据引理 1,它的任何工作实例一定不会在$[lp_i,(l+1)p_i)$间隔内执行,因此,无需考虑这些任务。

综上所述,可以得到如下定理。

定理 1　如果系统 $S=\{P_k\}$ 是和谐周期分区系统,函数 $f(k,i,j)$ 表示分区 P_k 在时间区间$[i,j)$内分配得到的时间片数量,则分区 P_k 中任务可调度的充分必要条件是对任何 $p_i \in \text{period}$,对于任何 $l=0,1,\cdots,\dfrac{p_{\max}}{p_i}-1$,都有

$$f(k,lp_i,(l+1)p_i) \geqslant \sum_{\tau\in A_{ki}} \tau.e\frac{p_i}{\tau.p} + \sum_{\tau\in B_{ki}\wedge lp_i \bmod \tau.p=0} \tau.e = \sum_{\tau\in C}\tau.e\left\lceil\frac{p_i}{\tau.p}\right\rceil \tag{1}$$

式中:集合 $C=\{\tau \mid \tau\in A_{ki}\cup B_{ki}\wedge lp_i \bmod \tau.p=0\}$,即 A_{ki} 中所有任务,以及 B_{ki} 中所有在 lp_i 时刻有新的工作实例被释放的任务构成的集合。

证明　根据上述分析,定理 1 的必要性是显然的。下面用反证法证明其充分性。如果式(1)成立,假设某个以 p_i 为周期的任务 τ 的工作实例 J 超时,其中 J 的释放时间是 lp_i,截止时间是 $(l+1)p_i$,则一定有

$$f(k,lp_i,(l+1)p_i) < \tau.e + \sum_{\tau'\in D}\tau'.e\frac{p_i}{\tau'.p} + \sum_{\tau'\in E}\tau'.e = \sum_{\tau'\in D\cup E\cup\{\tau\}}\tau'.e\left\lceil\frac{p_i}{\tau'.p}\right\rceil \tag{2}$$

式中:集合 D 为所有周期不比 $\tau.p$ 大、优先级比 $\tau.\text{pri}$ 高的任务的集合;集合 E 为所有周期比 $\tau.p$ 大、优先级比 $\tau.\text{pri}$ 高且有工作实例在 lp_i 时刻释放的任务构成的集合。显然有集合 C 包含 $D\cup E\cup\{\tau\}$,这样就有

$$f(k,lp_i,(l+1)p_i) < \sum_{\tau'\in D\cup E\cup\{\tau\}}\tau'.e\left\lceil\frac{p_i}{\tau'.p}\right\rceil \leqslant \sum_{\tau'\in C}\tau'.e\left\lceil\frac{p_i}{\tau'.p}\right\rceil \tag{3}$$

这与条件相矛盾,假设不成立,即没有任何任务的任何工作实例超时。因此,分区是可调度的,原命题得证。

为了下文叙述方便,定义:

$$\text{demand}(k,l,i) = \sum_{\tau\in A_{ki}}\tau.e\frac{p_i}{\tau.p} + \sum_{\tau\in B_{ki}\wedge lp_i \bmod \tau.p=0}\tau.e \tag{4}$$

即 $\text{demand}(k,l,i)$ 表示为满足任务可调度性,在时间间隔$[lp_i,(l+1)p_i)$内分区 P_k 最少得到的时间片数。

3　时间窗口分配算法

时间窗口分配算法在实质上是将主时间帧中的时间片分配给各个分区。假设主时间帧的长度为 F,即其中共有 F 个时间片。用数组 $\text{distribute}[0\cdots F-1]$ 表示时间片的分配情况:$\text{distribute}[n]=k$ 表示从 n 时刻开始的时间片被分配给分区 P_k;$\text{distribute}[n]=-1$ 表示该时间片尚未被分配给任何分区。

算法 distribute_operator 在给定时间间隔$[s,e)$内为分区 P_k 分配指定数量的时间片。它从

北航学报 赠阅

distribute 数组中的 s 位置开始寻找空闲时间片(即语句 distribute[i] = -1),并将此空闲时间片分配给分区 P_k(语句 distribute[i] = k)。

输入:分区号 k,时间间隔[s,e)和时间片数量 m。

输出:是否分配成功标识。

```
begin
    for i from s to e - 1
      if  distribute[i] = -1 then
          distribute[i] = k, m = m - 1
          if m = 0      return success
      end if
    end for
    return failure
end
```

从定理 1 可以看出,为满足某个分区 P_k 中任务的可调度性,必须为它在任意的时间间隔[lp_i,$(l+1)p_i$),$p_i \in$ period,$l = 0,1,\cdots,\frac{p_{\max}}{p_i}$ 内分配 demand(k,l,i)个时间片以满足其中任务的执行时间需求。由于时间间隔[lp_i,$(l+1)p_i$)($i>1$)内包括 p_i/p_{i-1} 个长度为 p_{i-1} 的时间间隔,当这些长度为 p_{i-1} 的时间间隔内的时间需求都得到满足时,还需要为该分区在[lp_i,$(l+1)p_i$)间隔内分配的时间片数 additionT 为

$$\text{additionT}(k,l,i) = \text{demand}(k,l,i) - \sum_{h=0}^{\frac{p_i}{p_{i-1}}-1} \text{demand}\left(k,h + l\frac{p_i}{p_{i-1}},i-1\right) \tag{5}$$

而对于时间间隔[lp_1,$(l+1)p_1$),分区在此间隔内必须至少得到 demand(k,l,1)个时间片。因此,可以考虑按照 $p_1,p_2,\cdots,p_{\max}$ 的顺序,在[lp_i,$(l+1)p_i$)间隔内为各个分区分配满足要求的最少数量的时间片。具体的分配算法如算法 Window_Distribute_algorithm 所示。

输入:分区集$\{P_k\}$。

输出:分配成功或失败的标识。

```
begin
    for i from 1 to max
      for j from 0 to p_max/p_i - 1
        for k from 1 to partitionNum
          if i = 1 then
            m = demand(k,j,1)
          else
            m = additionT(k,j,i)
          end if
          ret = distribute_operator(k, jp_i, (j + 1)p_i, m)
          if ret = failure then
          return failure
        end for 3
      end for 2
    end for 1
    return success
end
```

当算法 Window_Distribute_algorithm 返回 success 后,数组 distribute 中即存储了时间片的分配方案,也就是时间窗口的分配方案。

如果在某次迭代中调用的 distribute_operator 返回失败,表明在此时间区间内无法为所有任务都分配满足式(1)的时间片数,也就表明此分区系统是不可调度的,因此算法返回失败。

下面讨论 Window_Distribute_algorithm 算法的时间复杂度。假设和谐周期分区系统 $S=\{P_k\}$ 共包括 n 个分区,每个分区最多有 m 个任务,并且所有任务的最大周期为 $p_{\max}$。首先,根据 demand 函数的定义,计算一次 demand 函数要遍历分区中的所有任务。因此,计算 demand 函数的时间复杂度是 $O(m)$。而 additionT(k,j,i) 函数的计算涉及一次 demand(k,j,i)的计算,以及 p_i/p_{i-1} 个 demand(k,l,$i-1$)值的求和。如果上一次迭代计算得到的 demand(k,l,$i-1$)保存在数组中,则由于 p_i/p_{i-1} 的最大值是所有任务的最大周期 $p_{\max}$,则 additionT(k,j,i)的计算复杂度就是 $O(m+p_{\max})$。根据最大周期 p 可得系统中最多可能有 lb p 种不同的周期,这意味着算法最外层循环最多迭代 lb p 次。易知,第 2 层和第 3 层循环最多执行 $p_{\max}$ 次和 n 次。因此,Window_Distribute_algorithm 算法的时间复杂度是 $O((m+p_{\max})p_{\max}n\text{lb}\ p_{\max})$。

4 案例分析

本节通过一个例子来演示第 3 节算法的执行过程。假设某 IMA 系统包括 2 个应用:液压应用和环控应用。这 2 个应用分别位于分区 A 和分区 B。液压应用包括 2 个任务:健康管理任务 task1 和主控任务 task2。环控应用只包括主控任务 task3。task1、task2 和 task3 的周期、最坏执行时间、截止时间和优先级如表 1 所示。由于系统中所有任务的截止时间和周期相等,且任何大的任务周期是任何小的任务周期的倍数,因此,该系

统是和谐周期分区系统。

由表 1 可见,由于最大的任务周期是 8,因此将主时间帧设置为 8。现在需要将这 8 个时间片分配给 A、B 2 个分区。从表 1 中可见,最小周期是 2。根据算法,首先应在长度为 2 的时间间隔[0,2),[2,4),[4,6),[6,8)内根据任务时间需求为这2 个分区分配时间片。在分区 A 内,task2 的周期为 2。由于 0 时刻 task2 与比其优先级高的 task1 同时就绪,因此 demand(A,0,1) = 1 + 2/2 × 1 = 2。在剩余的 3 个长度为 2 的区间内,demand(A,l,1) = 0,l = 1,2,3,即分区 A 应当在这 3 个区间内分别得到 1 个时间片。分区 B 中没有周期为 2 的任务,所以 demand(B,l,1) = 0,l = 0,1,2,3,因此本轮不为分区 B 分配时间片。本轮分配完成后的结果如图 2 所示。

表 1　用于案例分析的分区系统的任务参数

Table 1　Parameters of tasks of partitioning systems for case analysis

任　务	分　区	周　期	截止时间	最坏执行时间	优 先 级
task1	A	8	8	1	H
task2	A	2	2	1	L
task3	B	4	4	1	H

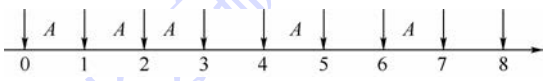


图 2　第 1 轮分配结果

Fig. 2　Distribution results of the first iteration

然后应当在长度为 4 的时间区间[0,4),[4,8)内为 2 个分区分配时间片。对于分区 A,demand(A,0,2) = 3,demand(A,1,2) = 2,所以有additionT(A,0,2) = 3 − 2 − 1 = 0,additionT(A,1,2) = 2 − 1 − 1 = 0,故在这 2 个区间内不为分区 A 分配额外时间片。在分区 B 中,task3 的周期为 4,可得demand(B,0,2) = demand(B,1,2) = 1,而任何 demand(B,0,1)都等于 0,因此 additionT(B,0,2) = additionT(B,1,2) = 1 − (0 + 0) = 1,即需要在[0,4)和[4,8)区间内分别为分区 B 分配一个时间片。本轮迭代分配结果如图 3 所示。

最后考虑长度为 8 的时间间隔。对于分区 A,demand(A,0,3) = 5,所以 additionT(A,0,3) = 5 − 3 − 2 = 0,因此,无需再为分区 A 分配额外的时间片;对于分区 B,demand(B,0,3) = 2,additionT(B,0,3) = 2 − 1 − 1 = 0,即也无需为分区 B 分配额外时间片。图 3 所示的分配结果即是最终的分配结果。根据分配结果生成的时间窗口布局如图 4所示。分区调度器可以通过此调度方案周期地调度分区,并且保证任何分区中的任何周期任务都不会超时。

从图 4 可以看出,在一个主时间帧内将发生 5 次分区切换。而如果将从 7 时刻开始的空闲时间片分配给分区 A,分区切换将只发生 4 次。因此,分区调度表中需要预先设置的条目至少是 4 条。根据此调度方案调度分区,各个分区中任务的工作实例的延迟情况如表 2 所示。

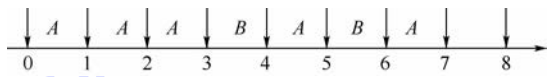


图 3　第 2 轮分配结果

Fig. 3　Distribution results of the second iteration



图 4　主时间帧内分区时间窗口布局

Fig. 4　Layout of partition time windows in main time frame

表 2　任务的工作实例的响应时间

Table 2　Response time of jobs instances of tasks

任　务	最大响应时间	最小响应时间	平均响应时间
task1	0	0	0
task2	1	0	0.25
task3	3	1	2

从表 2 可见,task3 的工作实例的响应时间较长。可以将在 4 时刻开始执行的 task1 的第 2 个实例和 5 时刻开始执行的 task2 的第 2 个实例的执行顺序交换,以降低 task2 的响应时间。这虽然增长了 task1 的响应时间,但却可以提高整个系统的性能。本文算法目前无法实现这样的优化。

5　与其他算法的比较

本节对本文算法和已有算法进行比较。引言已经提到,文献[11]给出了一种分区系统时间窗口分配算法,因此,本文主要与文献[11]中的算法进行比较。

文献[11]中的算法适合于分区内应用 RM 静态调度算法的分区系统。在这种算法下,每个分区在主时间帧内只被分配一个时间窗口,并且每个分区 P_k 在主时间帧内被分配的时间窗口长度占整个主时间帧长度的比例为

$$\alpha_k = 2 - 2(U_k/n_k + 1)^{-n_k} \tag{6}$$

式中:n_k 为分区 P_k 中的任务总数;U_k 为分区 P_k 中所有任务利用率的和,即

$$U_k = \sum_{\tau \in P_k} \frac{\tau.e}{\tau.p} \tag{7}$$

北航学报 赠阅

如果某个分区系统中所有分区的时间窗口占比 α_k 之和小于 1，此算法就可以成功为此分区系统生成调度表；相反，如果所有分区的 α_k 之和大于 1，此算法就对此分区系统失效。

现在考虑如表 3 所示的分区系统，此系统包括 2 个分区。其中，分区 B 中只包括一个周期为 4 的任务；分区 A 包括 2 个任务：周期为 8 的 task1 和周期为 2 的 task2。根据 RM 静态调度算法，task2 的优先级高于 task1 的优先级。

对于此系统，如果应用文献[11]的算法，根据式(6)可得，$\alpha_A = 0.839$，$\alpha_B = 0.4$。由于 $\alpha_A + \alpha_B > 1$，因此此算法失效。

而如果应用本文算法，将生成如图 5 所示的调度表。

表 3 用于算法对比的分区系统的任务参数

Table 3 Parameters of tasks of partitioning systems for algorithms comparison

任务	分区	周期	截止时间	最坏执行时间	优先级
task1	A	8	8	1	L
task2	A	2	2	1	H
task3	B	4	4	1	H

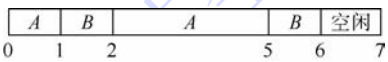


图 5 本文算法生成的调度表

Fig. 5 Schedule table of proposed algorithm

事实上，正如文献[11]已经指出的，当系统的总利用率大于 50% 后，本文算法的成功率将大幅下降。而本文算法只要保证系统在理论上可调度，就一定能够生成可行的调度表；并且文献[11]中的算法只适合于分区内部采用 RM 静态调度算法的系统，而本文算法适合于分区内部应用任何固定优先级调度算法的系统。当然，本文算法和文献[11]中算法相比，最大的局限是只适用于满足和谐周期条件的分区系统，对于不满足和谐周期条件的分区系统，本文算法不再适用。

6 结 论

本文以 IMA 系统为背景，基于传统的响应时间分析方法，研究了满足和谐周期条件的分区系统中各个分区中任务可调度的充分必要条件，并以此为基础给出了一种分区时间窗口分配算法。对于任何和谐周期分区系统，如果该系统在理论上存在一种可行的调度方案，这种算法就一定能够生成一种可行的调度方案。反之，如果该算法返回失败，就表明该和谐周期分区系统在理论上是不可调度的。

然而本文算法也存在一些不足。最大的局限就是要求分区系统必须满足和谐周期条件。另外，还存在分区切换次数过多、任务响应时间过长等问题。

虽然存在上述问题，本文给出的算法依然可以用于实际工程中的分区调度表设计。后续的工作将集中在优化分区切换次数及任务响应时间上。

参考文献（References）

[1] WATKINS C B, WALTER R. Transitioning from federated avionics architectures to integrated modular avionics[C]// Digital Avionics Systems Conference, 2007. Piscataway, NJ: IEEE Press, 2007: 2. A. 1-1-2. A. 1-10.

[2] 周强，熊华钢. 新一代民机航空电子互连技术发展[J]. 电光与控制，2009，16(4)：1-6.
ZHOU Q, XIONG H G. Development of the new generation civil avionic interconnection technology[J]. Electronics Optics & Control, 2009, 16(4): 1-6(in Chinese).

[3] 沈玉龙，崔西宁，马建峰，等. 综合化航空电子系统可信软件技术[J]. 航空学报，2009，30(5)：938-945.
SHEN Y L, CUI X N, MA J F, et al. Trust software technology in integrated avionics systems[J]. Acta Aeronautica et Astronautica Sinica, 2009, 30(5): 938-945(in Chinese).

[4] ARINC. avionics application software standards interface: ARINC 653[S]. Annapolis: AEEC, 2003.

[5] 杨霞，桑楠，雷剑，等. 嵌入式高可信架构中基于静态模型的调度研究[J]. 航空学报，2009，30(12)：2387-2394.
YANG X, SANG N, LEI J, et al. Scheduling based on static model in trusted architecture for embedded systems[J]. Acta Aeronautica et Astronautica Sinica, 2009, 30(12): 2387-2394(in Chinese).

[6] LIU J W S. Real-time sytem[M]. Upper Saddle River: Prentice Hall, 2003: 130-140.

[7] MOK A K, FENG X, CHEN D. Resource partition for real-time systems[C]// 7th IEEE Proceedings of Real-time Technology and Applications Symposium. Piscataway, NJ: IEEE Press, 2001: 75-84.

[8] 谭龙华，杜承烈，雷鑫. ARINC 653 分区实时系统的可调度分析[J]. 航空学报，2015，36(11)：3698-3705.
TAN L H, DU C L, LEI X. Schedulability analysis for ARINC 653 partitioned real-time systems[J]. Acta Aeronautica et Astronautica Sinica, 2015, 36(11): 3698-3705(in Chinese).

[9] CARNEVALI L, PINZUTI A, VICARIO E. Compositional verification for hierarchical scheduling of real-time systems[J]. IEEE Transactions on Software Engineering, 2013, 39(5): 638-657.

[10] 陈平，魏峰，李蜀瑜. ARINC 653 调度算法研究[J]. 现代电子技术，2015，38(15)：29-32.
CHEN P, WEI F, LI S Y. Study on ARINC 653 scheduling algorithm[J]. Modern Electronics Technique, 2015, 38(15): 29-32

(in Chinese).

[11] 李昕颖,顾健,何锋,等. 硬实时系统在强分区约束下的双层分区调度[J]. 计算机学报,2010,33(6):1032-1039.
LI X Y, GU J, HE F, et al. Two-level partition scheduling in hard real time system under strong partition constraints[J]. Chinese Journal of Computers, 2010, 33(6): 1032-1039 (in Chinese).

[12] 周天然,熊华钢. 航空电子系统混合实时任务的双层调度[J]. 航空学报,2011, 32(6):1067-1074.
ZHOU T R, XIONG H G. Two-level hierarchical scheduling for hybrid real-time tasks in avionics systems[J]. Acta Aeronautica et Astronautica Sinica, 2011, 32(6): 1067-1074 (in Chinese).

[13] LIPARI G, BINI E, NGUYEN C, et al. A methodology for designing hierarchical scheduling system[J]. Journal of Embedded Computing—Real-time System, 2005, 1(2): 257-269.

[14] WAN M, TIAN S. Research on schedulability of partition scheduling for IMA[C]//2011 4th International Symposium on Computational Intellingence and Design. Piscataway, NJ: IEEE Press, 2011, 2: 322-325.

[15] NASRI M, FOHLER G. An efficient method for assigning harmonic periods to hard real-time tasks with period ranges[C]// 2015 27th Euromicro Conference on Real-Time Systems. Piscataway, NJ: IEEE Press, 2015: 149-159.

[16] 桥乃强,徐涛,谷青范. ARINC 653 分区调度算法的研究与改进[J]. 计算机工程,2011,37(20):249-251.
QIAO N Q, XU T, GU Q F. Research and improvement of ARINC 653 partition schedule algorithm[J]. Computer Engineering, 2011, 37(20): 249-251 (in Chinese).

作者简介:

尚利宏　男,博士,副教授,硕士生导师。主要研究方向:嵌入式系统、容错技术。

谭特　男,硕士研究生。主要研究方向:嵌入式系统。

周密　男,博士。主要研究方向:嵌入式系统、容错技术。

Time windows distribution algorithm for real-time harmonic-period partition system on uniprocessor platform

SHANG Lihong[1,*], TAN Te[1], ZHOU Mi[2]

(1. School of Computer Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China;
2. Reliable Avionics Technology Co., Ltd., Beijing 100192, China)

Abstract: Recently the avionics system is quickly transferring to integrated modular architecture. To prevent the mutual interference between different applications, IMA software usually adopts partition mechanism. Due to the "time partition", the traditional real-time schedulability analytical method is not applicable. This paper researches a class of special partition system, which is called harmonic-period partition system on uniprocessor platform. This paper gives the formalized definition of harmonic-period partition system and the necessary and sufficient condition of schedulability of tasks in harmonic-period partition system. On this basis, this paper proposes an algorithm, which is called time windows distribution algorithm. This algorithm distributes multiple time windows for each partition in the main time frame. This algorithm must be able to find a feasible schedule table for a harmonic-period partition system if this system is schedulable theoretically, and all tasks in partitions will not timeout if the global scheduler schedules partitions according to this schedule table. The algorithm proposed in this paper can be applied to practical engineering.

Keywords: partition; real-time; harmonic-period; schedulability; time window

Received: 2016-12-29; **Accepted**: 2017-03-24; **Published online**: 2017-04-27 16:30
URL: kns.cnki.net/kcms/detail/11.2625.V.20170427.1630.002.html

* **Corresponding author**. E-mail: shanglh@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0033

共轴刚性双旋翼非定常气动干扰载荷分析

谭剑锋[1,*]，孙义鸣[1]，王浩文[2]，林长亮[3]
（1. 南京工业大学 机械与动力工程学院，南京 211816； 2. 清华大学 航天航空学院，北京 100084；
3. 航空工业哈尔滨飞机工业集团有限责任公司 飞机设计研究所，哈尔滨 150066）

摘 要：共轴刚性双旋翼系统提高直升机最大前飞速度，但旋翼振动载荷明显增大。为研究高速共轴刚性双旋翼系统振动载荷特性，须首先分析共轴刚性双旋翼气动干扰下的非定常气动载荷。基于非定常面元法建立满足桨叶前缘和后缘边界条件的旋翼反流区气动模型以体现高速共轴刚性双旋翼后行边反流区影响，且增加共轴刚性双旋翼桨尖涡-桨叶气动干扰模型以体现共轴刚性双旋翼非定常气动干扰影响，并结合基于黏性涡粒子法的共轴刚性双旋翼尾迹模型，构建高速共轴刚性双旋翼气动干扰下的气动载荷分析方法。通过计算前飞状态下的 X2 共轴刚性双旋翼特征剖面非定常气动载荷时间历程，并与 PRASADUM 以及基于 NASA OVERFLOW 和 CREATE AV Helios 的 CFD/CSD 计算结果对比，验证本文共轴刚性双旋翼非定常气动载荷分析方法的有效性。相比于 PRASADUM，本文分析更好地体现上、下旋翼在前行边和后行边非定常气动载荷的变化特性，并与 CFD/CSD 计算结果更吻合。分析 X2 上、下旋翼气动干扰对共轴刚性双旋翼桨叶非定常气动载荷的影响，以及单旋翼与共轴刚性双旋翼非定常气动载荷差异。分析表明，低速状态下的共轴刚性双旋翼非定常气动载荷受双旋翼桨尖涡干扰显著，而高速前飞状态受双旋翼桨叶干扰明显，且表现出桨叶片数整数倍的辐射状干扰特征。

关 键 词：共轴刚性双旋翼；非定常气动载荷；反流区；面元法；黏性涡粒子法
中图分类号：V211.52
文献标识码：A **文章编号**：1001-5965(2018)01-0050-13

高速直升机 X2 的巡航速度高达 463 km/h，是美国现役“黑鹰 UH-60A”直升机的 2 倍、“阿帕奇 AH-64D”直升机的 1.5 倍，而实现高速前飞的关键技术不仅在于尾部加装了推进螺旋桨，而且更重要的是使用了区别于常规单旋翼的共轴刚性双旋翼系统。与常规直升机单旋翼相比，共轴刚性双旋翼不仅显著提高高前进比的拉力，且去除用于平衡反扭矩的尾桨功率损耗，因而提高最大前飞速度。虽然共轴刚性双旋翼系统具有高速前飞所需具备的特征和潜力，但存在大速度前飞下旋翼振动载荷较大，且振动水平超出相同尺寸的铰接式旋翼问题[1]，而共轴刚性双旋翼气动干扰下的非定常气动载荷是振动载荷的主要来源。因此，为研究高速共轴刚性双旋翼系统振动载荷特性，须首先分析共轴刚性双旋翼气动干扰下的非定常气动载荷。

旋翼自由尾迹方法简单高效，适用于桨叶气动载荷分析[2-3]。通过建立共轴刚性双旋翼自由尾迹模型，Yeo 和 Johnson[4] 于 2013 年采用 CAMRAD Ⅱ（Comprehensive Analytical Model of Rotor-

收稿日期：2017-01-18；**录用日期**：2017-08-11；**网络出版时间**：2017-09-08 16:14
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170908.1614.001.html
基金项目：国家自然科学基金（11502105）；江苏省自然科学基金（BK20161537）；江苏省高校自然科学研究面上项目（15KJB130004）
* **通信作者**. E-mail: Jianfengtan@njtech.edu.cn

引用格式：*谭剑锋，孙义鸣，王浩文，等. 共轴刚性双旋翼非定常气动干扰载荷分析[J]. 北京航空航天大学学报，2018，44(1)：50-62. TAN J F，SUN Y M，WANG H W，et al. Analysis of rigid coaxial rotor unsteady interactional aerodynamic loads[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(1)：50-62 (in Chinese).*

craft Aerodynamics and Dynamics Ⅱ）研究高速共轴刚性双旋翼系统桨叶载荷特性。随后 Schmaus 和 Chopra[5] 于 2015 年采用基于自由尾迹模型的 UMARC（University of Maryland Advanced Rotor Code）研究高速共轴刚性双旋翼系统结构载荷。Quackenbush 等[6] 采用基于自由尾迹的 CHARM（Comprehensive Hierarchical Aeromechanics Rotorcraft Model）研究新型高速共轴刚性双旋翼复合式直升机性能。此外，Baeder 等[7] 基于自由尾迹模型的 PRASADUM（Parallelized Rotorcraft Analysis for Simulation and Design developed at the University of Maryland）研究高速共轴刚性双旋翼前飞状态下的桨叶载荷。然而，这些方法通过翼型静态气动特性描述旋翼反流区的升力变化，而未考虑高速共轴刚性双旋翼后行边反流区的前缘脱体涡，也未考虑共轴刚性双旋翼桨尖涡对桨叶非定常气动载荷的干扰影响。

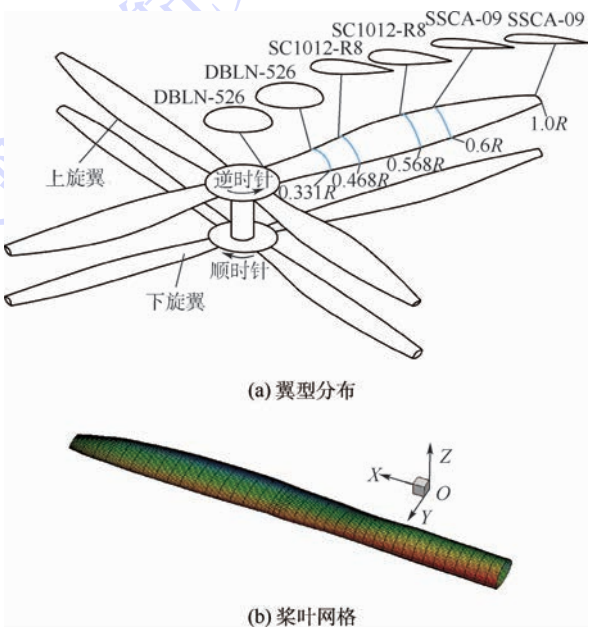
计算流体力学（Computational Fluid Dynamics，CFD）方法基于速度-压力形式的 Navier-Stokes 方程，适用于旋翼流场分析[8]。叶正寅和徐国华等基于嵌套 CFD 方法研究悬停状态共轴刚性双旋翼非定常流动干扰特性[9-10]。随后招启军等[11] 将此方法用于悬停状态共轴刚性双旋翼非定常流动干扰特性研究，但均未针对共轴刚性双旋翼高速状态的非定常气动载荷开展研究。Singh 和 Kang[12] 于 2015 年采用 CFD 方法研究了共轴刚性双旋翼桨叶时变载荷，表明共轴刚性双旋翼载荷分析需考虑双旋翼气动干扰因素。Baeder 等[7] 采用 CFD/CSD（Computaional Strutural Dyanmics）研究了高速共轴刚性双旋翼前飞状态下的桨叶载荷，表明桨叶载荷受上、下旋翼气动干扰影响明显，且需考虑后行边反流区影响。此外，Bhagwat 等[13] 基于 Helios（Helicopter overset simulations）和 RCAS（Rotorcraft Comprehensive Analysis System）的 CFD/CSD 混合方法研究了高速刚性共轴双旋翼桨叶载荷，表明共轴刚性双旋翼之间气动干扰影响共轴刚性双旋翼气动特性。虽然上述 CFD 方法能较好分析高速共轴刚性双旋翼桨叶气动载荷，但计算的共轴刚性双旋翼尾迹耗散较大，且需复杂的网格和较多的计算资源[14]。

黏性涡粒子法基于速度-涡量形式的 Navier-Stokes 方程，适用于旋翼尾迹、旋翼气动载荷[15-16]、旋翼与平尾[17] 以及旋翼与尾桨气动干扰分析[18]。为降低高速共轴刚性双旋翼前飞状态下的需要功率，高速状态下的 X2 旋翼转速下降 20%，最大前飞速度下的桨尖 *Ma* 数 M_{tip} 约为 0.82，抑制桨叶压缩性的影响[7]，因此，涡方法仍可适用于共轴刚性双旋翼桨叶气动特性分析，且已得到成功应用[4-7]。然而，此方法未考虑高速共轴刚性双旋翼后行边反流区以及共轴刚性双旋翼之间桨尖涡干扰影响。为此，本文基于非定常面元法建立满足桨叶前缘和后缘边界条件的旋翼反流区气动模型以体现高速共轴刚性双旋翼后行边反流区影响，且增加共轴刚性双旋翼桨尖涡-桨叶气动干扰模型以体现共轴刚性双旋翼非定常气动干扰的影响，并结合基于黏性涡粒子法的共轴刚性双旋翼尾迹模型，构建高速共轴刚性双旋翼气动干扰下的时变气动载荷分析方法。

1　计算方法

1.1　共轴刚性双旋翼桨叶气动基本原理

与单旋翼不同，前飞状态下的共轴刚性双旋翼系统存在较大反流区和升力偏置，且共轴刚性双旋翼系统的上、下旋翼桨尖涡干扰也将影响桨叶的非定常气动载荷。此外，共轴刚性双旋翼采用非线性弦长、非线性负扭、多段翼型桨叶，其中 X2 高速共轴刚性双旋翼桨叶采用钝头翼型 DBLN-526、SC1012-R8 以及 SSCA-09 翼型（见图 1(a)，其中 R 为桨叶半径），且各段翼型之间光滑过渡。基于非定常面元法和翼型曲线插值，将共轴刚性双旋翼桨叶离散为光滑表面单元，得到 X2 桨叶离散网格如图 1(b) 所示，且根据非定常面元法求得桨叶表面压力分布，从而获得桨



(a) 翼型分布

(b) 桨叶网格

图 1　X2 高速共轴刚性双旋翼桨叶翼型和网格

Fig. 1　Airfoil and grid of X2 high-speed rigid coaxial rotor blade

北航学报 赠阅

叶各截面的气动特性，而无需通过翼型气动特性查表修正，因而更好体现桨叶非定常气动特性。

根据非定常面元法基本理论，定义共轴刚性双旋翼流场速度势函数 ϕ_v 由共轴刚性双旋翼桨叶和尾迹涡面的汇 σ_s 和偶极子 μ_d 构成[15,17-18]：

$$\phi_v(x,y,z,t)=\frac{1}{4\pi}\int_{S_B}\mu_d\boldsymbol{n}\cdot\nabla\left(\frac{1}{|\boldsymbol{r}_t|}\right)\mathrm{d}S-\frac{1}{4\pi}\int_{S_B}\sigma_s\left(\frac{1}{|\boldsymbol{r}_t|}\right)\mathrm{d}S+\frac{1}{4\pi}\int_{S_W}\mu_d\boldsymbol{n}\cdot\nabla\left(\frac{1}{|\boldsymbol{r}_t|}\right)\mathrm{d}S \tag{1}$$

式中：S_B 为共轴刚性双旋翼桨叶表面；S_W 为共轴刚性双旋翼尾迹涡面；$\boldsymbol{n}$ 为共轴刚性双旋翼桨叶和尾迹涡面外法线；$\boldsymbol{r}_t$ 为共轴刚性双旋翼桨叶和尾迹涡面位置；S 为面积积分变量。

根据共轴刚性双旋翼桨叶表面无穿透和远场无扰动边界条件，则

$$\begin{cases}\dfrac{\partial\phi_v}{\partial n_B}-\boldsymbol{v}_B\cdot\boldsymbol{n}=0 & \text{物面边界}\\ \lim\ \nabla\phi_{v,\boldsymbol{r}_t\to\infty}=0 & \text{远场边界}\end{cases} \tag{2}$$

式中：$\boldsymbol{v}_B$ 为共轴刚性双旋翼桨叶速度；n_B 为桨叶表面单位法向。

1.2 共轴刚性双旋翼桨叶反流区气动模型

大速度前飞状态，共轴刚性双旋翼均存在较大的后行边反流区。前行边的空气从桨叶前缘向后缘移动，并满足后缘 Kutta 条件，根据桨叶 Neumann 和后缘 Kutta 条件，后缘处形成脱体涡（见图 2），涡面偶极子 μ_w^{TE} 为

$$\mu_w^{TE}=\mu_u^{TE}-\mu_d^{TE} \tag{3}$$

式中：μ_u^{TE} 和 μ_d^{TE} 分别为桨叶后缘上、下表面偶极子强度。

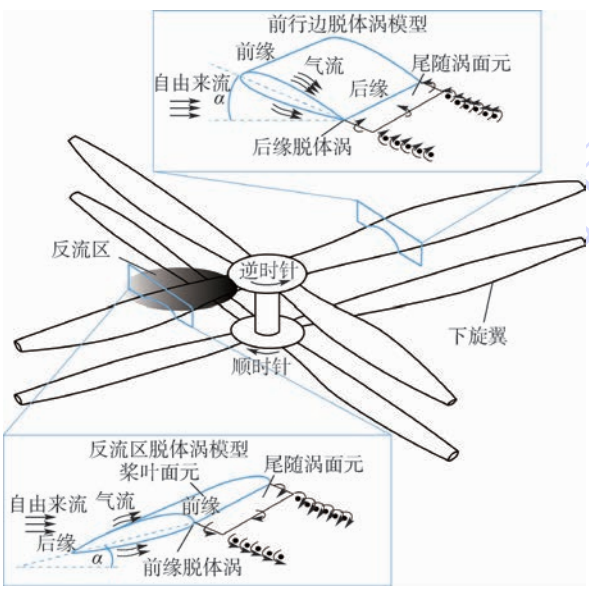


图 2 高速共轴刚性双旋翼反流区气动模型

Fig. 2 Reverse flow aerodynamic model of high-speed rigid coaxial rotor

与前行边不同，共轴刚性双旋翼后行边反流区的空气从桨叶后缘向前缘移动，并在前缘形成脱体涡（见图 2），满足前缘 Kutta 条件，因此桨叶前缘尾迹涡面偶极子 μ_w^{LE} 为

$$\mu_w^{LE}=\mu_u^{LE}-\mu_d^{LE} \tag{4}$$

式中：μ_u^{LE} 和 μ_d^{LE} 分别为桨叶前缘上、下表面偶极子强度。

将式（3）和式（4）代入式（1），并根据 Neumann 和前后缘 Kutta 条件，得到共轴刚性双旋翼桨叶偶极子求解方程：

$$\sum_{k=1}^{N}\mu_{d,k}\boldsymbol{A}_k=-\sum_{k=1}^{N}\sigma_{s,k}\boldsymbol{B}_k \tag{5}$$

式中：$\mu_{d,k}$ 和 $\sigma_{s,k}$ 分别为共轴刚性双旋翼桨叶第 k 个面元偶极子和汇强度；$\boldsymbol{A}_k$ 为共轴刚性双旋翼桨叶与从前缘或后缘脱出的尾迹偶极子影响系数之和；$\boldsymbol{B}_k$ 为共轴刚性双旋翼桨叶的汇影响系数。

$$\boldsymbol{A}_k=\begin{cases}\dfrac{1}{4\pi}\displaystyle\int_{\text{blade}}\boldsymbol{n}_k\cdot\nabla(1/|\boldsymbol{r}_k|)\mathrm{d}S_k & k\neq\text{LE 或 TE}\\ \dfrac{1}{4\pi}\displaystyle\int_{\text{blade}}\boldsymbol{n}_k\cdot\nabla(1/|\boldsymbol{r}_k|)\mathrm{d}S_k\pm\dfrac{1}{4\pi}\int_{\text{TEwake}}\boldsymbol{n}_{TE}\cdot\nabla(1/|\boldsymbol{r}_{TE}|)\mathrm{d}S_{TE} & k=\text{TE}\\ \dfrac{1}{4\pi}\displaystyle\int_{\text{blade}}\boldsymbol{n}_k\cdot\nabla(1/|\boldsymbol{r}_k|)\mathrm{d}S_k\pm\dfrac{1}{4\pi}\int_{\text{LEwake}}\boldsymbol{n}_{LE}\cdot\nabla(1/|\boldsymbol{r}_{LE}|)\mathrm{d}S_{LE} & k=\text{LE}\end{cases} \tag{6}$$

$$\boldsymbol{B}_k=-\frac{1}{4\pi}\int_{\text{blade}}(1/|\boldsymbol{r}_k|)\mathrm{d}S_k \tag{7}$$

其中：$\boldsymbol{n}_{TE}$ 和 $\boldsymbol{n}_{LE}$ 分别为后缘和前缘法向矢量；$\boldsymbol{r}_{TE}$ 和 $\boldsymbol{r}_{LE}$ 分别为后缘和前缘位置矢量；S_k 为第 k 个单元面积；下标 blade 表示桨叶面元，下标 TEwake 表示桨叶后缘涡面，下标 LEwake 表示桨叶前缘涡面；$\boldsymbol{r}_k$ 为第 k 个面元的位置矢量。

求解式（5）得到共轴刚性双旋翼桨叶偶极子 μ_d 分布。与前行边桨叶相似，根据式（4）得到反流区桨叶前缘的尾迹涡面面元强度 μ_w^{LE}，而后由尾迹涡面速度求出下一个时间步 $t+\Delta t$ 的涡面位置，并根据涡量等效原则积分得到在尾随偶极子面元中心的涡粒子涡量（见图 2）。通过速度-涡量形式的 Navier-Stokes 方程确定共轴刚性双旋翼尾迹涡量的空间分布，从而体现共轴刚性双旋翼反流区对共轴刚性双旋翼尾迹结构的影响。

1.3 共轴刚性双旋翼桨尖涡-桨叶气动干扰模型

共轴刚性双旋翼系统的上旋翼桨尖涡将穿过下旋翼桨叶表面，从而改变下旋翼桨叶表面压力分布，并影响下旋翼桨叶气动载荷的时变非线性特性。与此相同，下旋翼也将影响上旋翼桨叶非定常气动载荷。根据伯努利方程，共轴刚性双旋

北航学报 赠阅

翼桨尖涡干扰下的上、下旋翼桨叶非定常压力系数 C_p^{u} 和 C_p^{d} 为

$$\begin{cases} C_p^{\mathrm{u}} = \dfrac{p^{\mathrm{u}} - p_{\mathrm{ref}}}{\frac{1}{2}\rho(\boldsymbol{v}_{\mathrm{ref}}^{\mathrm{u}})^2} = 1 - \dfrac{(\boldsymbol{v}_{\mathrm{B}}^{\mathrm{u}})^2}{(\boldsymbol{v}_{\mathrm{ref}}^{\mathrm{u}})^2} - \dfrac{2}{(\boldsymbol{v}_{\mathrm{ref}}^{\mathrm{u}})^2} \cdot \\ \qquad \left(\dfrac{\partial \phi_{\text{down blade}}}{\partial t} + \dfrac{\partial \phi_{\text{down wake}}}{\partial t} \right) \\ C_p^{\mathrm{d}} = \dfrac{p^{\mathrm{d}} - p_{\mathrm{ref}}}{\frac{1}{2}\rho(\boldsymbol{v}_{\mathrm{ref}}^{\mathrm{d}})^2} = 1 - \dfrac{(\boldsymbol{v}_{\mathrm{B}}^{\mathrm{d}})^2}{(\boldsymbol{v}_{\mathrm{ref}}^{\mathrm{d}})^2} - \dfrac{2}{(\boldsymbol{v}_{\mathrm{ref}}^{\mathrm{d}})^2} \cdot \\ \qquad \left(\dfrac{\partial \phi_{\text{up blade}}}{\partial t} + \dfrac{\partial \phi_{\text{up wake}}}{\partial t} \right) \end{cases} \tag{8}$$

式中：p_{ref} 和 ρ 分别为参考压力和空气密度；$\boldsymbol{v}_{\mathrm{B}}^{\mathrm{u}}$、$p^{\mathrm{u}}$、$\boldsymbol{v}_{\mathrm{ref}}^{\mathrm{u}}$、$\phi_{\text{up blade}}$ 和 $\phi_{\text{up wake}}$ 分别为上旋翼桨叶表面空气速度、表面压力、当地参考速度以及上旋翼桨叶诱导速度势和尾迹诱导速度势；$\boldsymbol{v}_{\mathrm{B}}^{\mathrm{d}}$、$p^{\mathrm{d}}$、$\boldsymbol{v}_{\mathrm{ref}}^{\mathrm{d}}$、$\phi_{\text{down blade}}$ 和 $\phi_{\text{down wake}}$ 分别为下旋翼桨叶表面空气速度、表面压力、当地参考速度以及下旋翼桨叶诱导速度势和尾迹诱导速度势。

式(8)中 $\partial\phi_{\text{up blade}}/\partial t$ 描述上旋翼桨叶时变气动载荷对下旋翼桨叶表面压力的影响；而 $\partial\phi_{\text{down blade}}/\partial t$ 表示下旋翼桨叶时变气动载荷对上旋翼桨叶表面压力的影响，以体现共轴刚性双旋翼桨叶交叠时存在的载荷干扰。

$$\begin{cases} \dfrac{\partial \phi_{\text{up blade}}}{\partial t} = \dfrac{\phi_{\text{up blade}}^{t} - \phi_{\text{up blade}}^{t-\Delta t}}{\Delta t} \\ \dfrac{\partial \phi_{\text{down blade}}}{\partial t} = \dfrac{\phi_{\text{down blade}}^{t} - \phi_{\text{down blade}}^{t-\Delta t}}{\Delta t} \end{cases} \tag{9}$$

式(8)中 $\partial\phi_{\text{up wake}}/\partial t$ 表述上旋翼桨尖涡对下旋翼桨叶表面压力的影响，可转变为上旋翼桨尖涡对下旋翼桨叶的诱导速度与上旋翼桨尖涡速度的乘积；而 $\partial\phi_{\text{down wake}}/\partial t$ 描述下旋翼桨尖涡对上旋翼桨叶表面压力的影响，可转变为下旋翼桨尖涡对上旋翼桨叶诱导速度与下旋翼桨尖涡速度的乘积。

$$\begin{cases} \left.\dfrac{\partial \phi}{\partial t}\right|_{\text{up wake}} = -\sum \boldsymbol{v}_{\mathrm{u}}^{\mathrm{ind}}(\boldsymbol{x}_{\mathrm{u}}) \cdot \boldsymbol{v}_{\mathrm{dw}}(\boldsymbol{x}_{\mathrm{d}}') \\ \left.\dfrac{\partial \phi}{\partial t}\right|_{\text{down wake}} = -\sum \boldsymbol{v}_{\mathrm{d}}^{\mathrm{ind}}(\boldsymbol{x}_{\mathrm{d}}) \cdot \boldsymbol{v}_{\mathrm{uw}}(\boldsymbol{x}_{\mathrm{u}}') \end{cases} \tag{10}$$

式中：$\boldsymbol{x}_{\mathrm{u}}$、$\boldsymbol{v}_{\mathrm{uw}}$ 和 $\boldsymbol{x}_{\mathrm{u}}'$ 分别为上旋翼桨叶位置、桨尖涡速度和位置；$\boldsymbol{x}_{\mathrm{d}}$、$\boldsymbol{v}_{\mathrm{dw}}$ 和 $\boldsymbol{x}_{\mathrm{d}}'$ 分别为下旋翼桨叶位置、桨尖涡速度和位置；$\boldsymbol{v}_{\mathrm{u}}^{\mathrm{ind}}$ 和 $\boldsymbol{v}_{\mathrm{d}}^{\mathrm{ind}}$ 分别为下旋翼桨尖涡对上旋翼桨叶的诱导速度和上旋翼桨尖涡对下旋翼桨叶的诱导速度。

共轴刚性双旋翼桨尖涡干扰下的双旋翼桨叶气动载荷为

$$\Delta \boldsymbol{F}_k = -C_{p,k}(\rho v_{\mathrm{ref}}^2/2)_k \Delta S_k \boldsymbol{n}_k \tag{11}$$

式中：$\Delta \boldsymbol{F}_k$ 和 ΔS_k 分别为气动力和面积；$C_{p,k}$ 为面元压力系数；$\boldsymbol{n}_k$ 为面元外法线矢量。

1.4　共轴刚性双旋翼尾迹模型

与单旋翼不同，共轴刚性双旋翼系统的上、下旋翼桨叶均脱出桨尖涡，受到双旋翼桨尖涡之间的相互干扰作用，桨尖涡几何特征将发生改变，且上旋翼桨尖涡将穿过下旋翼旋转平面（见图 3(a)），从而改变共轴刚性双旋翼非定常气动载荷。

基于黏性涡粒子法[15-18]，共轴刚性双旋翼涡量分布由速度-涡量 $(\boldsymbol{u}, \boldsymbol{\omega})$ 形式的 Navier-Stokes 方程确定：

$$\frac{\partial \boldsymbol{\omega}}{\partial t} + \boldsymbol{u} \cdot \nabla \boldsymbol{\omega} = \nabla \boldsymbol{u} \cdot \boldsymbol{\omega} + \nu \nabla^2 \boldsymbol{\omega} \tag{12}$$

式中：ν 为运动黏性系数。

采用四阶 Runge-Kutta、PSE（Particle Strength Exchange）、直接求解法求解式(12)获得共轴刚性双旋翼涡量场空间分布，得到共轴刚性双旋翼桨尖涡结构。

共轴刚性双旋翼系统的上旋翼桨尖涡将穿过下旋翼旋转平面，从而改变下旋翼气动载荷，但上旋翼桨尖涡靠近下旋翼桨叶表面，并进入下旋翼桨叶内部，从而改变上旋翼桨尖涡对下旋翼桨叶气动载荷的干扰影响。为体现上、下旋翼桨叶对共轴刚性双旋翼桨尖涡的影响，基于镜面映射法，进入桨叶的涡强为涡量与体积 V_{ACB} 之积（见图 3(b)），则镜面涡量为

$$\begin{cases} \alpha_x^{\mathrm{i}} = \alpha_x \\ \alpha_y^{\mathrm{i}} = \alpha_y \\ \alpha_z^{\mathrm{i}} = -\alpha_z \end{cases} \tag{13}$$

式中：α_x、α_y 和 α_z 分别为涡量在桨叶表面的 2 个切分量和法分量；α_x^{i}、α_y^{i} 和 α_z^{i} 分别为镜面涡量在桨叶表面的 2 个切分量和法分量。

1.5　共轴刚性双旋翼配平

与常规旋翼系统不同，由于共轴刚性双旋翼之间存在总距、纵、横向周期操纵差异，因此定义双旋翼操纵量为

$$\begin{cases} \theta_{\mathrm{u}} = (\theta_0 + \Delta\theta_0) - (\theta_{1\mathrm{c}} + \Delta\theta_{1\mathrm{c}})\cos\psi - \\ \qquad (\theta_{1\mathrm{s}} + \Delta\theta_{1\mathrm{s}})\sin\psi \\ \theta_{\mathrm{d}} = (\theta_0 - \Delta\theta_0) - (\theta_{1\mathrm{c}} - \Delta\theta_{1\mathrm{c}})\cos\psi - \\ \qquad (\theta_{1\mathrm{s}} - \Delta\theta_{1\mathrm{s}})\sin\psi \end{cases} \tag{14}$$

式中：θ_{u}、θ_{d} 分别为上、下旋翼桨距；θ_0、$\theta_{1\mathrm{c}}$ 和 $\theta_{1\mathrm{s}}$ 分别为双旋翼系统总距、纵向和横向操纵量平均值；$\Delta\theta_0$、$\Delta\theta_{1\mathrm{c}}$ 和 $\Delta\theta_{1\mathrm{s}}$ 分别为双旋翼系统总距、纵向和横向操纵量差值；ψ 为旋翼桨叶方位角。

北航学报 赠阅

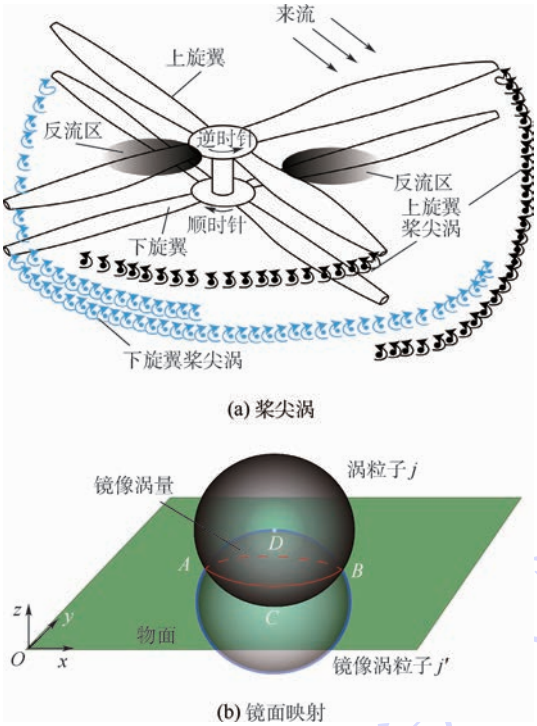


图 3 X2 高速共轴刚性双旋翼桨尖涡和镜面映射

Fig. 3 Tip vortex of X2 rigid coaxial rotor and mirror mapping of vorticity

根据共轴刚性双旋翼系统力矩和力平衡要求,稳定状态下的加速度均为零,因此平衡方程为

$$\begin{cases}\begin{bmatrix}F_x\\F_y\\F_z\end{bmatrix}=\begin{bmatrix}T\sin\theta\\-T\cos\theta\sin\phi\\-T\cos\theta\cos\phi\end{bmatrix}\\\begin{bmatrix}M_x\\M_y\\M_z\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}\end{cases}\tag{15}$$

式中:F_x、F_y、F_z 和 M_x、M_y、M_z 分别为共轴刚性双旋翼的气动力和力矩;T 为重力;θ 为双旋翼系统俯仰角;ϕ 为双旋翼系统滚转角。

与常规直升机配平不同,由于共轴刚性双旋翼操作量多于平衡方程,因而配平较为复杂。为此,文献[19]建立基于目标函数和约束条件的优化方法实现共轴刚性双旋翼飞行器全机配平。此外,文献[20]通过建立共轴刚性双旋翼飞行器飞行动力学模型,给定提前操纵角和纵、横向周期变距差动值实现整机配平。针对共轴刚性双旋翼系统,给定旋翼系统俯仰姿态,通过旋翼系统力和力矩平衡,从而实现双旋翼系统配平,并已应用于共轴刚性双旋翼系统性能和桨叶结构载荷分析[5-6,12-13]。为此,针对共轴刚性双旋翼系统,上述双旋翼系统平衡方程为 6 个,而变量为 8 个,双旋翼配平变量多于平衡方程数目,为实现对共轴刚性双旋翼系统配平方程封闭,基于文献[7]给定机体俯仰角 θ 和滚转角 ϕ,减少双旋翼系统操作变量,而后在给定飞行状态下,给定初始操纵量,通过计算共轴刚性双旋翼系统气动载荷,判定是否满足平衡方程。若不满足,则通过差分计算共轴刚性双旋翼系统 Jacobian 矩阵,并通过牛顿迭代法计算双旋翼操纵量,直到满足平衡方程为止,完成共轴刚性双旋翼系统配平。

为验证共轴刚性双旋翼系统配平的可靠性,计算得到 Harrington 共轴双旋翼悬停状态[21]和前飞状态下的性能[22]如图 4 所示。图中 P 为双旋翼功率,hp,1 hp = 745.7 W;C_T 为旋翼系统拉力系数;C_Q 为旋翼系统扭矩系数;μ 为前进比。

从图 4 中可以看出,计算得到的 Harrington 共轴双旋翼悬停状态下的 C_T-C_Q 曲线与实验测量结果[21]吻合的很好,悬停状态下,共轴双旋翼之间存在一定的总距差动。此外,计算得到 Harrington 共轴双旋翼在 C_T = 0.004 8 下各前飞速度的功率与实验测量结果[22]基本吻合,其中共轴双旋翼之间的总距、纵横向周期变距均存在差动,由此表明本文共轴刚性双旋翼气动载荷配平方法的可行性。

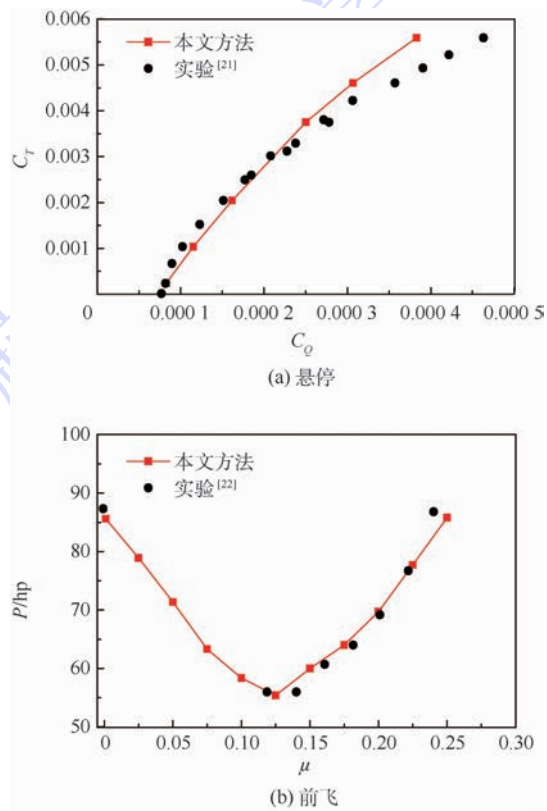


图 4 Harrington 共轴双旋翼性能

Fig. 4 Performance of Harrington coaxial rotor

2　高速共轴刚性双旋翼非定常气动干扰特性

2.1　X2 高速共轴刚性双旋翼非定常气动载荷

本算例计算前飞状态下 X2TD（X2 Technology Demonstrator）高速共轴刚性双旋翼桨叶非定常气动载荷[7,23]。上、下旋翼均由 4 片非均匀弦长和非线性负扭桨叶构成，旋翼半径为 4.023 m，桨尖 M_{tip} 为 0.554，桨叶翼型分布如图 1 所示。计算模型桨叶弦向为 60 段，展向为 40 段，双旋翼系统由 19 200 个面元构成，方位角步长为 2.5°，双旋翼系统操纵量如表 1 所示。

计算得到前飞速度为 102、185 和 278 km/h 状态下的旋翼特征剖面非定常气动载荷如图 5 所示。图中：C_n 为截面法向升力系数；ψ 为方位角。图中同时给出基于自由尾迹的综合分析方法 PRASADUM 计算结果[7]以及基于 NASA OVERFLOW 和 CREATE AV Helios 的 CFD / CSD 计算结果[7]，其中 PRASADUM 耦合了共轴刚性双旋翼自由尾迹模型和桨叶气动弹性模型，CFD/CSD 耦合了共轴刚性双旋翼 CFD 模型和桨叶气动弹性模型。由于共轴刚性双旋翼桨叶挥舞运动由桨叶弹性变形产生，为体现挥舞运动的影响，基于铰接式旋翼挥舞运动方程，通过等效挥舞外伸量和等效挥舞约束弹簧刚度[20,24]，保证挥舞频率满足旋翼弹性挥舞频率 1.42，并近似一阶弹性挥舞。3 种前飞状态下，本文分析方法计算得到特征剖面非定常气动载荷变化特征与 CFD/CSD 计算结果均吻合地较好，体现了上、下旋翼在前行侧和后行侧由干扰引起的非定常载荷减小特性，并较好计算 X2 共轴刚性双旋翼非定常气动载荷。由于 PRASADUM 的共轴刚性双旋翼自由尾迹模型通过桨尖涡模拟共轴刚性双旋翼系统尾迹结构，暂未考虑双旋翼桨叶整体涡结构对桨叶非定常气动载荷的干扰以及后行边反流区的前缘脱体涡特性，PRASADUM 计算结果总体偏大，且前行边和后行边升力系数均出现较大振荡。相比于 PRASADUM 计算结果，本文方法计算的上、下旋翼非定常气动载荷随时间变化幅值、相位均与 CFD/CSD 计算结果吻合地更好，且更好体现各前飞速度上、下旋翼在前行边和后行边升力系数下降特性。此外，由于在低速前飞状态下，后行边反流区较小，因而增加桨叶后行边反流区气动模型的计算结果改善不明显；然而，大速度前飞状态下，旋翼后行边反流区逐渐增大，增加桨叶后行边反流区气

表 1　共轴刚性双旋翼系统操纵量（185 km/h）

Table 1　Control variables of rigid coaxial rotor system（185 km/h）　（°）

变　量	PRASADUM[7]	CFD[7]	本　文
θ_0	5.84	6.38	8.87
$\Delta\theta_0$	0.32	0.1	0.1
θ_{1c}	3.38	1.77	0.27
$\Delta\theta_{1c}$	0	0	0
θ_{1s}	0.21	0.34	0.33
$\Delta\theta_{1s}$	4.29	4.1	5.7

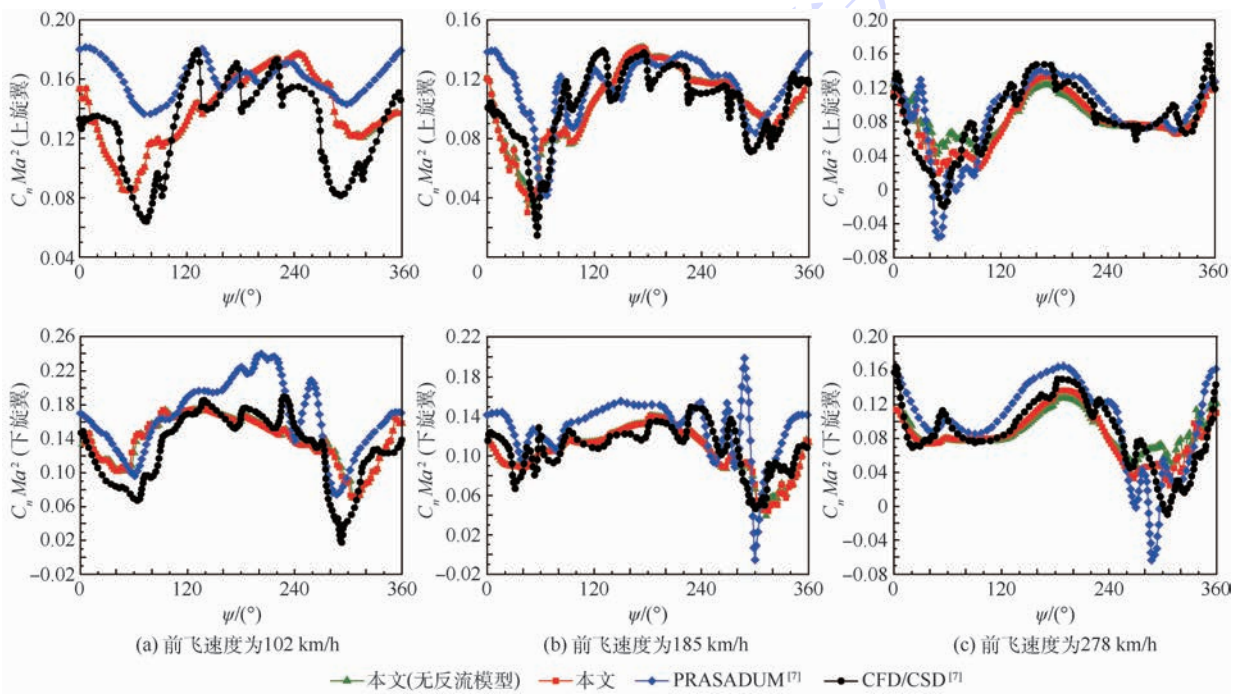


图 5　不同前飞速度下的 X2 旋翼截面载荷

Fig. 5　Sectional load of X2 rotor at different forward speeds

北航学报 赠阅

动模型的计算结果明显与 CFD/CSD 计算结果更吻合。上旋翼后行边桨叶根部截面涡量如图 6 所示。增加桨叶后行边反流区气动模型后,后行边的桨叶内段前缘脱出涡,从而增加桨根区域涡量强度,并与之后的桨叶干扰,增加方位角 300°桨叶非定常气动载荷变化幅度。而未考虑桨叶后行边反流区气动模型,涡量则从桨叶后缘脱出,而后叠加在桨叶表面,削弱桨根区域涡量强度,减小与之后桨叶干扰产生的非定常气动载荷。因此,后行边涡量截面分布和气动载荷变化均可体现后行边反流区气动模型的可行性。

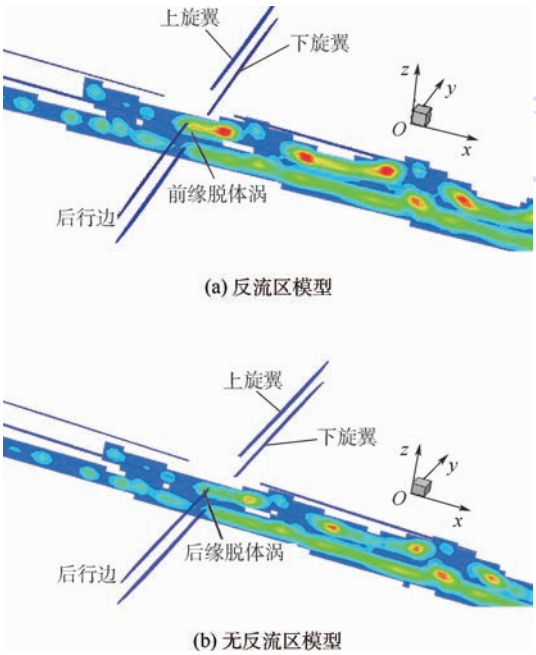


图 6 上旋翼后行边桨叶截面涡量(278 km/h)

Fig. 6 Vorticity of blade section at retreating side of upper rotor (278 km/h)

2.2 上、下旋翼之间的非定常气动干扰特性

图 7 给出 3 种前飞速度状态下 X2 共轴刚性双旋翼系统上、下旋翼特征剖面非定常气动载荷变化历程(按上旋翼时间历程)。前飞速度 102 km/h 状态,相比于上旋翼,下旋翼非定常气动载荷在方位角 60°和 300°处明显下降,且在后行边 300°处下降更显著。主要原因在于下旋翼受到上旋翼桨尖涡干扰的影响。上旋翼桨尖涡将穿过下旋翼后半部分(见图 8(a),其中上旋翼桨尖涡采用红色表示,下旋翼桨尖涡采用淡蓝色表示),且在方位角 60°和 300°处受到上旋翼卷起桨尖涡干扰显著,因而导致下旋翼诱导入流 λ 在方位角 60°和 300°处明显大于上旋翼(见图 9(a1)、(a2))。随着前飞速度的增加,双旋翼桨尖涡向下速度逐渐减小,上旋翼桨尖涡沿着上旋翼桨盘平面移动,而下旋翼桨尖涡沿着下旋翼桨盘平面移动(见图 8(b)、(c)),从而对下旋翼干扰减弱,因此下旋翼的非定常气动载荷时间历程与上旋翼基本相似(见图 7(b)、(c)),上旋翼诱导速度分布与下旋翼诱导速度分布也基本相似(见图 9(b1)、(b2))。因此上、下旋翼气动干扰对桨盘诱导速度和非定常气动载荷影响明显。

2 种前飞速度的上、下旋翼总距分别为(8.85°、9.08°)和(11.5°、11.4°),且均存在一定的横向周期操纵量,分别为(7.9°、-7.9°)和(5.2°、-4.7°),因此图 10 中 2 种前飞速度下的拉力主要分布在前半侧。低速状态下,共轴刚

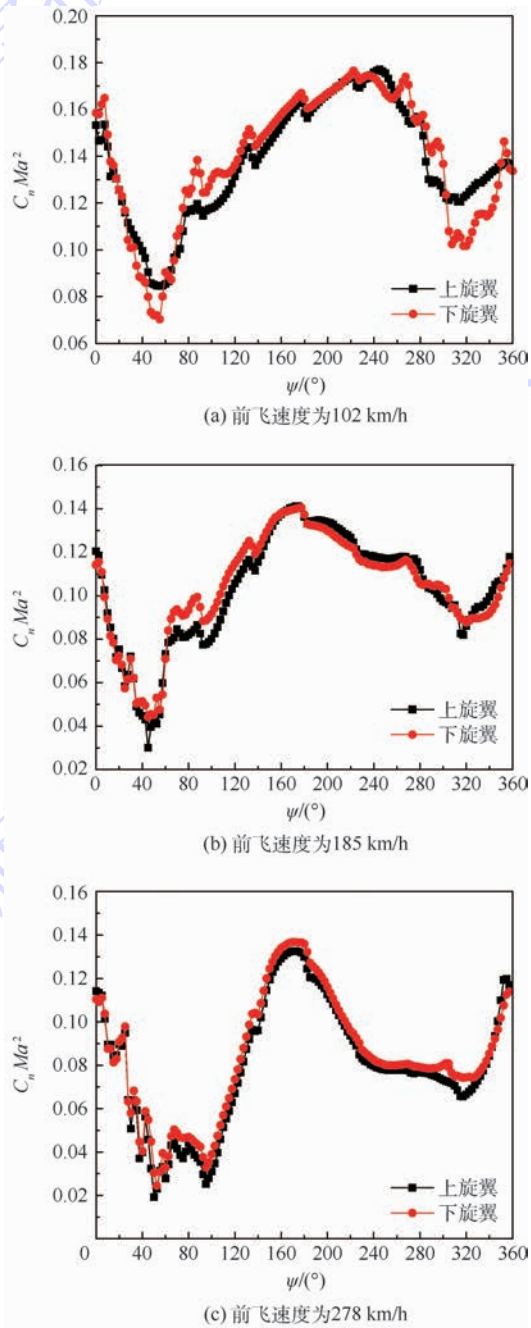


图 7 不同前飞速度下上、下旋翼截面载荷

Fig. 7 Sectional load of upper and lower rotor at different forward speeds

性双旋翼与常规旋翼相似，旋翼反流区较小，升力偏置较小，因而旋翼前半部分和后部分为旋翼主要升力区域（见图10（a1）、（a2））。上、下旋翼前半部分均受到旋翼桨尖涡的干扰影响（见图 8（a）），由此引起上、下旋翼载荷波动，且上、下旋翼桨尖涡位置存在交叉，因而上、下旋翼载荷波动交叉出现，并导致上、下旋翼升力差表现出由桨尖涡干扰导致的丝带特性；此外，上旋翼桨尖涡将穿过下旋翼后部（见图 8（a）），因而上、下旋翼升力差表现出 60°与 300°处卷起桨尖涡干扰特性（见图 10（a3））。随着前飞速度增加，与常规旋翼不同，共轴刚性双旋翼反流区增大，升力偏置明显向前行边移动，因而前行边为双旋翼主要升力区域（见图 10（b1）、（b2））。此外，上、下旋翼升力差呈现出与桨叶片数相关的辐射状特性，而桨尖涡干扰导致的丝带特性和 60°与 300°处卷起桨尖涡干扰特性减弱（见图 10（b3））。主要原因在于高速前飞状态下，上旋翼桨尖涡远离下旋翼（见图 8（c）），桨尖涡干扰效应降低，而旋翼载荷增加，上旋翼对下旋翼时变干扰载荷增加（见式（9）），因而升力特性表现出 8 片桨叶的辐射状干扰。

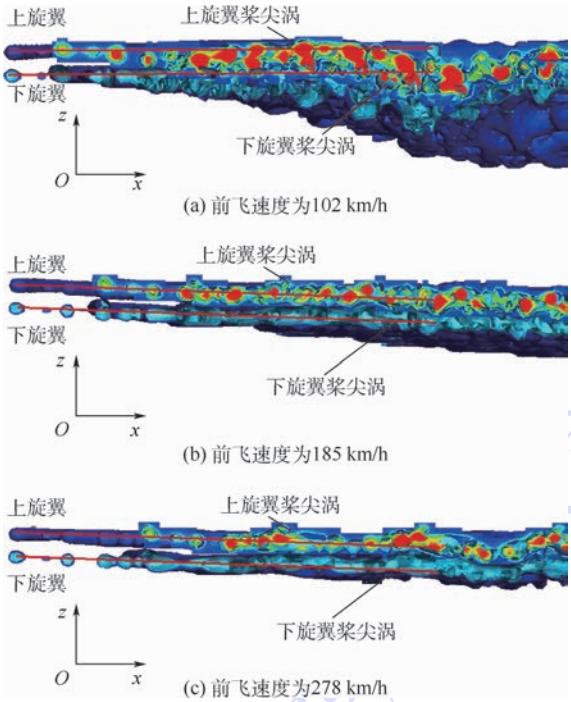


图 8　不同前飞速度下的 X2 旋翼尾迹结构

Fig. 8　Wake structure of X2 rotor at different forward speeds

前飞状态下的共轴刚性双旋翼尾迹结构与单旋翼略有不同。受下旋翼后行边和前行边卷起桨尖涡的诱导影响，上旋翼前行边和后行边桨尖

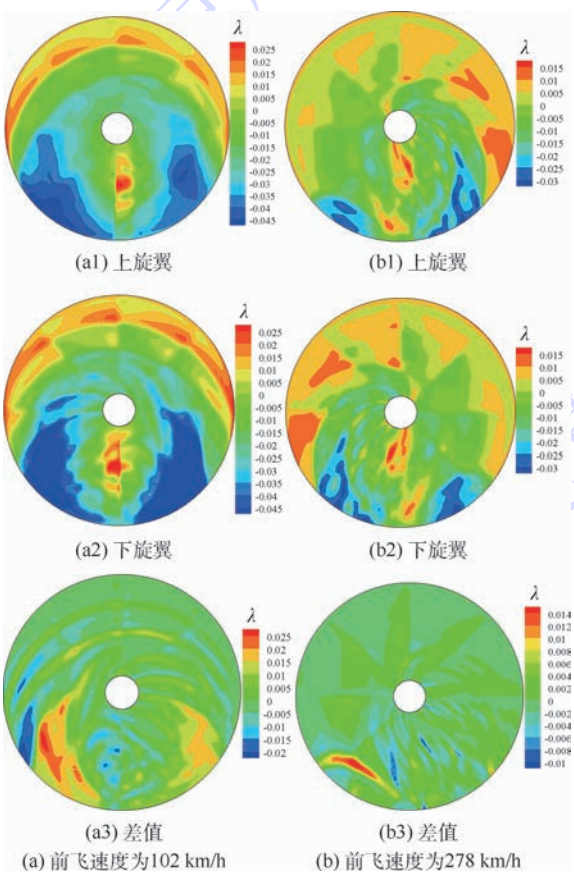


图 9　X2 旋翼诱导入流分布

Fig. 9　Induced flow distribution of X2 rotor

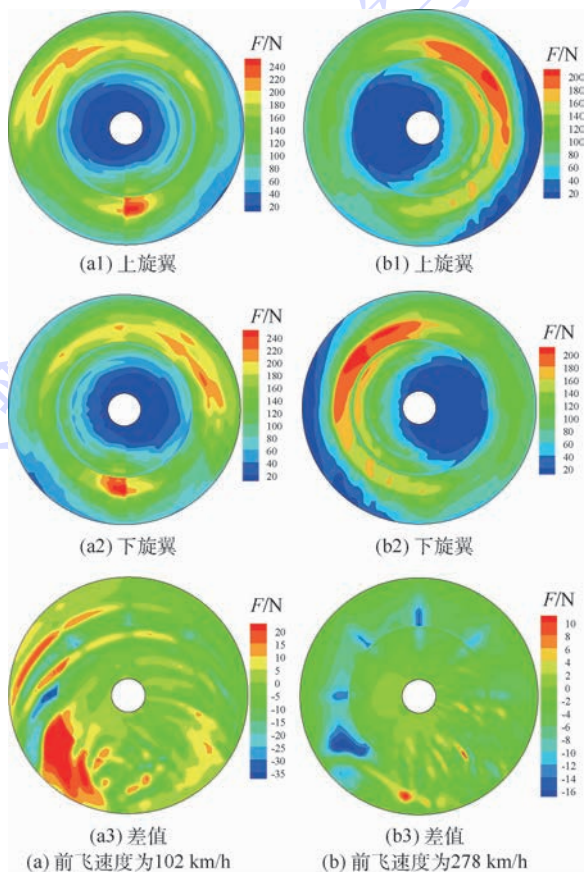


图 10　X2 旋翼截面升力分布

Fig. 10　Sectional force distribution of X2 rotor

北航学报 赠阅

涡向内侧和下方移动，并与下旋翼桨尖涡位置互换（见图11（a），其中上旋翼桨尖涡采用红色表示，下旋翼桨尖涡采用橙色表示）。在 $x=0.25R$ 处上旋翼前行边和后行边脱出桨尖涡①，下旋翼后行边和前行边桨叶脱出桨尖涡②（见图11（b））。受桨尖涡之间的相互干扰作用，在 $x=0.5R\sim0.75R$ 处，上旋翼桨尖涡①开始向旋翼内侧和下方移动，而下旋翼桨尖涡②则向上旋翼方向移动（见图11（c）、（d））；在 $x=1.0R$ 处，下旋翼桨尖涡②达到上旋翼旋转平面高度，并在上旋翼桨尖涡①之上，实现桨尖涡①和②位置互换（见图11（e））。图12给出了此状态下桨叶特征剖面处诱导速度随方位角的变化历程和频率分布。上、下旋翼诱导速度在方位角 80°～240°之间基本相似，然而上、下旋翼桨尖涡位置互换导致下旋翼桨尖涡上移，从而使得前行边 0°～80°和后行边 240°～360°的下旋翼诱导速度显著大于上旋翼。此外，相比于上旋翼诱导速度谐波分量，下旋翼诱导速度不仅 1 Ω 谐波分量增加 17.5%，且3 Ω、4 Ω、5 Ω 谐波分量显著增大（其中 Ω 为旋翼转速的频率单位），分别为 30.9%、144.2%、194.7%。与此同时，上、下旋翼特征剖面非定常气动载荷总体差异小于诱导速度差异，但非定常气动载荷的 3 Ω、4 Ω、5 Ω 谐波分量却显著增大。

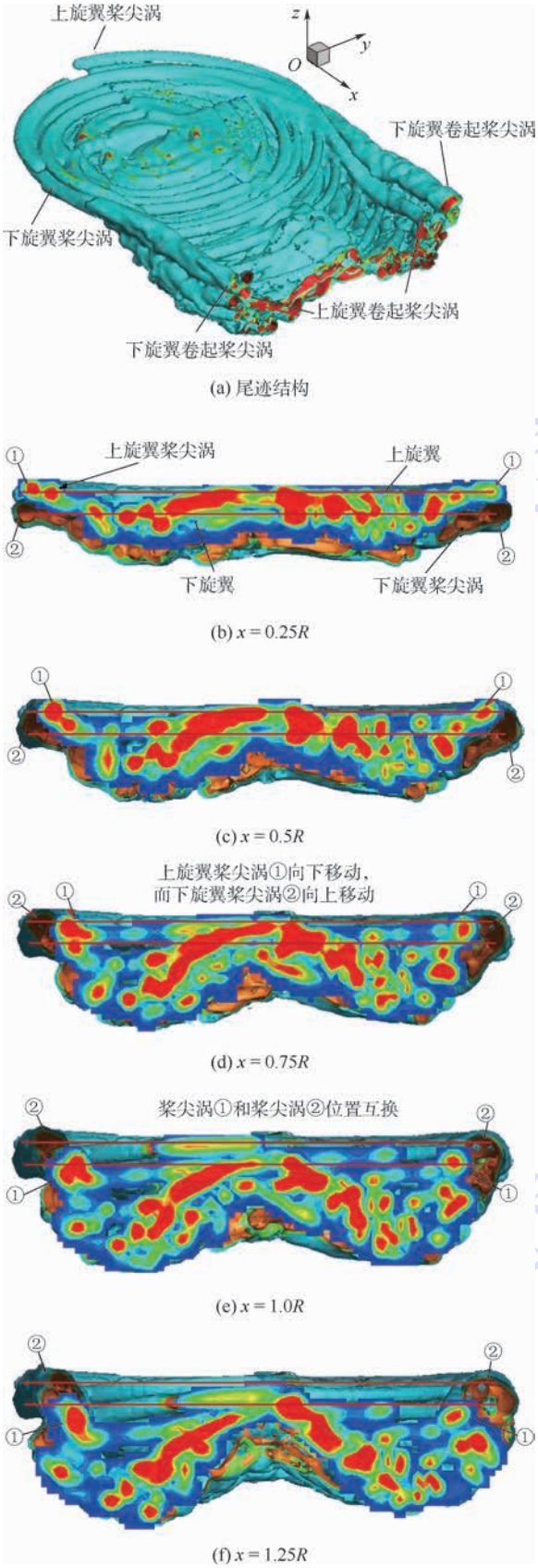


图 11 X2 旋翼桨尖涡位置互换（102 km/h）

Fig. 11 Interchange of tip vortex position of X2 rotor (102 km/h)

2.3 共轴刚性双旋翼与单旋翼气动载荷特性差异

图13给出3种前飞速度状态下单旋翼和共轴刚性双旋翼截面载荷与频率。前飞速度为102 km/h状态，共轴刚性双旋翼在 60°和 300°处升力明显小于孤立旋翼，主要原因为下旋翼在方位角 60°和 300°处受上旋翼卷起桨尖涡的干扰影响（见图9（b）），同时下旋翼卷起桨尖涡向上旋翼上方移动，并对上旋翼也存在较明显的干扰。随着前飞速度的增加，由于上、下旋翼桨尖涡向下移动速度减小，双旋翼桨尖涡之间干扰作用减弱，因而共轴刚性双旋翼和单旋翼之间的气动载荷时间历程差异减小。

相比于单旋翼，共轴刚性双旋翼升力系数 1 Ω、3～10 Ω 谐波分量增大。受上旋翼卷起桨尖涡对下旋翼在 60°和 300°处显著干扰作用影响，共轴刚性双旋翼升力系数 1 Ω 分量明显增加，且随前飞速度的增加而减小。但共轴刚性双旋翼升力系数 8 Ω 谐波分量随前飞速度增加而增加，主要原因为下旋翼桨叶受到上旋翼桨叶干扰作用增强（见式（9）），且下旋翼旋转一周将与上旋翼桨叶叠加 8 次，因而产生 8 Ω 谐波分量。

图14给出共轴刚性双旋翼与单旋翼的诱导速度和截面升力差值。各种状态下，诱导速度和截面升力差值均表现出桨叶片数的辐射特征和旋

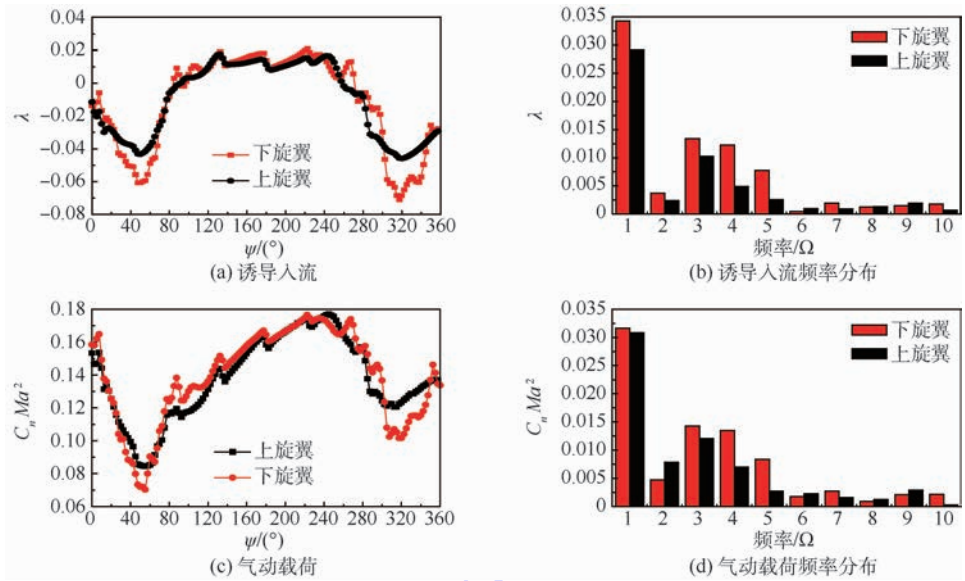


图 12　上、下旋翼诱导入流和截面载荷的频谱特性

Fig. 12　Frequency spectrum of induced flow and sectional load of upper and lower rotor

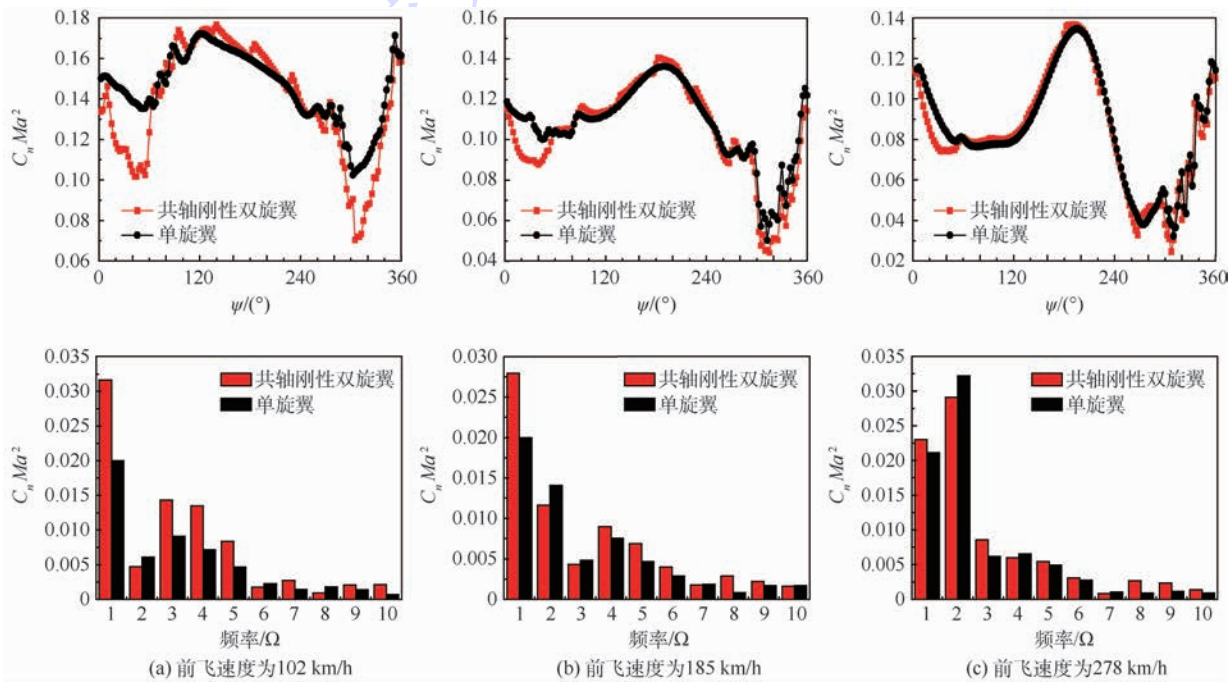


图 13　共轴刚性双旋翼和单旋翼截面载荷与频率

Fig. 13　Sectional load and frequency of rigid coaxial rotor and single rotor

翼后部干扰特性。低速前飞状态，受上旋翼卷起桨尖涡的显著影响，双旋翼诱导速度和截面升力差值均表现出显著的卷起桨尖涡干扰和桨叶片数整数倍（8Ω）的辐射状干扰特性。随着前飞速度增加，由于上、下旋翼桨尖涡干扰作用减小，共轴刚性双旋翼和单旋翼之间的诱导速度和截面升力差异所呈现的卷起桨尖涡干扰特征逐渐减弱，而受到上、下旋翼桨叶干扰更加明显，因而呈现桨叶片数整数倍辐射状的桨叶干扰特征更加突出。

上述研究分析表明，本文分析方法能较好分析共轴刚性双旋翼复杂干扰下的非定常气动载荷，且共轴刚性双旋翼系统的桨尖涡干扰和桨叶干扰对双旋翼桨叶非定常气动载荷存在明显的影响。由于推进螺旋桨也将处于共轴刚性双旋翼桨尖涡干扰区域，为此，将进一步研究共轴刚性双旋翼桨尖涡干扰下的推进螺旋桨性能和载荷，以及由此产生的全机操作特性变化。

北航学报 赠阅

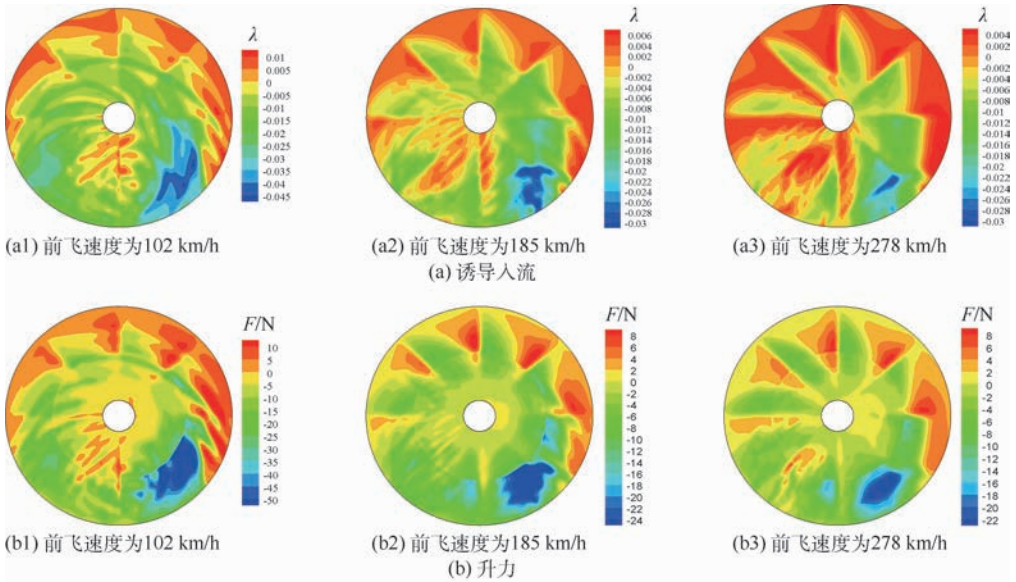


图 14 单旋翼与共轴双旋翼的诱导速度和截面升力差异

Fig. 14 Change in induced velocity and sectional force due to single rotor and rigid coaxial rotor

3 结 论

1）建立满足桨叶前缘和后缘边界条件的高速共轴刚性双旋翼反流区气动模型，并结合共轴刚性双旋翼桨尖涡-桨叶气动干扰模型，构建高速共轴刚性双旋翼非定常气动载荷分析方法，相比于基于自由尾迹的 PRASADUM，本文方法计算更好地体现上、下旋翼在前行边和后行边非定常气动载荷的变化特性，与 CFD/CSD 计算结果更吻合。

2）低速状态，下旋翼桨叶载荷在 60°和 300°方位角出明显下降，但随着前飞速度增加，上、下旋翼气动载荷特性差异缩小。

3）低速状态，上、下旋翼气动载荷差异呈现出由桨尖涡干扰导致的丝带特性和卷起桨尖涡干扰特性，而高速前飞状态则呈现出桨叶片数相关的辐射状特性。

4）受双旋翼桨尖涡相互诱导作用，共轴刚性双旋翼出现桨尖涡位置互换现象。

5）相比于单旋翼，低速状态的共轴刚性双旋翼升力系数 1Ω 谐波分量显著增加，并随前飞速度增加而减小，而高速状态 8Ω 谐波分量增加明显，且桨叶片数整数倍辐射状的桨叶干扰特征更加突出。

参考文献（References）

[1] ELLER E. X2™ load alleviating controls[C]// The American Helicopter Society 68th Annual Forum. Fairfax, VA: AHS, 2012:1578-1587.

[2] TAN J F, WANG H W. Highly efficient unsteady panel time-marching free wake for aerodynamics of rotorcraft[J]. Journal of Aircraft, 2014, 51(1): 54-61.

[3] GOVINDARAJAN B M, LEISHMAN J G. Curvature corrections to improve the accuracy of free-vortex methods[J]. Journal of Aircraft, 2016, 53(2): 378-386.

[4] YEO H, JOHNSON W. Investigation of maximum blade loading capability of lift-offset rotors[C]// The AHS 69th Annual Forum. Fairfax, VA: AHS, 2013: 782-797.

[5] SCHMAUS J, CHOPRA I. Aeromechanics for a high advance ratio coaxial helicopter[C]// The AHS 71st Annual Forum. Fairfax, VA: AHS, 2015: 1139-1153.

[6] GAFFEY T M, ZHANG C, QUACKENBUSH D T, et al. Aeromechanics of the coaxial compound helicopter[C]// 56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. Reston: AIAA, 2015: 1-19.

[7] PASSE B, SRIDHARAN A, BAEDER J. Computational investigation of coaxial rotor interactional aerodynamics in steady forward flight[C]// 33rd AIAA Applied Aerodynamics Conference. Reston: AIAA, 2015: 1-29.

[8] CAO Y H, ZHAO M. Numerical simulation of rotor flowfields based on several spatial discretization schemes[J]. Journal of Aircraft, 2012, 49(5): 1535-1539.

[9] 许和勇，叶正寅. 悬停共轴双旋翼干扰流动数值模拟[J]. 航空动力学报，2011，26(2)：453-457.
XU H Y, YE Z Y. Numerical simulation of interaction unsteady flows around co-axial rotors in hover[J]. Journal of Aerospace Power, 2011, 26(2): 453-457(in Chinese).

[10] 叶靓，徐国华. 共轴式双旋翼悬停流场和气动力的 CFD 计算[J]. 空气动力学学报，2012，30(4)：437-442.
YE L, XU G H. Calculation on flow field and aerodynamic force of coaxial rotors in hover with CFD method[J]. Acta Aerodynamic Sinica, 2012, 30(4): 437-442(in Chinese).

[11] 朱正，招启军，李鹏. 悬停状态共轴刚性双旋翼非定常流动干扰机理研究[J]. 航空学报，2016，37(2)：568-578.

ZHU Z,ZHAO Q J,LI P. Investigations on unsteady flow interaction mechanism of coaxial rigid rotors in hover[J]. Acta Aeronautica et Astronautica Sinica,2016,37(2):568-578(in Chinese).

[12] SINGH R,KANG H. Computational investigations of transient loads and blade deformations on coaxial rotor systems[C]// 33rd AIAA Applied Aerodynamics Conference. Reston:AIAA,2015:101-111.

[13] SINGH R,KANG H,BHAGWAT M,et al. Computational and experimental study of coaxial rotor steady and vibratory loads[C]//54th AIAA Aerospace Sciences Meeting. Reston:AIAA,2016:1-10.

[14] KOMERATH N M,SMITH M J,TUNG C. A review of rotor wake physics and modeling[J]. Journal of the American Helicopter Society,2011,56(2):022006-1-022006-21.

[15] TAN J F,WANG H W. Simulating unsteady aerodynamics of helicopter rotor with panel/viscous vortex particle method[J]. Aerospace Science and Technology,2013,30(1):255-268.

[16] HE C J,ZHAO J G. Modeling rotor wake dynamics with viscous vortex particle method[J]. AIAA Journal,2009,47(4):902-915.

[17] 谭剑锋,王浩文,吴超,等. 基于非定常面元/黏性涡粒子混合法的旋翼/平尾非定常气动干扰研究[J]. 航空学报,2014,35(3):643-656.
TAN J F,WANG H W,WU C,et al. Rotor/empennage unsteady aerodynamic interaction with unsteady panel/viscous vortex particle hybrid method[J]. Acta Aeronautica et Astronautica Sinica,2014,35(3):643-656(in Chinese).

[18] 谭剑锋. 直升机旋翼对尾桨非定常气动载荷的影响[J]. 航空学报,2015,36(10):3228-3240.
TAN J F. Influence of helicopter rotor on tail rotor unsteady aerodynamic loads[J]. Acta Aeronauica et Astronautica Sinica,2015,36(10):3228-3240(in Chinese).

[19] 陈全龙,韩景龙,员海玮. 前行桨叶概念旋翼动力学分析方法[J]. 航空学报,2014,35(9):2451-2460.
CHEN Q L,HAN J L,YUN H W. Analytical method for advancing blade concept rotor dynamics[J]. Acta Aeronautica et Astronautica Sincia,2014,35(9):2451-2460(in Chinese).

[20] 袁野,陈仁良,李攀. 共轴刚性旋翼飞行器配平特性及验证[J]. 南京航空航天大学学报,2016,48(2):186-193.
YUAN Y,CHEN R L,LI P. Trim characteristics and verification of coaxial rigid rotor airfraft[J]. Journal of Nanjing University of Aeronautics & Astronautics,2016,48(2):186-193(in Chinese).

[21] HARRINGTON R D. Full-scale-tunnel investigation of the static-thrust performance of a coaxial helicopter rotor: TN-2318[R]. Washington,D. C.:NASA,1951.

[22] DINGELADEIN R C. Wind-tunnel studies of the performance of multirotor configurations: TN-3236[R]. Washington,D. C.: NASA,1954.

[23] WALSH D,WEINER S,ARIFIAN K,et al. High airspeed testing of the Sikorsky X2 technology TM demonstrator[C]// The 67th Annual Forum of the AHS International. Fairfax,VA: AHS,2011:2999-3010.

[24] 高正,陈仁良. 直升机飞行动力学[M]. 北京:科学出版社,2003:40-42.
GAO Z,CHEN R L. Flight dynamics of helicopter[M]. Beijing:Science Press,2003:40-42(in Chinese).

作者简介:

谭剑锋　男,博士,讲师。主要研究方向:旋翼空气动力学与结构动力学、风机空气动力学。

孙义鸣　男,硕士研究生。主要研究方向:旋翼空气动力学、风机空气动力学。

王浩文　男,博士,教授,博士生导师。主要研究方向:旋翼动力学、结构强度及振动载荷分析。

林长亮　男,博士,高级工程师。主要研究方向:直升机气动设计、总体设计、旋翼结构动力学。

北京航空航天大学学报

北航学报 赠阅

Analysis of rigid coaxial rotor unsteady interactional aerodynamic loads

TAN Jianfeng[1,*], SUN Yiming[1], WANG Haowen[2], LIN Changliang[3]

(1. School of Mechanical and Power Engineering, Nanjing Tech University, Nanjing 211816, China;
2. School of Aerospace Engineering, Tsinghua University, Beijing 100084, China;
3. Institute of Aircraft Design, AVIC Harbin Aircraft Industry Group Co., Ltd., Harbin 150066, China)

Abstract: The maximum forward speed for helicopter increases by adopting rigid coaxial rotor system, while the vibration load in rotor system obviously increases. In order to analyze the vibration characteristics of high-speed rigid coaxial rotor system, it is necessary to investigate unsteady aerodynamic loads of rigid coaxial rotor with aerodynamic interaction. Therefore, a rotor reverse flow aerodynamic model is established based on an unsteady panel method through satisfying boundary conditions of blade leading-edge and trailing-edge to reflect the influence of the reverse flow on the retreating side of the high-speed coaxial rotor. Moreover, a rigid coaxial rotor tip-vortex-blade aerodynamic model is added to describe the influence of aerodynamic interaction between the coaxial rotors. Coupling those models with the wake model of coaxial rotor based on a viscous vortex particle method, an unsteady aerodynamic analysis method under aerodynamic interaction of high-speed rigid coaxial rotor is established. The aerodynamic load at characteristic span of X2 rigid coaxial rotor is simulated during forward flight, and compared with the results of PRASADUM and CFD/CSD based on NASA OVERFLOW and CREATE AV Helios to validate the effectiveness of the present unsteady aerodynamic analysis method. Compared to PRASADUM, the present method better describes the variation characteristics of unsteady airloads of the upper and lower rotors on the advancing and retreading sides, and the results agree better with the computational results of CFD/CSD. Finally, the influence of aerodynamic interaction between the X2 upper and lower rotors on the unsteady aerodynamic loads is analyzed, and the difference of unsteady aerodynamic load between the single rotor and coaxial rotor is also investigated. It is shown that the unsteady aerodynamic load of rigid coaxial rotor is affected obviously by the tip vortex of coaxial rotor at low speed, while it is influenced by coaxial rotor blade at high speed. The characteristic of aerodynamic load of coaxial rotor is radial distribution with number of blades at high speed.

Keywords: rigid coaxial rotor; unsteady aerodynamic load; reverse flow; panel method; viscous vortex particle method

北京航空航天大学学报

Received: 2017-01-18; **Accepted**: 2017-08-11; **Published online**: 2017-09-08 16:14

URL: kns.cnki.net/kcms/detail/11.2625.V.20170908.1614.001.html

Foundation items: National Natural Science Foundation of China (11502105); Natural Science Foundation of Jiangsu Province (BK20161537); The General Project of Natural Science Research of Higher Education Institutions of Jiangsu Province, China (15KJB130004)

* **Corresponding author**. E-mail: Jianfengtan@njtech.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0037

基于抛物方程的低空空域监测雷达城市环境地杂波强度分布建模

雷鹏[1,*]，冉志强[1]，王俊[1]，刘晓敏[2]
（1. 北京航空航天大学 电子信息工程学院，北京 100083； 2. 中国电子科学研究院，北京 100041）

摘 要：地杂波强度是影响雷达低空空域监测性能的重要因素之一。尤其在城市环境下，高层建筑和大气结构将使雷达信号传播及地表电磁散射特性产生复杂变化。提出一种基于抛物方程（PE）的城市环境地杂波强度分布建模方法，能够为低空空域监测雷达系统性能预估、站址选择和杂波特性分析提供理论基础。首先，通过宽角 PE 模型，预测城市高层建筑及大气结构引起的雷达信号反射、绕射、折射和多径效应；其次，将宽角 PE 模型扩展到三维空间，结合雷达方程，实现各杂波单元的强度计算；最后，利用仿真结果分析了不同建筑外形和高层建筑群对雷达信号传播和地杂波强度的影响。

关 键 词：低空空域监测；雷达信号传播；抛物方程（PE）；传播因子；地杂波强度
中图分类号：TN951；TN955；V19
文献标识码：A **文章编号**：1001-5965（2018）01-0063-08

随着低空民用航空器在商业服务中的应用日益增长，低空空域监测已成为完善空中交通安全保障的重要措施之一。特别在城市环境中，民用无人机的蓬勃发展给低空空域监测带来新的挑战。目前，主要的低空空域安全监测方法包括一次雷达、二次雷达、浮空器探测和无源探测等。其中，一次雷达具有架设方便、成本低、可靠性高等优势，通常仅依靠地面设备，独立工作性强，可实现远距离探测，因此，已成为一种典型的低空空域监测手段[1]。

近年来，利用雷达系统进行低空空域目标监测的研究发展迅速，并取得一定成果。文献[1-3]考虑了雷达低空目标测量中海面和复杂地形造成的多径干扰，通过建立相应的多径信号模型，分别利用最大似然估计和稀疏贝叶斯学习方法实现了对低空目标高度和到达角的估计。文献[4]给出了一种低空目标的机载预警雷达回波模型，通过几何关系建立 4 种多径信号传播路径，最终采用概率密度假设和蒙特卡罗方法仿真了低空目标的雷达回波信号。文献[5]以城市复杂环境下的雷达目标监测为目的，利用射线追踪法分析了城市多径效应对雷达目标到达角估计的影响。

然而，除了多径效应，杂波强度也是影响雷达低空目标测量效能的重要因素之一。尤其在城市环境下，建筑物带来的遮挡与起伏、大气折射等因素，将使电磁传播和地表电磁散射特性产生复杂变化[6]。因此，建立真实、灵活的地杂波强度分布模型是进行雷达系统性能预估、站址选择和杂波特性分析的基础。本文针对低空空域监测中城市复杂地形对雷达信号传播的影响，利用宽角抛物方程（Parabolic Equation，PE）模型，考虑城市高层建筑引起的雷达信号反射、绕射和多径效应，

收稿日期：2017-01-18；**录用日期**：2017-02-24；**网络出版时间**：2017-04-05 14:46
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170405.1446.001.html
基金项目：国家自然科学基金（61501011，61671035）
* **通信作者**. E-mail：peng.lei@buaa.edu.cn

引用格式：雷鹏，冉志强，王俊，等. 基于抛物方程的低空空域监测雷达城市环境地杂波强度分布建模[J]. 北京航空航天大学学报，2018，44（1）：63-70. LEI P，RAN Z Q，WANG J，et al. Parabolic equation based land clutter power map modeling for low-altitude surveillance radar in urban areas[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44（1）：63-70（in Chinese）.

以及大气结构造成的折射效应,通过三维高斯型天线方向图获得立体空间内的传播因子;在此基础上,计算雷达杂波单元的后向散射系数和有效照射面积,最终实现地杂波强度分布模型的有效建立。

1 PE 模型

PE 模型是一种确定性传播模型,是由二维标量波动方程得到的一种近似,可应用于海洋[7-8]以及陆地环境[9]的电磁传播特性预测。PE 模型认为电磁波的主能量在抛物线轴向锥形的区域前向传播[9]。考虑直角坐标系下的电磁传播过程,传统的麦克斯韦方程可化简为二维波动方程:

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial z^2}+k_0^2n^2\psi=0 \tag{1}$$

式中:x 为传播区域水平方向;z 为高度方向;$k_0=2\pi/\lambda$ 为真空中的波数,λ 为雷达信号波长;n 为介质中的折射指数。定义传播方向上的波函数为 $u(x,z)=\psi(x,z)\mathrm{e}^{-\mathrm{i}k_0x}$,则式(1)可表示为

$$\frac{\partial^2u}{\partial x^2}+2\mathrm{i}k_0\frac{\partial u}{\partial x}+\frac{\partial^2u}{\partial z^2}+k_0^2(n^2-1)u=0 \tag{2}$$

对式(2)进行因式分解可得

$$\frac{\partial u}{\partial x}=\begin{cases}-\mathrm{i}k_0(1-Q)u & \text{前向传播}\\ -\mathrm{i}k_0(1+Q)u & \text{后向传播}\end{cases} \tag{3}$$

式中:伪微分算子 Q 定义为

$$Q=\sqrt{\frac{1}{k_0^2}\cdot\frac{\partial^2}{\partial z^2}+n^2} \tag{4}$$

式(3)第 1 个式子表征了直角坐标系下的电磁波前向传播的 PE 模型。

常用的 PE 模型求解方法包括有限差分法和分步傅里叶法(Split-Step Fourier Method,SSFM)。后者在求解速度和稳定性上具有明显优势,且迭代步长几乎不受电磁波波长影响[10]。此外,由于城市低空空域监测易受高层建筑等影响,通常需要增加俯仰向范围以实现对指定空域的有效覆盖。

因此,本文的研究将采用 Feit-Fleck 近似法[11]得到的宽角 PE 模型,即

$$\frac{\partial u}{\partial x}=\mathrm{i}k_0\left(\sqrt{1+\frac{1}{k_0^2}\cdot\frac{\partial^2}{\partial z^2}}-1\right)u+\mathrm{i}k_0(n-1)u \tag{5}$$

在此基础上,采用 SSFM 算法得到理想导体边界条件下的 PE 模型的解为

$$u(x+\Delta x,z)=\mathrm{e}^{\mathrm{i}k_0\Delta x(n-1)}\cdot F^{-1}\left(\mathrm{e}^{\mathrm{i}\Delta x\left(\sqrt{k_0^2-p^2}-k_0\right)}F(u(x,z))\right) \tag{6}$$

式中:p 为空间频率。

SSFM 算法的迭代求解过程可以描述为:根据距离向 x 处的场值 $u(x,z)$,结合大气折射项因子以及障碍物绕射项因子,以距离向采样间隔 Δx 为步进迭代单位,利用傅里叶正反变换得到距离向 $x+\Delta x$ 处的场值 $u(x+\Delta x,z)$,依次进行迭代得到整个计算域的场值。其迭代过程如图 1 所示。可以看出使用 SSFM 算法在 PE 模型的迭代求解过程中存在 2 个关键问题:初始场设置、边界条件。

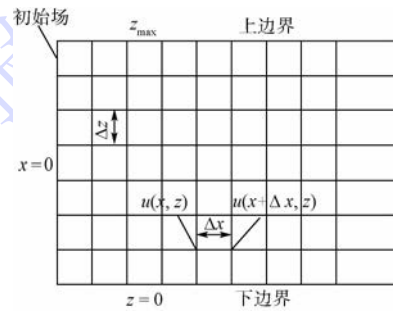


图 1 PE 迭代求解过程

Fig. 1 Iterative procedure of PE solution

由于远场天线方向图 $f(p)$ 与孔径场分布 $u(0,z)$ 之间构成一组傅里叶变换对[12],结合镜像原理可以得到水平和垂直极化情况下地表以上空间初始场对应的傅里叶变换 $f_{g/\!/}(p)$ 和 $f_{g\perp}(p)$ 为:

$$f_{g/\!/}(p)=f(p)\mathrm{e}^{-\mathrm{i}pH}-f^*(-p)\mathrm{e}^{\mathrm{i}pH} \tag{7}$$

$$f_{g\perp}(p)=f(p)\mathrm{e}^{-\mathrm{i}pH}+f^*(-p)\mathrm{e}^{\mathrm{i}pH} \tag{8}$$

式中:H 为天线高度。

通过傅里叶逆变换就可以得到初始场分布。对于高斯型天线方向图有 $f(p)=\mathrm{e}^{-p^2\omega^2/4}$,式中 $\omega=\sqrt{2\ln 2}\Big/\left(k_0\sin\left(\frac{\theta_{3\,\mathrm{dB}}}{2}\right)\right)$,其中 $\theta_{3\,\mathrm{dB}}$ 为天线方向图 3 dB密度。

上边界采用 Tukey 窗函数以满足 Sommerfeld 辐射条件,即电磁波在到达上边界时,不能透射出计算域也不能在上边界发生反射[10]。Tukey 窗函数如式(9)所示,其设置区域为 $\frac{3}{4}z_{\max}\leqslant z\leqslant z_{\max}$。

$$W(z)=\frac{1}{2}+\frac{1}{2}\cos\left(4\pi\left(z-\frac{3}{4}z_{\max}\right)\Big/z_{\max}\right) \tag{9}$$

实际地表对入射到表面的能量有一定的吸收,下边界条件通常用阻抗边界条件[13]来描述,即

$$\frac{\partial u(x,z)}{\partial z}\Big|_{z=0}+\beta' u(x,z)\Big|_{z=0}=0 \tag{10}$$

式中:阻抗系数 β' 为

$$\beta'=\begin{cases}\mathrm{i}k_0\dfrac{\sqrt{\varepsilon_{\mathrm{r}}'-\cos^2\alpha_i}}{\varepsilon_{\mathrm{r}}'} & \text{垂直极化}\\ \mathrm{i}k_0\sqrt{\varepsilon_{\mathrm{r}}'-\cos^2\alpha_i} & \text{水平极化}\end{cases} \tag{11}$$

其中：ε_r'为复相对介电常数；α_i 为掠射角。

对于阻抗边界条件使用离散混合傅里叶变换算法能够有效提高 PE 模型的稳定性与计算效率[14]。

相比于其他电磁波传播特性预测模型，如射线追踪法和波导模理论，PE 模型具有如下主要特点[10]：①包含了传播过程中的反射、折射、绕射等效应，并且不用考虑其发生位置及判定机制；②考虑了大气环境和复杂地形影响；③不仅能够预测点对点的传播关系，还能预测整个计算区域的电磁波传播特性，从而得到空间场分布。

2　雷达地杂波强度分布模型

雷达在城市环境下的低空空域监测性能很大程度上受到地杂波强度的影响。杂波图能够记录雷达站周围环境的杂波强度分布，在目标监测和跟踪等方面具有重要应用。雷达地杂波强度分布主要与雷达信号传播衰减特性以及地表的后向散射能力有关。在城市环境下，首先，通过将 PE 模型理论与地表建筑物特征相结合，建立城市环境下的电磁传播 PE 模型，预测雷达信号的传播衰减特性；其次，根据地表散射强度的相关原理，最终可实现对地杂波强度分布的有效计算。主要步骤如图 2 所示。

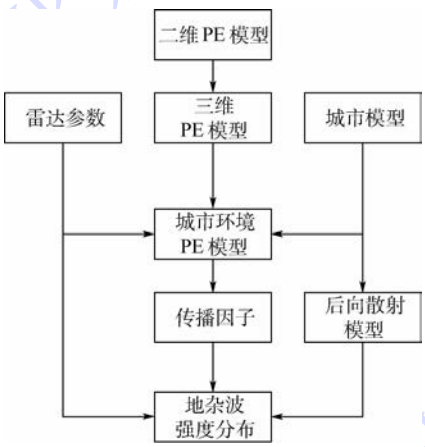


图 2　地杂波强度分布建模流程图

Fig. 2　Flowchart of land clutter power map modeling

2.1　城市环境 PE 模型

人造建筑遮挡、建筑表面和地面反射、大气折射将是影响城市环境下雷达信号传播衰减特性的主要因素。因此，本文在使用 PE 模型对城市环境下的电磁波传播特性进行预测时，将结合城市建筑模型、阻抗边界条件和大气折射指数影响进行俯仰向上的传播因子计算，在此基础上，通过方位采样处理得到最终的三维 PE 模型。

2.1.1　城市建筑模型

建筑物遮挡是城市环境中雷达信号传播衰减的主要因素。常见的城市建筑模型包括阶梯形、矩形、帆形、橄榄形等，其二维剖面简化模型如图 3所示。此外，城市建筑在空间上多呈现为带状或集团状。

由于计算地杂波强度分布需要在雷达坐标系下进行，因此，需将包含城市建筑模型的地面网格变换到雷达坐标系下完成。具体处理步骤将在 2.2 节阐述。

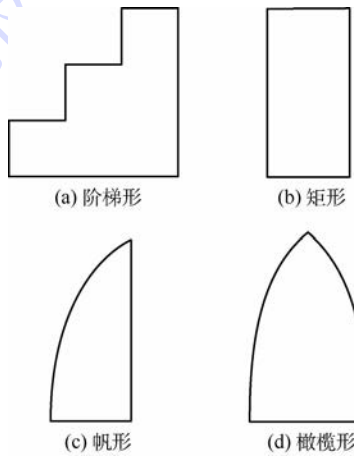


图 3　简化建筑模型

Fig. 3　Simplified building models

2.1.2　边界平移法

边界平移法[15]能够在 PE 模型的应用中有效地考虑建筑物的高低起伏。边界平移法根据前后步进海拔差，在高度向上进行场值平移，实现过程简单，在大坡度条件下同样适用。

边界平移法在帆形建筑的实现过程如图 4 所示。

高度差确定的点数为 N_t。当坡度为正时，由 X_1 处的场迭代计算得到的 X_2 处的场将向下移动 N_t 个单元，即 X_1 处的 u_{N_t+1} 对应 X_2 处的 u_1，依次类推。当坡度为负时，场值将向上移动 N_t 个单元。边界平移法不需要对地形做转换处理，不必

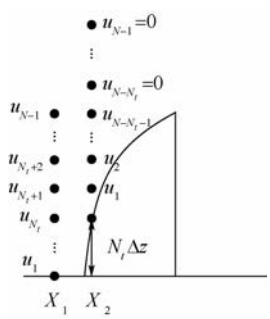


图 4　边界平移法

Fig. 4　Boundary shift method

北航学报 赠阅

改变大气折射指数项因子以及障碍物绕射项因子,实现过程最为简单。能够有效解决复杂的大气环境中由建筑物高低起伏所引起的不规则地形条件下的雷达信号传播问题。

2.1.3 大气结构影响

大气结构是影响雷达信号传播的重要因素之一,其对电磁传播特性的作用主要由大气折射指数 n 确定。随着高度改变,大气折射指数会改变雷达信号在介质中的传播速度,从而使得雷达信号在大气传播中产生折射,当雷达信号传播距离较远的时候,需要考虑地球曲率影响,可采用修正的大气折射率 M 将地球曲面近似处理为平面。

$$\frac{\mathrm{d}M}{\mathrm{d}z}=\frac{\mathrm{d}K}{\mathrm{d}z}+0.157 \tag{12}$$

式中:K 为大气折射率;$\mathrm{d}/\mathrm{d}z$ 为高度向上的折射率梯度。修正的大气折射率在高度向上的梯度不同,雷达信号在大气中会产生不同的折射效应。

2.1.4 三维 PE 模型

传统的二维 PE 模型仅考虑了雷达信号在给定方位向上的传播特性。对于城市环境雷达信号传播,需要预测整个三维区域的雷达信号覆盖情况,因此需要利用二维 PE 模型解决三维的估计问题。

为了获取雷达信号在方位向上的传播特性,本文采用文献[16-17]方法将传统的二维 PE 模型扩展到三维空间。在三维 PE 模型中,雷达位置不变,根据天线指向和波束宽度,在有效方位角范围内,按照固定方位向间隔形成一系列的二维空间计算域切片,从而提取对应方位向上的地形,在每一个地形上分别利用二维 PE 模型进行求解。其中,三维 PE 模型中初始场的设置根据三维高斯天线方向图,分别在每一个角度间隔单元上获取方向图切片进行傅里叶逆变换得到对应二维空间的初始场分布。可见,在该三维 PE 模型方法中,相邻切片对应的二维计算区域之间并无电磁传播影响。

2.2 地杂波强度计算

根据雷达方程,地杂波强度与雷达系统参数、目标雷达散射截面积(Radar Cross Section, RCS)、传播损耗等因素有关,可表示为[18]

$$P_{\mathrm{c}}=\frac{P_{\mathrm{t}}G^2\lambda^2\sigma_0A_0F^4}{(4\pi)^3R^4L} \tag{13}$$

式中:P_{c} 为杂波功率;σ_0 为杂波单元平均反射系数;A_0 为雷达有效照射面积;F^2 为传播因子;P_{t} 为雷达发射功率;G 为雷达增益;L 为系统损耗;R 为目标径向距离。

考虑杂波单元大于地面网格单元的场景,本文将采用正向法[19]将地面直角坐标系转换到雷达坐标系,如图 5 所示。

使用正向法进行坐标系转换时,需要在雷达极坐标系下求出雷达杂波单元所包含的地面杂波块面积和后向散射系数等参数,再通过求和方式得到整个雷达杂波单元的地杂波强度。

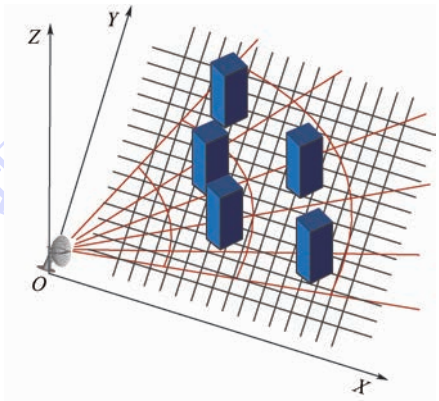


图 5 直角坐标系到雷达极坐标系的正向转换

Fig. 5 Forward transformation from Cartesian grid to radar polar coordinate system

2.2.1 后向散射模型

雷达散射截面积 σ_{RCS} 与杂波单元平均反射系数 σ_0、雷达有效照射面积 A_0 的关系可以表示为:$\sigma_{\mathrm{RCS}}=\sigma_0A_0$。不同媒质对雷达信号的后向散射能力通常不同。本文采用 Constant-γ 模型进行后向散射系数的计算,其表达式为 $\sigma_0=\gamma\sin\alpha$,$\alpha$ 为入射余角,γ 为散射系数。

2.2.2 有效照射面积

在地基雷达低空空域监测条件下,擦地角通常较小,因此,天线波束照射范围一般大于雷达脉冲距离门对应范围[20],如图 6 所示。β 为建筑曲面坡度,φ 为擦地角,η 为建筑曲面法向量与雷达指向之间的夹角,$\Delta\theta$ 为雷达天线波束的方位向宽度,B 为发射信号带宽,c 为光速。

根据图 6 所示关系,照射面积计算公式为

$$A=\frac{Rc\Delta\theta}{2B\cos\varphi} \tag{14}$$

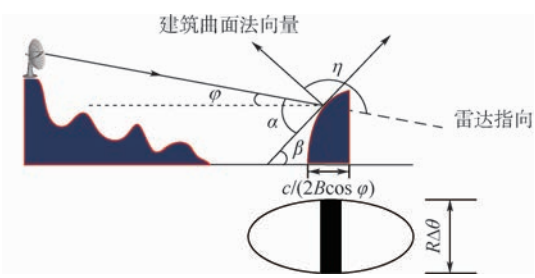


图 6 照射面积计算示意图

Fig. 6 Illustration of clutter irradiated area calculation

上述面积为雷达波束在水平面上的投影面积,在计算 RCS 时,应求取系统的有效照射面积。由于建筑物高低起伏影响,雷达有效照射面积需要根据建筑物参数和擦地角等,由投影面积 A 得到有效照射面积 A_0,即

$$A_0 = \frac{A}{\cos\beta}|\cos\eta| = \frac{A}{\cos\beta}|\sin\alpha| \tag{15}$$

3　仿真结果与分析

设雷达高度为 100 m,雷达发射频率为3 GHz,最大传播仰角为 15°,采用高斯型天线方向图,波束宽度为 2°,大气结构为标准大气环境。

在 PE 模型中,环境因素对电磁波传播的影响主要通过不同的电磁参数体现,包括复相对介电常数 ε_r'以及电导率 σ。在本节仿真中,假设建筑表面材料为良导体,其复相对介电常数 ε_r' = 80 + i79. 1[16],地表媒质假设为中等干燥地表类型。

对于二维 PE 模型,设建筑物与雷达水平距离为 10 km,径向宽度为 200 m,帆形建筑的曲线轮廓满足如下方程:

$$t(x') = 5\sqrt{x' - 10^4} \tag{16}$$

其中:x'表示曲线轮廓上各点与雷达之间的水平径向距离。为保证矩形建筑与帆形建筑高度相同,根据式(16),设置矩形建筑高度约为 70.7 m。

图 7 和图 8 分别给出了在矩形和帆形建筑条件下基于二维 PE 模型的雷达信号传播特性预测结果。观察可知,在建筑物之前的区域,雷达信号传播主要受大气波导折射效应影响,传播因子随高度在 0 值附近振荡变化,对应于图 7(a)和图 8(a)中明暗相间的条状分布。在建筑物遮挡区内,传播因子大幅衰减,表明建筑物对于雷达信号传播具有显著的阻碍作用。但由于大气折射、信号绕射等现象的存在,建筑物并未完全阻断雷达信号传播,并且部分遮挡区域的雷达信号强度依然较强。此外,在约 70 m 以上的高度区域,遮挡逐渐减弱,大气波导折射效应再次成为影响雷达信号传播的主要因素,此时传播因子变化趋势与建筑物之前区域相同,即在 0 值附近振荡。建筑外形对传播因子的影响则主要表现为遮挡区域衰减强度的不同。比较图 7(b)和图 8(b)可知,矩形建筑比帆形建筑对电磁波传播的衰减更大,但在建筑高度以上的区域,衰减特性几乎相同。

对于三维 PE 模型,考虑雷达天线方位向变化范围为 -30° ~ 30°,以 0°为中心,左右变化间隔

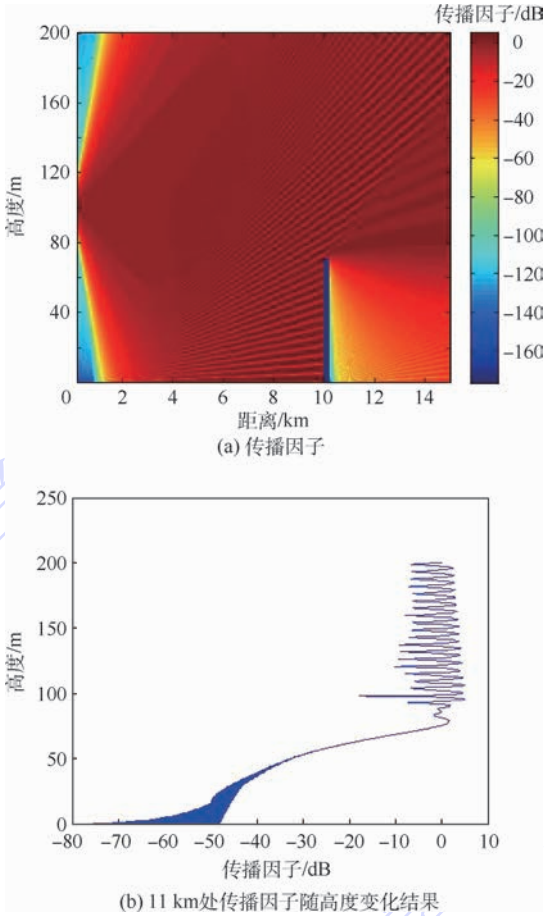


图 7　矩形建筑条件下的雷达信号传播特性

Fig. 7　Propagation characteristics of radar signals over a rectangular building

均为 2°,且在每一方位指向上的二维 PE 计算域切片间隔为 0. 5°;设 2 组建筑群中心分别处于 ±2°方位向上,其各包含 9 栋独立建筑,如图 9 所示。各建筑的方位向宽度为 60 m,径向宽度为 200 m,距离雷达位置分别为 9、10 和 11 km;帆形建筑剖面方程为

$$t(x') = (j+2)\sqrt{x' - 10^3(8+j)} \tag{17}$$

式中:j = 1,2,3,分别对应第 1 ~ 3 排帆形建筑。同理于单建筑场景,根据式(17),选取对应的每排矩形建筑高度分别约为 42.4、56.6 和 70.7 m。

图 10 给出了不同形状建筑群条件下的雷达信号传播特性预测结果。该结果表明,经多个建筑物遮挡,雷达信号传播损耗更大。此外,帆形建筑对信号传播的绕射效应较矩形建筑更为明显,从而导致帆形建筑前表面部分以及遮挡区域的雷达信号传播因子大于矩形建筑。而在建筑物高度以上区域的传播衰减则主要受大气折射效应影响。

设雷达发射功率为 200 kW;根据文献[21],对于城市建筑,取 Constant-γ 模型中 γ = -5 dB。

北航学报 赠阅

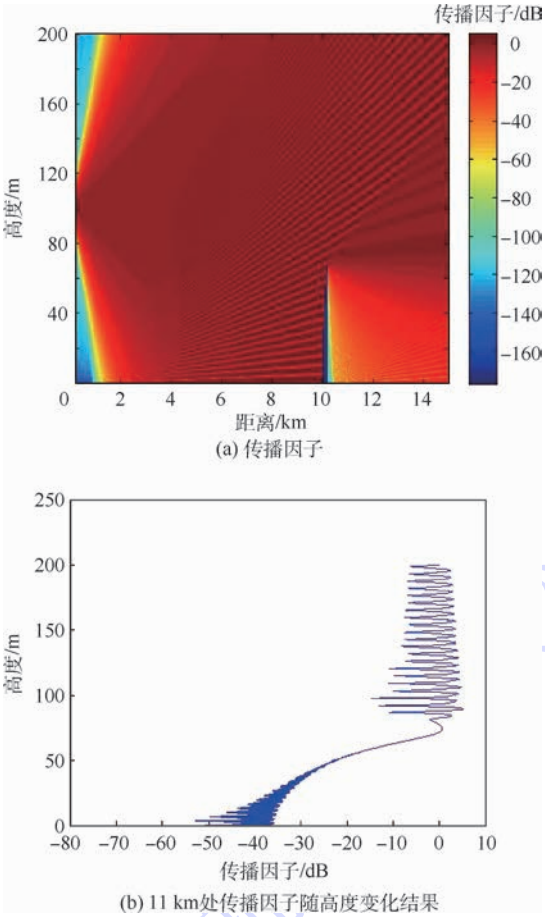


图 8 帆形建筑条件下的雷达信号传播特性

Fig. 8 Propagation characteristics of radar signals over a sail-shaped building

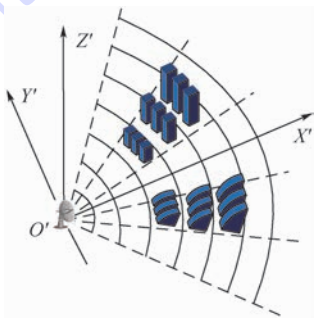


图 9 建筑群分布示意图

Fig. 9 Illustration of building distribution

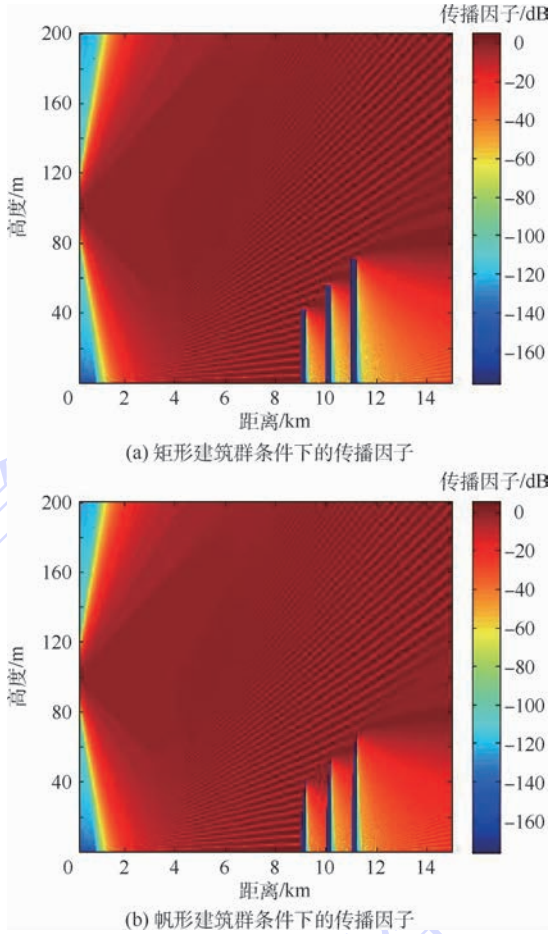


图 10 城市建筑群环境下的雷达信号传播特性

Fig. 10 Propagation characteristics of radar signals over urban buildings

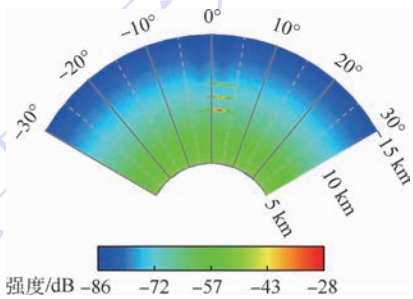


图 11 地杂波强度分布结果

Fig. 11 Result of land clutter power map

图 11 给出了雷达极坐标系下的地杂波强度分布结果。为了突出局部细节特征,这里采用 1∶2 比例尺将分布结果方位向角度进行均匀放大。由于地表和大气结构的同质性,无人造建筑区域的地杂波强度在同等距离条件下相同。而当存在建筑物时,雷达天线与建筑的空间几何关系、传播因子大小差异等因素导致帆形建筑区域的杂波强度明显增大。此外,因为建筑物遮挡区域面积小于杂波单元面积,并且遮挡区域的地杂波强度非常小,所以其所在杂波单元中其他地表区域对地杂波的贡献更为显著。

4 结 论

本文以地基雷达低空空域监测为背景,研究了基于 PE 的城市地杂波强度分布建模方法,其结果表明:

1) 由于反射、绕射和多径效应的存在,城市建筑遮挡虽然会对雷达信号传播产生明显的衰减作用,但不会完全阻断,并且建筑越高,衰减效果越明显。

2）在同等条件下,矩形建筑遮挡比帆形建筑遮挡对雷达信号的衰减作用更大。

3）在地基雷达低空空域监测条件下,传播因子是影响地杂波强度的重要因素之一。

参考文献（References）

[1] 陈唯实,宁焕生. 利用一次雷达实现低空空域的安全监视[J]. 北京航空航天大学学报,2012,38(2):143-148.
CHEN W S, NING H S. Security surveillance of low-altitude airspace with primary radar[J]. Journal of Beijing University of Aeronautics and Astronautics, 2012, 38(2): 143-148(in Chinese).

[2] TAKAHASHI R, HIRATA K, MANIWA H. Altitude estimation of low elevation target over the sea for surface based phased array radar[C]// Proceedings of IEEE Radar Conference. Piscataway, NJ: IEEE Press, 2010: 123-128.

[3] ZHENG Y S, CHEN B X. Altitude measurement of low-angle target in complex terrain for very high-frequency radar[J]. IET Radar Sonar Navigation, 2015, 9(8): 967-973.

[4] MAN M Y, LEI Z Y, XIE Y J, et al. Monte Carlo simulation of the echo signals from low-flying targets for airborne radar[J]. International Journal of Antennas & Propagation, 2014: 416985.

[5] GREENBERG E, NAOR M. Direction of arrival estimation in urban multipath environments[C]// Proceedings of European Conference on Antennas and Propagation. Piscataway, NJ: IEEE Press, 2016: 1-5.

[6] ANDERSON K D. Radar detection of low-altitude targets in a maritime environment[J]. IEEE Transactions on Antennas and Propagation, 1995, 43(6): 609-613.

[7] SIRKOVA I. Brief review on PE method application to propagation channel modeling in sea environment[J]. Central European Journal of Engineering, 2012, 2(1): 19-38.

[8] ZHAGN P, BAI L, WU Z, et al. Applying the parabolic equation to tropospheric groundwave propagation: A review of recent achievements and significant milestones[J]. IEEE Antennas & Propagation Magazine, 2016, 58(3): 31-44.

[9] DONOHUE D J, KUTLER J R. Modeling radar propagation over terrain[J]. Johns Hopkins APL Technical Digest, 1997, 18(2): 279-287.

[10] 胡绘斌. 预测复杂环境下电波传播特性的算法研究[D]. 长沙:国防科学技术大学,2006:4-8.
HU H B. Study on the algorithms of predicting the radio propagation characteristics in complex environments[D]. Changsha: National University of Defense Technology, 2006: 4-8(in Chinese).

[11] THOMSON D J, CHAPMAN N R. A wide-angle split-step algorithm for the parabolic equation[J]. Journal of the Acoustical Society of America, 1983, 74(6): 1848-1854.

[12] BARRIOS A E. A terrain parabolic equation model for propagation in the troposphere[J]. IEEE Transactions on Antennas and Propagation, 1994, 42(1): 90-98.

[13] LI L, LIN L K, WU Z S, et al. Study on the maximum calculation height and the maximum propagation angle of the troposcatter wide-angle parabolic equation method[J]. IET Microwaves Antennas & Propagation, 2016, 10(6): 686-691.

[14] KUTTLER J R, JANASWAMY R. Improved Fourier transform methods for solving the parabolic wave equation[J]. Radio Science, 2002, 37(2): 1-11.

[15] BARRIOS A E. Considerations in the development of the advanced propagation model (APM) for U. S. navy applications[C]// Proceedings of the International Radar Conference. Piscataway, NJ: IEEE Press, 2003: 77-82.

[16] AWADSALLAH R S, GEHMAN J Z, KUTTLER J R, et al. Modeling radar propagation in three-dimensional environments[J]. Johns Hopkins APL Technical Digest, 2004, 25(2): 101-111.

[17] NORMAN E D. Assessment of the wind farm impact on the radar: 1002. 2654v1[R]. Limours: Thales Air Systems, 2010: 1-55.

[18] LIN C C, REILLY J P. A site-specific model of radar terrain backscatter and shadowing[J]. Johns Hopkins APL Technical Digest, 1997, 18(3): 432-447.

[19] 米切尔. 雷达系统模拟[M]. 北京:科学出版社,1982:124-127.
MITCHELL R L. Radar signal simulation[M]. Beijing: Science Press, 1982: 124-127(in Chinese).

[20] 许小剑,黄培康. 雷达系统及其信息处理[M]. 北京:电子工业出版社,2010:156-158.
XU X J, HUANG P K. Radar system and signal processing[M]. Beijing: Publishing House of Electronics Industry, 2010: 156-158(in Chinese).

[21] FENG S, CHEN J. Law-angle reflectivity modeling of land clutter[J]. IEEE Geosciences and Remote Sensing Letters, 2006, 3(2): 254-258.

作者简介:

雷鹏　男,博士,讲师。主要研究方向:信号处理、模式识别。

北航学报 赠阅

Parabolic equation based land clutter power map modeling for low-altitude surveillance radar in urban areas

LEI Peng[1,*], RAN Zhiqiang[1], WANG Jun[1], LIU Xiaomin[2]

(1. School of Electronic and Information Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China;
2. China Academy of Electronics and Information Technology, Beijing 100041, China)

Abstract: Land clutter power exerts a strong impact on the performance of low-altitude surveillance radar. Especially in urban areas, skyscrapers and atmosphere conditions complicate the radar signal propagation and the electromagnetic scattering from land surfaces. This paper presents a land clutter power map modeling approach for low-altitude surveillance radar in urban areas by using the parabolic equations (PE). It could theoretically contribute to the performance prediction and field deployment of such radar systems as well as the analysis of urban clutter characteristics. The proposed approach exploits wide-angle PE to take into account the reflection, diffraction, refraction and multipath effects in low-grazing-angle radar propagation related to tall buildings and atmosphere conditions. After the 3D approximation of aforementioned 2D wide-angle PE, the propagation factors could be obtained in the 3D environment. Then according to the radar equation, the power in every clutter cell is calculated. Finally, numerical simulations are carried out to demonstrate the influence of different architectural appearances and urban skyscrapers on radar signal propagation and land clutter powers.

Keywords: low-altitude surveillance; radar signal propagation; parabolic equation (PE); propagation factor; land clutter power

北京航空航天大学学报

Received: 2017-01-18; **Accepted**: 2017-02-24; **Published online**: 2017-04-05 14:46

URL: kns.cnki.net/kcms/detail/11.2625.V.20170405.1446.001.html

Foundation items: National Natural Science Foundation of China (61501011, 61671035)

* **Corresponding author**. E-mail: peng.lei@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0027

基于分布式动力的翼身融合飞机整流罩气动设计

项洋，吴江浩*，熊峻江

（北京航空航天大学 交通科学与工程学院，北京 100083）

摘 要：整流罩设计对基于分布式动力的翼身融合(BWB)飞机气动特性会产生显著影响。为了揭示在边界层吸入(BLI)效应下整流罩的设计参数对飞机气动特性的影响及其原因，采用计算流体力学(CFD)方法和 Morris 敏感度分析法对此布局飞机气动特性进行了详细研究，得到了整流罩主要设计参数对飞机气动特性影响的敏感度和耦合关系，并对典型设计参数下的流动特性进行分析。结果表明：对飞机气动特性影响较大的参数是整流罩特征截面 2 和 3 的最大厚度，这是因为其增大了当地截面的厚度和弯度，进而影响了整流罩表面的压力分布；在流量系数减小和进气边界弦向位置前移时，最大厚度增大会造成背风面发生局部分离；整流罩特征截面 2 和 3 的最大厚度对气动特性具有较强的耦合影响。

关 键 词：翼身融合(BWB)布局；边界层吸入(BLI)；计算流体力学(CFD)；敏感度分析；整流罩

中图分类号：V221.3

文献标识码：A **文章编号**：1001-5965(2018)01-0071-11

近年来，翼身融合(Blended-Wing-Body，BWB)布局飞行器由于具有提高气动效率、降低油耗、排放和噪声等方面的潜力[1-3]而逐渐成为研究热点，是一种有希望取代常规布局的未来民机布局形式。为了能进一步发挥其优势，研究者提出在 BWB 布局上应用一种被称为分布式动力(distributed propulsion)的新的发动机布局形式[4-6]。该布局相比传统的集中吊挂式布局，具有更低的冲压阻力、浸润面积、结构重量和噪声等多项潜在优势。BWB 与分布式动力系统耦合布局的特点是在 BWB 机身上表面后部沿翼展方向布置若干个进气口靠近或半埋入机身的发动机或风扇，这会使机身上表面的边界层被发动机摄入，即产生了边界层吸入(Boundary Layer Ingestion，BLI)效应。该效应会使得分布式动力 BWB 飞机的气动特性显著不同于集中吊挂式布局[7-10]。已有一些研究者针对这一问题展开了相关研究。

Rodriguez[11]研究了带 3 台发动机的 BWB 飞机动力系统的布置形式，比较了吊挂式和埋入式进气口对气动特性的影响，并进行了初步的多学科优化设计，研究表明采用埋入式进气口布局的飞机具有较高的推进效率和气动效率。Lundbladh 和 Grönstedt[12]研究了不同进气形式的分布式动力 BWB 飞机在巡航条件下的效率问题，发现埋入式进气口对降低飞机的起飞重量和油耗、提升巡航效率是有利的。Liou 等[13]对分布式动力 BWB 飞机整流罩的气动设计问题进行了初步研究，发现在重新设计整流罩形状后，整流罩表面的激波减弱，全机升力系数提高。闫万方等[14]研究了部分设计参数对分布式动力 BWB 飞机全机气动特性的影响，发现在这些参数中，流量系数(Mass Flow Rate，MFR)和进气口弦向位置对飞

收稿日期：2017-01-16；**录用日期**：2017-02-06；**网络出版时间**：2017-03-22 11:40
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170322.1140.001.html
* **通信作者**. E-mail：buaawjh@buaa.edu.cn

引用格式：项洋，吴江浩，熊峻江. 基于分布式动力的翼身融合飞机整流罩气动设计[J]. 北京航空航天大学学报，2018，44(1)：71-81. XIANG Y，WU J H，XIONG J J. Aerodynamic design of nacelle of blended-wing-body aircraft with distributed propulsion[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(1)：71-81 (in Chinese).

北航学报 赠阅

机气动特性影响显著。

相比吊挂式发动机布局,分布式动力 BWB 飞机具有更长、面积更大的整流罩,且更靠近机身,会影响到飞机上表面流场,因此研究其对飞机机体气动特性的影响是有意义的。本文采用数值模拟与敏感度分析相结合的方法对分布式动力 BWB 飞机的整流罩设计参数进行研究,获得其对飞机气动特性影响的敏感度并分类,从中选择典型参数分析单变量对气动特性影响的原因,并分析多参数对气动特性的耦合影响。

1 模型和方法

1.1 模型定义

本文所研究的分布式动力 BWB 飞机的气动布局如图 1 所示[14]。该布局由中心体、融合段和外翼段组成,在中心体后部沿展向布置有分布式动力系统。动力系统前后部分别为进、排气边界,上表面为整流罩。飞机的主要几何参数标注于图 1,其展长 $b=68.2$ m,机身对称面弦长(即机身长度) $c_{\text{root}}=44.6$ m,重心到机头的距离 $L_{\text{cg}}=27.3$ m。动力系统的位置和外形尺寸取值如下(均以与机身对称面弦长 c_{root} 的比值表示):进气边界弦向位置(定义为动力系统进气边界到机头的距离) $L_{\text{f}}=0.8c_{\text{root}}$,整流罩长度 $L_{\text{nacelle}}=0.13c_{\text{root}}$,进气边界高度 $H=0.02c_{\text{root}}$,整流罩展向长度 $b_{\text{nacelle}}=0.48c_{\text{root}}$。

对于整流罩的参数化描述,本文参照文献[15]给出整流罩外形的定义方式。从对称面沿展向等距离布置 4 个控制截面作为设计变量,每个截面的几何参数定义如图 2 所示。

从图 2 可见,整流罩截面外形由前后 2 段组成。前段外形的设计参照传统的发动机整流罩,为 NACA-1 系翼型[16],其长度为 L_{front},高度为 $H_{\max}-H$($H_{\max}$ 为整流罩的最大高度)。后段为圆弧,半径为 R_{aft},表达式为

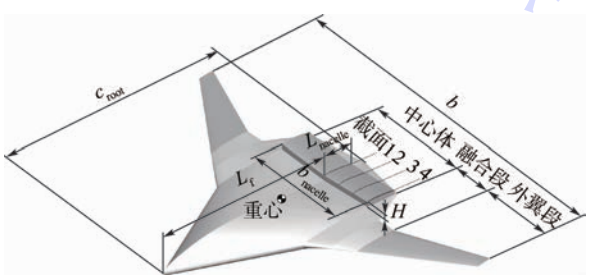


图 1 分布式动力 BWB 飞机的气动布局

Fig. 1 Aerodynamics configuration of distributed propulsion BWB aircraft

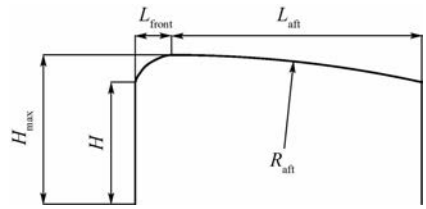


图 2 整流罩截面几何参数定义

Fig. 2 Geometric parameter definition of nacelle section

$$R_{\text{aft}}=\frac{L_{\text{aft}}}{\sin\left(2\dfrac{H_{\max}-H}{L_{\text{aft}}}\right)} \tag{1}$$

在本研究中,给定动力系统的 H 和整流罩长度 L_{nacelle}($L_{\text{nacelle}}=L_{\text{front}}+L_{\text{aft}}$)不变,并选定 L_{nacelle} 作为无量纲参考长度。于是可以定义 2 个新的几何参数:整流罩厚度 t 和整流罩厚度位置 x。

$$t=\frac{H_{\max}-H}{L_{\text{nacelle}}} \tag{2}$$

$$x=\frac{L_{\text{front}}}{L_{\text{nacelle}}} \tag{3}$$

由上述定义可知,由每个截面的 t 和 x 的组合可唯一确定该截面形状。整流罩曲面形状由这 4 个截面放样给出,可知由 t_i 和 x_i($i=1,2,3,4$)总共 8 个参数即可唯一确定整流罩形状。为了便于研究,需要选取一个基准构型。参照文献[15]中对二维情形下整流罩参数进行研究,设定基准构型的 $t_i=0.047$,$x_i=0.29$($i=1,2,3,4$)。此外,为确定动力系统的边界条件,还需用到流量系数 MFR 和排气方向 τ。MFR 的定义为

$$\text{MFR}=\frac{\dot{m}}{\rho_\infty V_\infty H b_{\text{nacelle}}} \tag{4}$$

式中:$\dot{m}$ 为流入进气口的空气的质量流率;ρ_∞ 为远前方来流的密度;V_∞ 为远前方来流的速度。基准条件下 MFR = 0.68。排气方向 τ 的定义为排气气流方向与排气边界法向的夹角,当排气方向斜向上时为正值,基准条件下 $\tau=0°$。

1.2 数值方法及验证

本文的数值计算以三维定常 RANS 方程作为控制方程,使用有限体积法对其进行空间离散,并使用具有二阶精度的耦合方法求解,湍流模型为一方程 S-A 模型。动力系统进气口的边界条件类型为压力出口,动力系统喷口的边界条件类型为质量流量入口。计算所使用的网格模型为 C-H 型网格,其结构如图 3 所示。

本文研究中出现的气动力和力矩系数包括升力系数 C_L、阻力系数 C_D 和俯仰力矩系数 C_M,其定义分别为:$C_L=L/(0.5\rho_\infty V_\infty^2 S_{\text{ref}})$,$C_D=D/(0.5\rho_\infty V_\infty^2 S_{\text{ref}})$,$C_M=M/(0.5\rho_\infty V_\infty^2 S_{\text{ref}} c)$。式中:

北航学报 赠阅

L、D 和 M 分别为飞机的升力、阻力和俯仰力矩(包含整流罩表面);力矩 M 的取矩点为重心;c 为飞机的平均气动弦长;S_{ref} 为飞机的参考面积。

以下对本文所用的计算网格进行验证。为确定合适的网格规模,对网格规模递增的 3 套网格分别进行计算,对比结果见表 1,基准计算条件为飞行高度 $H_{\mathrm{flight}}=11\,000$ m,马赫数 $Ma=0.85$,迎角 $\alpha=2.8°$。可见,网格 1 的计算结果与网格 3 相差较大,而网格 2 的计算结果与网格 3 较为接近,且计算开销小于网格 3。

另外,需要选取合适的壁面网格高度,因为壁面网格高度足够小才能准确地描述边界层内流动,但过小的壁面网格高度会降低网格质量。对 3 套规模相同、壁面网格高度递减的网格计算结果进行了对比,对比结果如表 2 所示。可见,网格 4的计算结果与网格 2 相差较大,而网格 2 与网格 5 的计算结果较为接近。从表 1 和表 2 的验证结果可以看出,网格 2 的数值模拟结果在精度与效率之间取得了平衡,因此本文选取网格 2 作为数值模拟的基准网格。

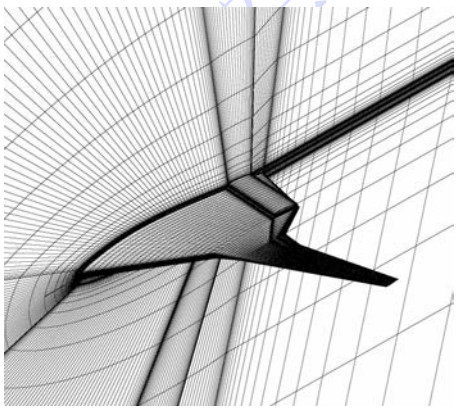


图 3　计算网格结构

Fig. 3　Structure of computation mesh

表 1　网格规模验证

Table 1　Validation of mesh size

算例名称	网格规模	C_L	C_D
网格 1	907 563	0.411 2	0.021 63
网格 2	1 378 251	0.413 7	0.020 71
网格 3	2 055 448	0.416 0	0.020 56

表 2　壁面网格高度验证

Table 2　Validation of wall grid height

算例名称	壁面网格高度/m	C_L	C_D
网格 2	5×10^{-4}	0.413 7	0.020 71
网格 4	1×10^{-3}	0.414 8	0.020 52
网格 5	3×10^{-4}	0.413 9	0.020 76

2　结果和讨论

2.1　参数敏感度分析

本文采用 Morris 法分析整流罩外形设计参数对飞机气动系数的敏感度。Morris 法是一种全局敏感度分析法,具体流程为:①将 n 个参数的定义域映射到[0,1]区间并离散化为 p 个水平,构成 n 维 p 水平的采样空间;②选定重复采样次数 r;③进行 r 次随机抽样,每次抽样都可以获得 1 组 n 个参数的基本效应(elementary effect)值;④对每个参数,计算 r 个基本效应的平均值 d 和标准差 S,这 2 个量分别表示该参数对目标函数的敏感度和该参数与其他参数的耦合作用程度[17]。

根据 Morris 法的步骤,首先选择 1.1 节给出的决定整流罩形状的 8 个参数作为研究对象,并给出这 8 个参数在各自的基准值 P_i 附近的变化区间。由于各参数的量纲和量级可能各异,因此,需要对参数的变化区间作归一化[15]。本文将归一化后的参数变化区间称为参考变化区间,第 i 个参数的参考变化区间 $[l^*,u^*]_i$ 定义为

$$[l^*,u^*]_i=\frac{[l,u]_i}{P_i^{\mathrm{ref}}} \tag{5}$$

式中:P_i^{ref} 为第 i 个参数对应的参考特征量;u、l 分别为该参数实际变化区间的上、下界。本文以整流罩长度 L_{nacelle} 作为整流罩各截面厚度和厚度位置对应的参考特征量。考虑几何约束和参数的一般变化范围,将参数的参考变化区间上下界设为 6% 和 −4%,由此可知参数的实际变化区间,如表 3所示。

此外还应为 Morris 法选定合适的抽样次数,若抽样次数太小,则无法获得有意义的统计平均值;反之则会导致过大的计算开销。综合考虑这 2 个因素,并参考二维翼型 Morris 敏感度分析的经验[15],将本研究的抽样次数取为 $r=50$。

使用 Morris 法计算了上述 8 个参数在巡航状态下(条件见 1.2 节)对气动系数的敏感度和耦合作用,计算结果如图 4 所示。

从图中可知,总体而言,各截面厚度 t_i 对气动系数影响较大,厚度位置 x_i 影响较小,这与二

表 3　参数的实际变化区间

Table 3　Actual changing interval of parameters

参　数	P_i	$[P_i-l,\ P_i+u]$
$t_i(i=1,2,3,4)$	0.047	[0.007,0.107]
$x_i(i=1,2,3,4)$	0.29	[0.25,0.35]

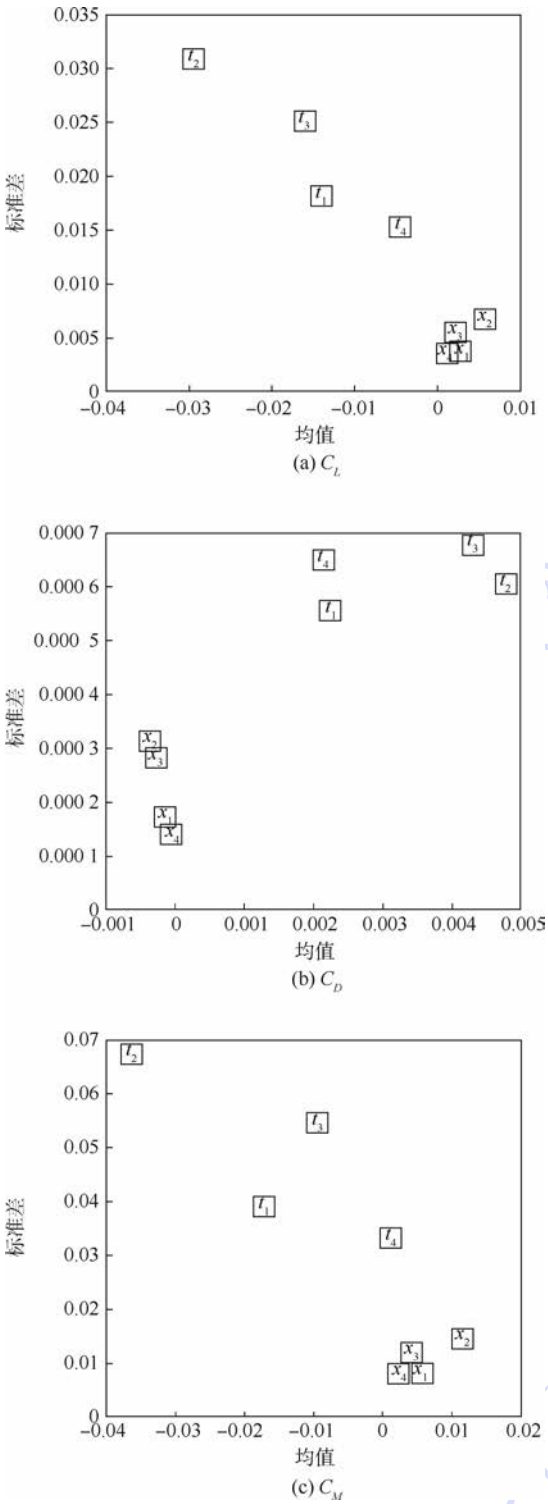


图 4 Morris 法计算的各参数对气动系数影响的均值与标准差

Fig. 4 Mean values and standard deviations of effects of parameters on aerodynamic coefficients with Morris method

维结果是一致的。从均值绝对值的对比可以看出,对 C_L 敏感度较大的参数排序为 t_2、t_3、t_1、x_2;对 C_D 敏感度较大的参数排序为 t_2、t_3、t_1、t_4;对 C_M 敏感度较大的参数排序为 t_2、t_1、x_2、t_3。从标准差的对比可以看出,对气动系数影响较大的参数与其他参数的相互作用也较明显,反之亦然,这与二维研究的结果也是一致的。另外,从图 4 中均值的符号可以看出,C_L 敏感度较大的参数对 C_L 的基本效应均值为负数。

图 5 为参数 t_2 对 C_L 的基本效应分布图,可以看出在样本中的大多数点处,基本效应是绝对值较大的负值,即在大多数点处,厚度的增加造成升力减少。这是因为这些点所对应的构型下,整流罩厚度的变化跨越了线性和非线性区间的分界点,厚度的增加会使得整流罩发生局部失速,或者使得失速区域扩大,从而造成升力的下降。这与采用了相同整流罩构型的二维翼型 Morris 法分析结果中,C_L 敏感度较大的参数对 C_L 的基本效应均值为正数[15]不同。因此在本设计中,为了获得较好的气动特性,应当对三维整流罩的厚度加以限制使其保持在线性区间。

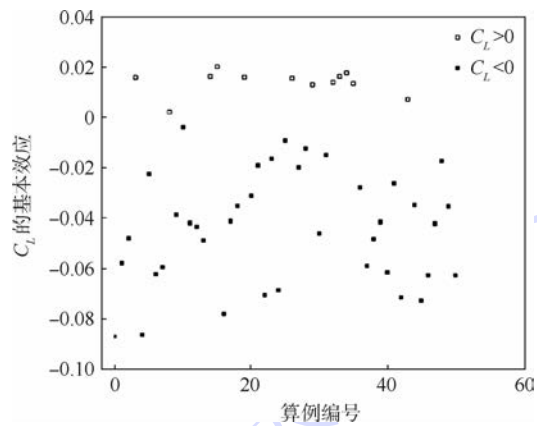


图 5 t_2 对 C_L 的基本效应分布

Fig. 5 Distribution of C_L elementary effect with respect to t_2

2.2 整流罩设计参数对气动特性的影响

2.2.1 截面厚度 t_2 的影响

t_2 是整流罩参数中对所有气动系数敏感度最大的一个。基准构型下,气动系数随 t_2 的变化规律见表 4。可见,随 t_2 的增大,C_L、C_D 和 C_M 都单调增大。

因整流罩的剖面形状类似圆头尖尾的低速翼型,在分析整流罩几何参数对气动特性的影响时,可以把整流罩视为 BWB 飞机机身后部上表面固定放置的一段展弦比不大(约 3.7)的机翼表面。从对带BLI效应的二维翼型的研究中[15]可知,在

表 4 t_2 对气动系数的影响

Table 4 Effect of t_2 on aerodynamic coefficients

t_2	C_L	C_D	C_M
0.007	0.4113	0.01972	0.0376
0.047	0.4137	0.02071	0.0449
0.087	0.4164	0.02202	0.0527

北航学报 赠阅

设定的参数变化区间内，随整流罩厚度的增大，整流罩发生了失速，C_L 先增大后减小，压差阻力系数一直增大，C_M 变化趋势和 C_L 一致。而从表 4 中气动系数随 t_2 的变化规律可以看出，在设定的相同变化区间内整流罩未发生失速，这与二维研究结果是不同的。

图 6 为 t_2 变化时的压力云图和流线图，图 7 为 t_2 变化对各截面压力分布的影响。从图中可以看出，t_2 变化主要影响整流罩表面的压力分布，基本不影响进气边界前的机体压力分布。在 t_2 所在的截面 2，t_2 增大使得当地的压力峰值增大，背风面的压力也随之增大。t_2 增大也使得其他截面的压力峰值增加了，且增加的幅值随与截面 2 的距离增加而降低。这是由于 t_2 的变化不仅改变了当地截面的形状，还影响了展向流动，从而影响到整个截面的压力分布。另外，从整流罩表面流线还可以看出，靠近对称面的整流罩表面流线较为平顺，与平直翼的流线相似；靠近整流罩端面的流线向后缘收缩，体现出三维效应；t_2 增大还使得流线略向内收缩，表明此时三维效应略微增强。

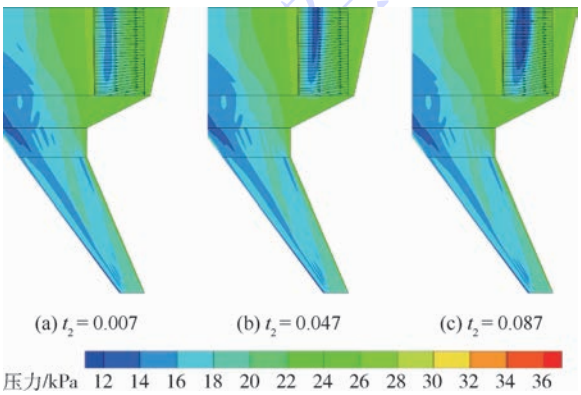


图 6　t_2 变化时的压力云图和流线图

Fig. 6　Pressure contours and stream lines with variation of t_2

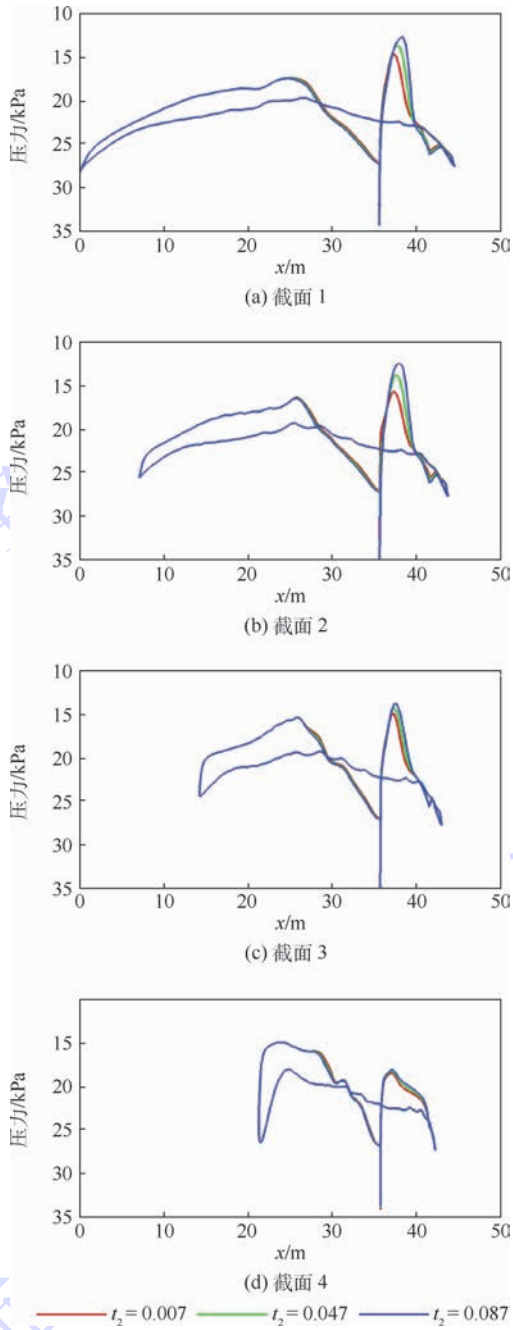


图 7　t_2 变化时的各截面压力分布

Fig. 7　Pressure distribution on cross sections with variation of t_2

2.2.2　截面最大厚度位置 x_4 的影响

x_4 是 4 个截面的最大厚度位置变量中对气动系数敏感度最小的一个。气动系数随 x_4 的变化规律见表 5，可见随 x_4 增大，3 个气动系数的变化量很小。与 2.2.1 节的结果对比可知，整流罩厚度位置的影响比厚度要弱，这也与 Morris 敏感度分析法的结果一致。

图 8 为 x_4 变化时的压力云图和流线图，图 9 为 x_4 变化对各截面压力分布的影响。从图中可以看出，x_4 的后移使得当地截面的压力峰值略微降低，这与二维结果是一致的。从其影响范围看，除对当地截面整流罩压力峰值有一定影响外，对机体表面和其他截面压力分布的影响很小。

表 5　x_4 对气动系数的影响

Table 5　Effect of x_4 on aerodynamic coefficients

x_4	C_L	C_D	C_M
0.25	0.413 5	0.020 70	0.044 7
0.29	0.413 7	0.020 71	0.044 9
0.33	0.413 7	0.020 68	0.044 8

2.3　其他设计条件对气动特性的影响

分布式动力推进 BWB 布局其他基本设计参数在基准值附近变化时[14]，整流罩对气动特性的影响规律可能不同。由于推进系统的布局参数众多，为了找到其中的关键参数，首先对基本设计参数进行了初步的敏感度区间分析，结果见表 6。

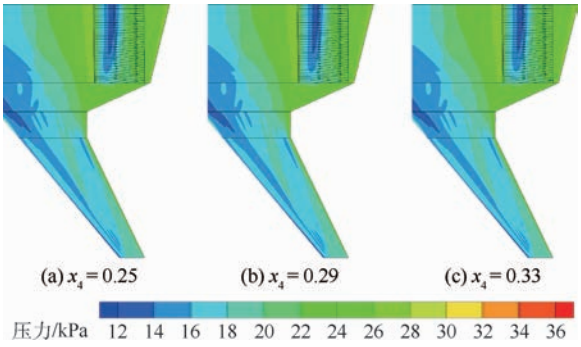


图 8 x_4 变化时的压力云图和流线图

Fig. 8 Pressure contours and stream lines with variation of x_4

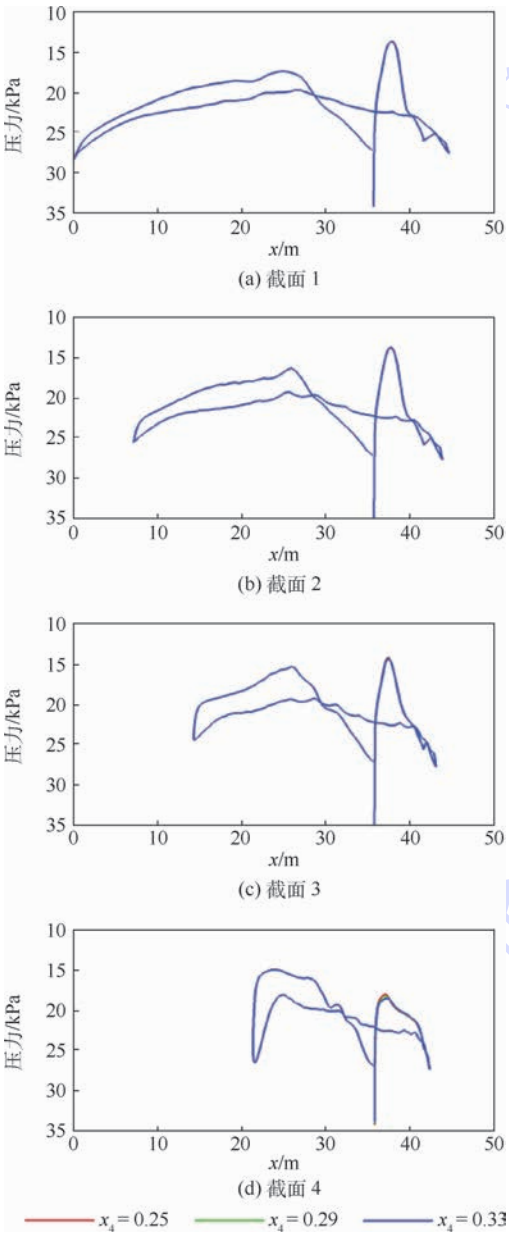


图 9 x_4 变化时的各截面压力分布

Fig. 9 Pressure distribution on cross sections with variation of x_4

表 6 基本设计参数的区间敏感度排序

Table 6 Interval sensitivity order of basis design parameters

参 数	C_L	C_D
流量系数	1.8	0.525
排气方向	1.34	0.32
进气边界弦向位置	0.392	0.28
进气边界高度	0.0686	0.04
弦向整流罩长度	0.0257	0.0386
展向流量分布	0.0195	0.024
展向进气位置分布	0.0068	0.027

从表 6 中可以看出,流量系数、排气方向和进气边界弦向位置这 3 个参数区间敏感度较大。以下分析这些参数在基准值附近变化时,整流罩关键参数 t_2 对气动特性的影响规律。

2.3.1 排气方向

推进系统的排气对飞机尾部附近的流场有引射作用,因此排气方向 τ 会影响飞机的气动特性。图 10 为不同 τ 下,C_L、C_D 和 C_M 随 t_2 的变化。

从图 10 中可以看出,在 τ 取不同值时,C_L、C_D 和 C_M 随 t_2 单调增加的趋势是不变的。且 C_L- t_2 和 C_M- t_2 曲线的斜率也基本不变,也就是说 τ 的变化不改变气动系数随 t_2 的变化规律。当 $\tau=20°$时,C_D 随 t_2 增加的斜率明显下降,此时 C_D 几乎不随 t_2 变化而变化。另外,当 $\tau=20°$时,C_L 和 C_M 随 t_2 变化的曲线都大幅向下移动,表明斜向上方的排气会使得气动特性大幅恶化。

图 11 为不同 τ 对应的局部压力分布和流线图。从图 11 (a)中可见,由于布局的原因,$\tau=20°$时排气方向与排气口后的机体表面产生了一个较大的夹角,在夹角处产生了分离,造成了气动特性的恶化。从图 11(c)中可见,$\tau=-20°$时由于排气口后机体表面约束了气流的偏转,排气方向与 $\tau=0°$时相差不大,因此气动特性与 $\tau=0°$时相差不大。另一方面由于气流上沿仍产生了角度不大的偏转,因此,整流罩表面的流动也受其影响,低压区略有扩大,从而 C_L、C_D 均有所增大。

2.3.2 流量系数

流量系数是 BLI 耦合布局中的重要参数,其变化会显著影响耦合布局的气动特性。图 12 为不同 MFR 下气动系数随 t_2 的变化。

从图 12 中可以看出,与基准状况相比,在 MFR 增大时,C_L、C_D 和 C_M 随 t_2 单调增加的趋势不变且斜率基本一致。而 MFR 减小至 0.40 时,C_L、和 C_M 随 t_2 先基本不变后减小,C_D 随 t_2 先增加后减小。图 13 为 MFR = 0.40 时,不同 t_2 对应的压力分布和流线图。从图13中可见,当t_2减

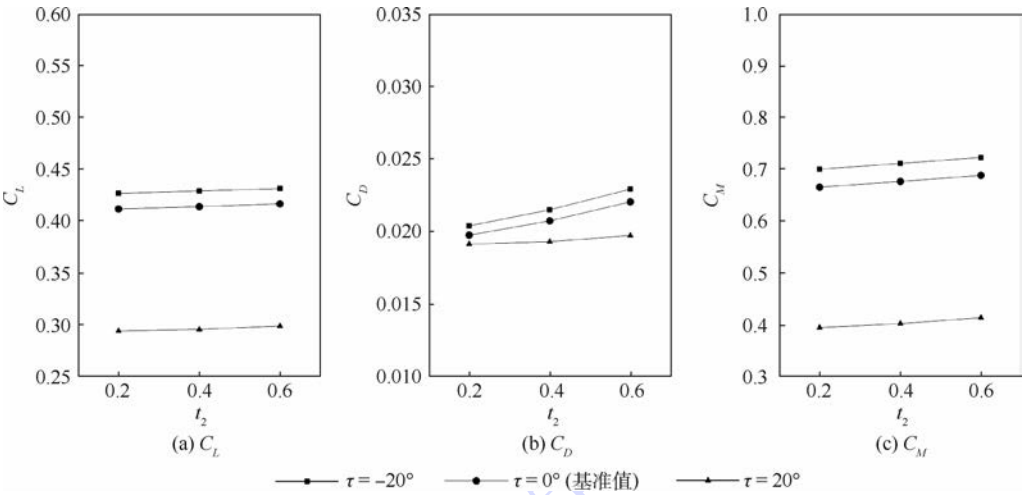


图 10　不同 τ 下气动系数随 t_2 的变化

Fig. 10　Variation of aerodynamic coefficients with t_2 at different τ

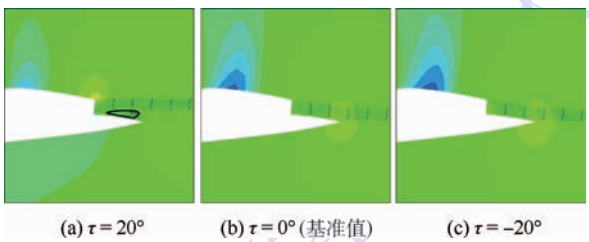


图 11　不同 τ 下的压力云图和流线图

Fig. 11　Pressure contours and stream lines at different τ

小时,整流罩表面低压区的前缘局部后移,低压区略有变小。t_2 增加时,低压区的后缘局部前移,低压力区明显变小。观察图 13(c)的流线可知,在 t_2 截面对应的位置,机身后部的流线明显向上偏折,这是因为 t_2 增加使得当地气流发生了后缘分离,机身后部的气流减速,从而喷流的流线向上偏折。可见整流罩局部厚度增加的构型在MFR过小时,会导致当地的气流发生后缘分离,进而造成升力降低、阻力增加以及喷口流线偏折。

2.3.3　进气边界弦向位置

进气边界弦向位置 L_f 是 BLI 耦合布局中进排气系统的重要位置参数。图 14 为不同 L_f 下气动系数随 t_2 的变化。由于几何约束,进排气系统无法再向后移动,因此研究 L_f 向前移动不同距离时 t_2 的影响规律。

从图 14 中可以看出,L_f 减小至 0.75 时,C_L 和 C_M 随 t_2 先增加后减小;L_f 减小至 0.70 时,C_L 和 C_M 随 t_2 的增加而减小;所有 L_f 下 C_D 均随 t_2 单调增加,且 L_f 越靠前 C_D 越大。这说明随着进排气系统的前移,整流罩的失速特性和气动特性也变差了。图 15 为 $L_f=0.70$ 时,不同 t_2 对应的压力分布和流线图。可见在 $t_2=0.007$ 时,整

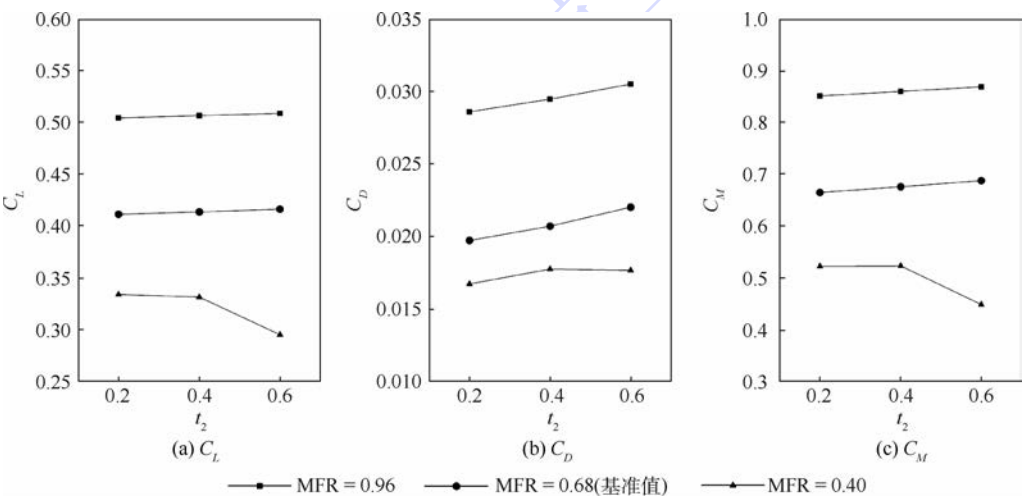
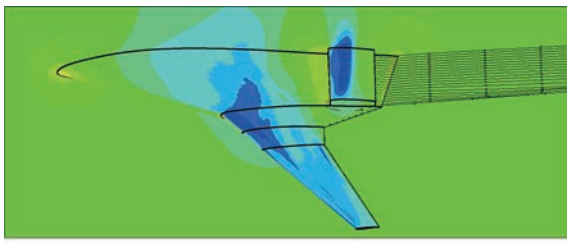


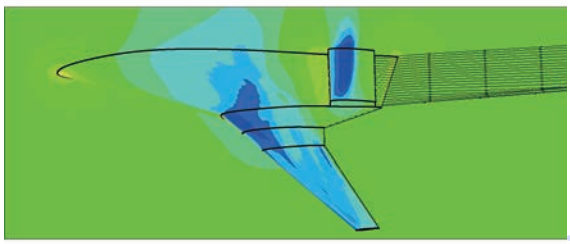
图 12　不同 MFR 下气动系数随 t_2 的变化

Fig. 12　Variation of aerodynamic coefficients with t_2 at different MFR

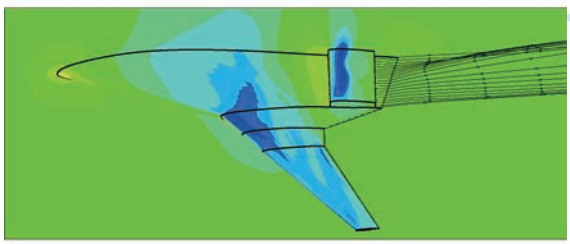
北航学报 赠阅



(a) t_2=0.007



(b) t_2=0.047



(c) t_2=0.087

图 13 MFR = 0.40 时飞机表面和对称面压力云图及喷口流线图

Fig. 13 Pressure contours of aircraft surface and symmetric plane and stream lines of nozzle at MFR = 0.40

流罩表面后部即已经存在分离区，t_2 的增大使得分离区也随之增大，从而愈加恶化了气动特性，t_2 = 0.087 时截面 2 后部的喷口气流也出现了偏折现象。这是由于进气边界的前移使得整流罩前缘过早地遭遇上翼面未经充分加速的气流，整个整流罩表面的气流能量不足，从而更容易失速。

2.4 参数耦合影响分析

Morris 敏感度分析结果表明，在整流罩设计参数中，t_2 和 t_3 对气动系数的耦合作用较强，下面分析原因。定义这 2 个参数同时产生增量（增量可以是正值或负值）时，气动系数的增量为 d_{coup}，这 2 个参数单独产生增量时气动系数的增量之和为 d_{line}。易知 d_{line} 即为 d_{coup} 的线性主部。再定义两者的相对差别为 $(d_{\text{line}} - d_{\text{coup}})/d_{\text{coup}}$，可知相对差别越大，代表这 2 个变量的耦合作用越强。考虑到参数参考变化区间的长度，将参数增量设为 ±4%。基于以上分析，将这 2 个参数产生增量时相对差别的变化规律列于表 7。

从表 7 中可见，t_2 和 t_3 同增时，两者对气动系数的耦合作用比两者同减时更强。这是因为整流罩厚度增加后发生了失速，其表面难以维持附着流动，气动系数的非线性增强。这与二维情形下 t 与 L_f 的耦合变化影响趋势是相同的[15]。

图 16 为 t_2 和 t_3 耦合变化时的压力分布图，每图的上半部分为 t_2 和 t_3 同时变化；下半部分为 t_2 单独变化，小图为 t_3 单独变化。从图中可以看出，t_2 和 t_3 同增时的影响更强，低压区显著减小。t_2 和 t_3 同减时类似，整流罩表面的低压区几乎消失，其效果比 t_2 和 t_3 单独变化时更强。

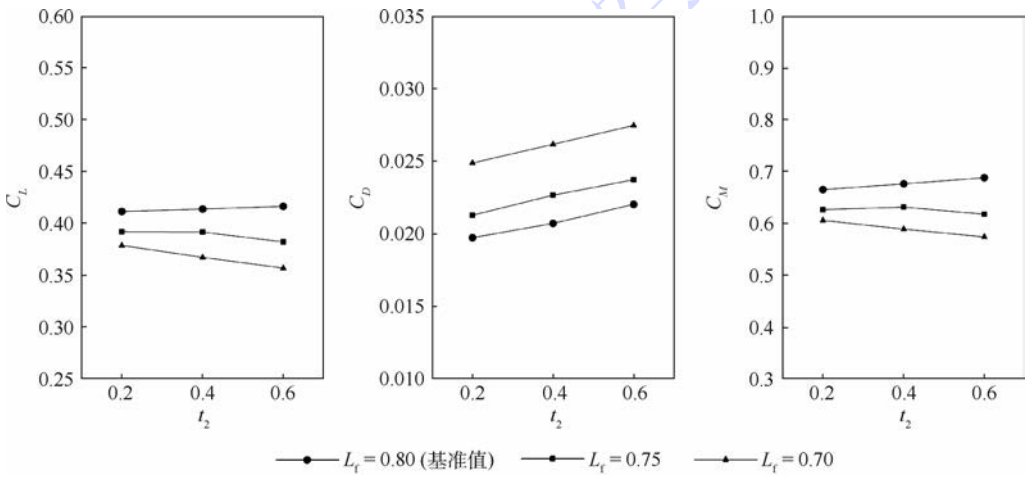


图 14 不同 L_f 下气动系数随 t_2 的变化

Fig. 14 Variation of aerodynamic coefficients with t_2 at different L_f

北航学报 赠阅

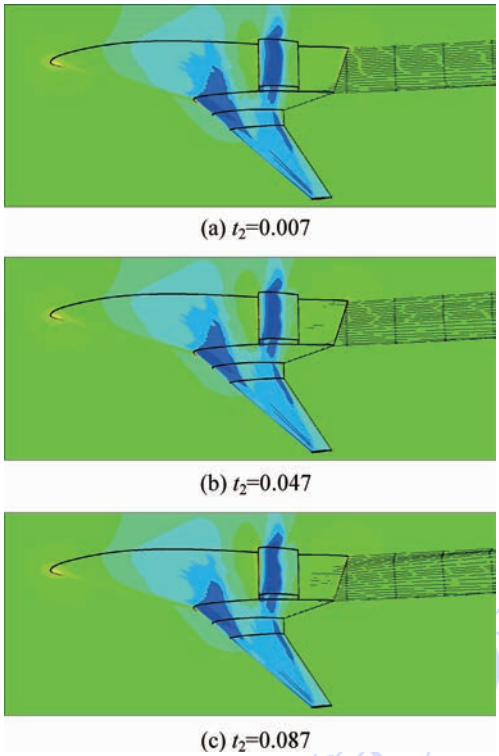


图 15　L_f = 0.70 时飞机表面和对称面压力分布及喷口流线图

Fig. 15　Pressure contours of aircraft surface and symmetric plane and stream lines of nozzle at L_f = 0.70

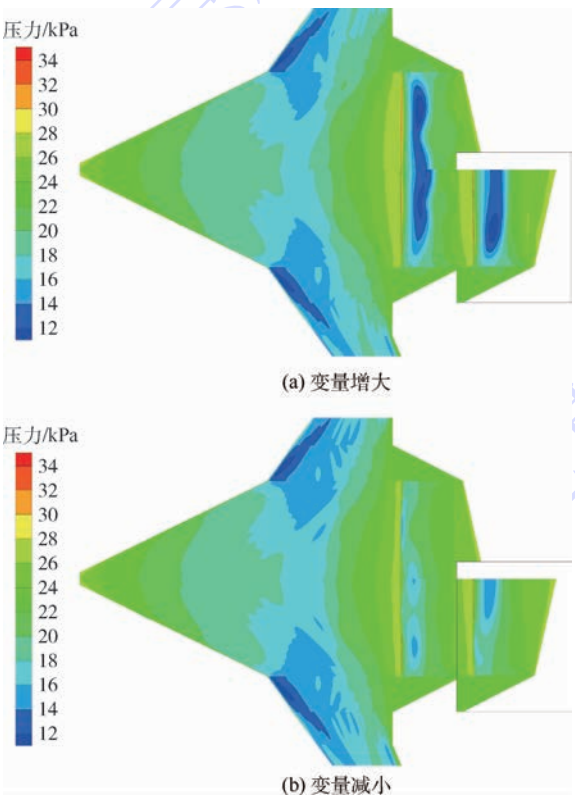


图 16　t_2 和 t_3 耦合变化时的压力分布

Fig. 16　Pressure distribution with coupled variation of t_2 and t_3

表 7　t_2 和 t_3 对气动系数的耦合影响

Table 7　Coupled effect of t_2 and t_3 on aerodynamic coefficients

气动系数	算例	d_{coup}	d_{line}	$\frac{d_{line}-d_{coup}}{d_{coup}}$/%
C_L	1	−0.012 4	−0.009 8	−21.1
	2	−0.028 7	−0.002 1	−92.8
C_D	1	−0.002 6	0.003 2	24.4
	2	−0.003 8	−0.004 7	21.3
C_M	1	0.045 2	−0.041 8	−7.5
	2	−0.040 5	−0.019 1	−147.1

注：算例 1、2 分别代表 t_2 和 t_3 同时减少和同时增加。

3　结　论

本文使用 Morris 敏感度分析法，结合数值模拟，研究了 BLI 效应下分布式动力 BWB 布局飞机的整流罩设计参数对气动特性的影响，获得以下结论：

1）对整流罩 4 个截面的厚度和最大厚度位置共 8 个参数进行了气动特性敏感度分析，结果表明：特征截面 2 和 3 的最大厚度对气动特性影响相对较大，其他参数影响相对较小。总体而言，厚度参数比最大厚度位置参数的敏感度更大且与其他参数耦合作用更强，这与二维的分析结果一致。

2）特征截面 2 最大厚度对气动特性影响较大的主要原因为：截面厚度变化使得当地截面翼型的厚度和弯度增加，从而改变本地截面的压力分布，增大升力和压差阻力系数；截面厚度的变化还改变了展向流动，使得其他截面的压力分布产生变化，并影响整流罩沿翼展的流线分布。

3）当流量系数和进气口弦向位置减小时，截面最大厚度增大对气动特性影响更为显著，这是因为在这两种情况下，整流罩表面来流速度较低，截面最大厚度增大会使得整流罩发生失速，影响整流罩表面压力分布和喷口后气流方向，从而影响升力和阻力。

4）特征截面 2 和 3 的最大厚度同时变化时对气动特性的耦合作用较强，这是因为两者同时增大使整流罩表面的失速区域进一步扩大，增加了气动特性的非线性。

参考文献（References）

[1] LIEBECK R H. Design of the blended wing body subsonic transport[J]. Journal of Aircraft, 2004, 41(1): 10-25.

[2] QIN N, VAVALLE A, MOIGNE L A, et al. Aerodynamic considerations of blended wing body aircraft[J]. Progress in Aerospace Sciences, 2004, 40(6): 321-343.

北航学报 赠阅

[3] LABAN M, ARENDSEN P, ROUWHORST W, et al. A computational design engine for multi-disciplinary optimisation with application to a blended wing body configuration[C]// 9th AIAA/ASSMO Multidisciplinary Analysis and Optimization Conference. Reston: AIAA, 2002.

[4] KIM H, LIOU M F, LIOU M S. Mail-slot nacelle shape design for N3-X hybrid wing body configuration[C]// 51st AIAA/SAE/ASEE Joint Propulsion Conference. Reston: AIAA, 2015: 3805.

[5] KO A. The multidisciplinary design optimization of a distributed propulsion blended-wing-body aircraft[D]. Blacksburg: Virgina Polytechnic Institute and State University, 2003.

[6] GOHARDANI A S, DOULGERIS G, SINGH R. Challenges of future aircraft propulsion: A review of distributed propulsion technology and its potential application for the all electric commercial aircraft[J]. Progress in Aerospace Sciences, 2011, 47(5): 369-391.

[7] KIM H D, BROWN G V, FELDER J L. Distributed turboelectric propulsion for hybrid wing body aircraft[C]// 9th International Powered Lift Conference. London: Royal Aeronautical Society, 2008.

[8] HILEMAN J I, SPAKOVSZKY Z S, DRELA M, et al. Airframe design for silent fuel-efficient aircraft[J]. Journal of Aircraft, 2010, 47(3): 956-969.

[9] KO A, LEIFSSON L T, SCHETZ J A, et al. MDO of a blended-wing-body transport aircraft with distributed propulsion: AIAA-2003-6732[R]. Reston: AIAA, 2003.

[10] KO A, SCHETZ J A, MASON W H. Assessment of the potential advantages of distributed-propulsion for aircraft[C]// XVIth International Symposium on Air Breathing Engines (ISABE). Reston: AIAA, 2003: 71-79.

[11] RODRIGUEZ D L. Multidisciplinary optimization method for designing boundary-layer-ingesting inlets[J]. Journal of Aircraft, 2009, 46(3): 883-894.

[12] LUNDBLADH A, GRÖNSTEDT T. Distributed propulsion and turbofan scale effects[C]// ISABE 2005, 17th Symposium on Airbreathing Engine. Reston: AIAA, 2005.

[13] LIOU M S, KIM H J, LIOU M F. Aerodynamic design of the hybrid wing body with nacelle: N3-X propulsion-airframe configuration[C]// 34th AIAA Applied Aerodynamics Conference. Reston: AIAA, 2016: 3875.

[14] 闫万方, 吴江浩, 张艳来. 分布式推进关键参数对 BWB 飞机气动特性影响[J]. 北京航空航天大学学报, 2015, 41(6): 1055-1065.
YAN W F, WU J H, ZHANG Y L. Effects of distributed propulsion crucial variables on aerodynamic performance of blended wing body aircraft[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(6): 1055-1065 (in Chinese).

[15] 项洋, 吴江浩, 张艳来. BLI 效应下整流罩设计对翼型气动特性的影响[J]. 北京航空航天大学学报, 2016, 42(5): 945-952.
XIANG Y, WU J H, ZHANG Y L. Effects of cowling design on aerodynamic performance of airfoil with BLI[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(5): 945-952 (in Chinese).

[16] NICHOLS M R, KEITH A L. Investigation of a systematic group of NACA 1-series cowlings with and without spinners: NACA-Report-950[R]. Washington, D. C.: U. S. Government Printing Office, 1950.

[17] MORRIS M D. Factorial sampling plans for preliminary computational experiments[J]. Technometrics, 1991, 33(2): 161-174.

作者简介：

项洋 男，博士研究生。主要研究方向：飞行器空气动力学。

吴江浩 男，博士，教授，博士生导师。主要研究方向：飞行器空气动力学。

熊峻江 男，博士，教授，博士生导师。主要研究方向：飞机结构力学。

北京航空航天大学学报

北航学报 赠阅

Aerodynamic design of nacelle of blended-wing-body aircraft with distributed propulsion

XIANG Yang, WU Jianghao*, XIONG Junjiang

(School of Transportation Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: Nacelle design has a significant effect on aerodynamic performance of blended-wing-body (BWB) aircraft with distributed propulsion. To clarify the effect and its reason of primary nacelle design parameters on aerodynamic performance of BWB aircraft with boundary layer ingestion (BLI) effect, a detailed study was conducted by computational fluid dynamics (CFD) method and Morris sensitivity analysis method. Sensitivity order and coupled effect of primary design parameters on aerodynamic performance were obtained. Flow details of higher sensitivity and greater coupled effect parameters were analyzed under baseline and alternative condition. The results show that the relatively most significant parameters are the maximum thickness of section 2 and 3. The main reason is that local thickness and camber increase, and pressure distribution of whole nacelle surface is changed. Leeward local stall will occur as the maximum thickness increases configuration when mass flow rate decreases and inlet location along the chord direction moves forward. The coupled effect of the maximum thickness of section 2 and 3 on aerodynamic performance is relatively significant.

Keywords: blended-wing-body (BWB) configuration; boundary layer ingestion (BLI); computational fluid dynamics (CFD); sensitivity analysis; nacelle

北京航空航天大学学报

Received: 2017-01-16; **Accepted**: 2017-02-06; **Published online**: 2017-03-22 11:40

URL: kns.cnki.net/kcms/detail/11.2625.V.20170322.1140.001.html

* **Corresponding author**. E-mail: buaawjh@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2016.0943

频率步进 SAR 虚拟阵列模型成像方法

冉达[1,2], 尹灿斌[2,*], 贾鑫[2]
(1. 航天工程大学 航天信息学院, 北京 101416; 2. 航天工程大学 航天指挥学院, 北京 101416)

摘 要: 针对频率步进合成孔径雷达(SAR)采用经典逆傅里叶变换成像方法时距离向无模糊测绘带宽度有限的问题,提出一种将频率步进 SAR 脉冲串信号等效为沿航迹分布的虚拟阵列雷达信号的模型及成像处理方法,并利用改进的后向投影方法实现对目标的无模糊成像。建立了频率步进 SAR 虚拟阵列模型,给出了基于该模型的高分辨距离像合成方法,并通过在原始后向投影方法的基础上引入距离偏移校正和二次相位补偿,实现了对目标的精确二维成像。结果表明:频率步进 SAR 虚拟阵列模型成像方法不受频率步进雷达无模糊测绘带宽度的理论限制,可以实现较宽测绘带内各目标的无模糊、快速成像。

关 键 词: 频率步进合成孔径雷达; 虚拟阵列; 后向投影; 快速傅里叶变换(FFT); 空时自适应处理(STAP)
中图分类号: V243.2; TN957.52
文献标识码: A **文章编号**: 1001-5965(2018)01-0082-07

频率步进信号通过多脉冲相参合成处理获得距离维高分辨成像能力,信号瞬时带宽小。不仅大大降低了对系统采样率的要求及工程实现难度,而且可以在一定程度上提高雷达的抗干扰能力,因此被广泛应用于雷达目标检测、成像等领域[1-6]。

利用频率步进信号成像主要存在 2 个问题:一是频率步进信号对运动目标比较敏感。频率步进信号要求序贯脉冲之间保持严格的相位关系,雷达与目标之间的相对径向运动会造成目标距离像的偏移和成像质量的退化,为得到运动目标的正确成像结果,通常在脉冲综合前需要首先完成速度测量与补偿。文献[7-12]对上述问题已经进行了较为详细的研究,相关问题已得到较好的解决。二是频率步进信号受信号参数设计的理论制约,存在距离向最大无模糊测绘带宽度的限制问题,严重制约了该信号在高分辨-宽测绘带 SAR 成像探测领域的应用。为解决该问题,文献[13-14]分别提出利用参差重频和脉冲波形分集技术实现距离向解模糊处理,但是,由于需要改变雷达的系统参数,增加了系统实现的复杂度。文献[15]利用仰角维自由度,提出了一种解距离模糊的三维空时自适应处理(Space Time Adaptive Processing, STAP)算法,但是算法复杂,难以实际应用。文献[16]基于频率分集阵列雷达的天线方向图,提出了一种改善距离模糊的 STAP 算法。文献[17]利用频率分集阵列雷达在距离向的空变自由度,实现了距离模糊杂波抑制和运动目标距离解模糊,算法简单有效,无需改变系统参数。实质上,频率步进信号可以看成是一种频率分集信号,在“走-停”模型假设下,可以将频率步进合成孔径雷达(SAR)的各组脉冲串等效为沿航迹分布的虚拟线性阵列信号。于是,基于频率分集阵列雷达解距离模糊的思想,通过将频率步进 SAR

收稿日期: 2016-12-15; **录用日期**: 2017-04-24; **网络出版时间**: 2017-08-31 14:18
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170831.1418.002.html
* **通信作者**. E-mail: ExpressESP@126.com

引用格式: 冉达, 尹灿斌, 贾鑫. 频率步进SAR 虚拟阵列模型成像方法[J]. 北京航空航天大学学报, 2018, 44(1): 82-88.
RAN D, YIN C B, JIA X. Imaging approach for frequency-stepped SAR with virtual array model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(1): 82-88 (in Chinese).

等效为虚拟频率分集阵列雷达并进行相应的成像处理，可解决频率步进 SAR 距离向无模糊测绘带宽度有限的问题。

本文首先给出了频率步进 SAR 虚拟阵列模型及其高分辨距离像合成方法；然后，为实现目标的精确二维成像，给出了适用于虚拟阵列模型的改进后向投影方法，解决了经典方法距离向无模糊测绘带宽度受限的问题；最后，通过实验和结果分析验证了本文方法的有效性。

1　虚拟阵列模型的频率步进 SAR 高分辨距离像合成方法

图 1 给出了基于虚拟阵列模型的频率步进 SAR 成像空间几何示意图，其中频率步进脉冲串被等效为虚拟的天线阵列信号。以场景中心点 O 为原点建立直角坐标系 O-xyz，雷达沿航线 y' 以速度 $\boldsymbol{v}$ 匀速运动（y' 轴与 y 轴平行），发射信号的子脉冲时间间隔为 T_r，宽度为 T_p，则相邻子脉冲之间的距离为 $d=|\boldsymbol{v}|T_r$。雷达在合成孔径时间内共发射 N_a 组 N_p 个子脉冲的频率步进脉冲串。以脉冲串的第 1 个子脉冲为参考，记其对应的雷达位置为 O'，发射信号载频为 f_0，则第 n 个子脉冲发射信号载频为 f_n，有 $f_n=f_0+n\Delta f, n=0,1,\cdots,N_p-1$，其中 Δf 为频率步进间隔。O' 到场景中心点 O 的距离矢量为 $\boldsymbol{r}_0$，其擦地角与方位角分别记为 $\varphi_{\eta 0}$ 和 $\theta_{\eta 0}$，大小与场景中心点 O 和方位慢时间 η 有关；O' 到目标 p 的距离矢量为 $\boldsymbol{r}_{p0}$，第 n 个子脉冲到目标 p 的距离矢量为 $\boldsymbol{r}_{np}$，其中 $\boldsymbol{r}_{p0}$ 的擦地角与方位角分别记为 $\varphi_{\eta p}$ 和 $\theta_{\eta p}$，其大小与目标 p 和方位慢时间 η 有关。

假设雷达发射信号为

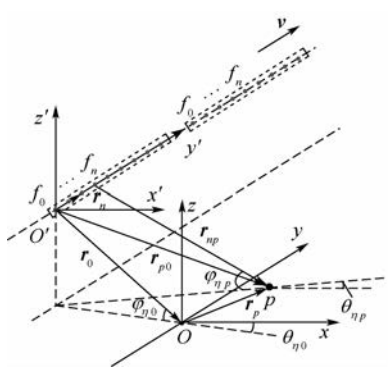


图 1　虚拟阵列模型的频率步进 SAR 成像空间几何示意图

Fig. 1　Spatial geometry of frequency-stepped SAR imaging of virtual array model

$$s_{\bar{k}n,\mathrm{t}}(t)=\mathrm{rect}\left(\frac{t-(n+\bar{k}N_\mathrm{p})T_\mathrm{r}}{T_\mathrm{p}}\right)\cdot \exp(\mathrm{j}(2\pi f_n t+\theta_{\bar{k}n})) \tag{1}$$

式中：t 为全时间，$-T_\mathrm{p}/2+(n+\bar{k}N_\mathrm{p})T_\mathrm{r}\leqslant t\leqslant(n+\bar{k}N_\mathrm{p})T_\mathrm{r}+T_\mathrm{p}/2$；$\bar{k}=0,1,\cdots,N_\mathrm{a}-1$ 为脉冲串序号；$n=0,1,\cdots,N_\mathrm{p}-1$ 为脉冲串的子脉冲序号；$\theta_{\bar{k}n}$ 为每个发射脉冲的初始相位。

假设理想点目标 p 的散射系数为 σ_p，则频率步进 SAR 的接收信号可表示为

$$s_{\bar{k}n,\mathrm{r}}(t)=\sigma_p\mathrm{rect}\left(\frac{t-(n+\bar{k}N_\mathrm{p})T_\mathrm{r}-\tau_p}{T_\mathrm{p}}\right)\cdot \exp(\mathrm{j}[2\pi f_n(t-\tau_p)+\theta_{\bar{k}n}]) \tag{2}$$

式中：τ_p 为目标 p 与雷达之间的双程时延；$-T_\mathrm{p}/2+(n+\bar{k}N_\mathrm{p})T_\mathrm{r}\leqslant t-\tau_p\leqslant(n+\bar{k}N_\mathrm{p})T_\mathrm{r}+T_\mathrm{p}/2$，方位慢时间 $\eta=\bar{k}N_\mathrm{p}T_\mathrm{r}$，$\bar{k}=0,1,\cdots,N_\mathrm{a}-1$。

在“走-停”模型下，频率步进 SAR 的一组脉冲串可等效为一个沿航迹分布的虚拟天线阵列信号。为方便表述，本节在推导频率步进 SAR 虚拟阵列回波模型时，仅以一组脉冲串为例，其余脉冲串的信号表达式可类推获得。假设频率步进 SAR 脉冲串信号的第 n 个子脉冲的发射信号为

$$s_{n,\mathrm{t}}(t')=\mathrm{rect}\left(\frac{t'-nT_\mathrm{r}}{T_\mathrm{p}}\right)\exp(\mathrm{j}(2\pi f_n t+\theta_n)) \tag{3}$$

式中：t' 为快时间，$-T_\mathrm{p}/2+nT_\mathrm{r}\leqslant t'\leqslant nT_\mathrm{r}+T_\mathrm{p}/2$；$\theta_n$ 为信号的初始相位；全时间 $t=t'+\eta$。由已知条件可知，第 n 个子脉冲到目标 p 的距离矢量

$$\boldsymbol{r}_{np}=\boldsymbol{r}_{p0}-\boldsymbol{r}_n \tag{4}$$

式中：$\boldsymbol{r}_n$ 为第 1 个子脉冲到第 n 个子脉冲位置的距离矢量。故第 n 个子脉冲对应的目标回波为

$$s_{n,\mathrm{r}}(t',\eta)=\sigma_p\mathrm{rect}\left(\frac{t'-nT_\mathrm{r}-2|\boldsymbol{r}_{\mathrm{np}}|/c}{T_\mathrm{p}}\right)\cdot \exp(\mathrm{j}(2\pi f_n t-2\boldsymbol{k}_{np}\cdot\boldsymbol{r}_{np})) \tag{5}$$

式中：$\boldsymbol{k}_{np}=k_n\cdot\hat{\boldsymbol{r}}_{np}$ 为距离矢量 $\boldsymbol{r}_{np}$ 对应的波数，$k_n=2\pi f_n/c$，$\hat{\boldsymbol{r}}_{np}$ 为 $\boldsymbol{r}_{np}$ 的单位向量，c 为光速；$-T_\mathrm{p}/2+nT_\mathrm{r}+2|\boldsymbol{r}_{np}|/c\leqslant t'\leqslant nT_\mathrm{r}+2|\boldsymbol{r}_{np}|/c+T_\mathrm{p}/2$。远场条件下，$\hat{\boldsymbol{r}}_{np}\approx\hat{\boldsymbol{r}}_{p0}$，$\hat{\boldsymbol{r}}_{p0}$ 为 $\boldsymbol{r}_{p0}$ 的单位向量。因此，波数 $\boldsymbol{k}_{np}=k_n\cdot\hat{\boldsymbol{r}}_{np}\approx k_n\cdot\hat{\boldsymbol{r}}_{p0}$，$k_n=2\pi f_n/c$。忽略信号包络项，将 $\boldsymbol{k}_{np}=k_n\cdot\hat{\boldsymbol{r}}_{p0}$ 和 $\boldsymbol{r}_{np}=\boldsymbol{r}_{p0}-\boldsymbol{r}_n$ 代入式(5)，则第 n 个子脉冲的回波信号可表示为

$$s_{n,\mathrm{r}}(t',\eta)=\sigma_p\exp\left(\mathrm{j}2\pi f_0\left(t-\frac{2\hat{\boldsymbol{r}}_{p0}\cdot\boldsymbol{r}_{p0}}{c}\right)\right)\cdot \exp\left(\mathrm{j}2\pi\left[n\Delta f\left(t-\frac{2\hat{\boldsymbol{r}}_{p0}\cdot\boldsymbol{r}_{p0}}{c}\right)+f_n\frac{2\hat{\boldsymbol{r}}_{p0}\cdot\boldsymbol{r}_n}{c}\right]\right) \tag{6}$$

式(6)中最后一个相位随着子脉冲发生改

北航学报 赠阅

变,反映了子脉冲位置与目标之间的相对几何关系以及子脉冲发射信号载频对回波相位的影响。由图 1 可知,$\hat{\boldsymbol{r}}_{np}=\hat{\boldsymbol{r}}_{p0}$ 和 $\boldsymbol{r}_n$ 可分别表示为

$$\begin{cases}\hat{\boldsymbol{r}}_{p0}=\cos\varphi_{\eta p}\cos\theta_{\eta p}\hat{\boldsymbol{x}}+\cos\varphi_{\eta p}\sin\theta_{\eta p}\hat{\boldsymbol{y}}+\\ \qquad\sin\varphi_{\eta p}\hat{\boldsymbol{z}}\\ \boldsymbol{r}_n=[0,n|\boldsymbol{v}|T_{\mathrm{r}}\cdot\hat{\boldsymbol{y}},0]\quad n=0,1,\cdots,N_{\mathrm{p}}-1\end{cases}\tag{7}$$

式中:$\hat{\boldsymbol{x}}$、$\hat{\boldsymbol{y}}$和$\hat{\boldsymbol{z}}$分别对应 x 轴、y 轴和 z 轴的单位向量。将式(7)代入式(6),同时对 N_{p} 个子脉冲回波进行叠加,有

$$\begin{aligned}s_{\mathrm{r}}(t',\eta)=&\sigma_p\exp\left(\mathrm{j}2\pi f_0\left(t-\frac{2|\boldsymbol{r}_{p0}|}{c}\right)\right)\cdot\\&\sum_{n=0}^{N_{\mathrm{p}}-1}\exp\left(\mathrm{j}n^2\left(\frac{4\pi\Delta f}{c}|\boldsymbol{v}|T_{\mathrm{r}}\cos\varphi_{\eta p}\sin\theta_{\eta p}\right)\right)\cdot\\&\exp\left(\mathrm{j}n\left[\frac{4\pi f_0}{c}|\boldsymbol{v}|T_{\mathrm{r}}\cos\varphi_{\eta p}\sin\theta_{\eta p}+\right.\right.\\&\left.\left.2\pi\Delta f\left(t-\frac{2|\boldsymbol{r}_{p0}|}{c}\right)\right]\right)\end{aligned}\tag{8}$$

当式(8)中二次相位项小于 $\pi/4$ 时,可忽略其对成像质量的影响。令最后一个相位项为 $\exp(\mathrm{j}n\Phi_1)$,其中 Φ_1 为

$$\begin{aligned}\Phi_1=&\frac{4\pi f_0}{c}|\boldsymbol{v}|T_{\mathrm{r}}\cos\varphi_{\eta p}\sin\theta_{\eta p}+\\&2\pi\Delta f\left(t-\frac{2|\boldsymbol{r}_{p0}|}{c}\right)\end{aligned}\tag{9}$$

将 $\sum_{n=0}^{N_{\mathrm{p}}-1}\exp(\mathrm{j}n\Phi_1)$ 的结果代入式(8),并利用参考信号 $s_{\mathrm{ref}}(t)=\exp(\mathrm{j}2\pi f_0 t)$ 对回波信号进行解调,有

$$\begin{aligned}s_{\mathrm{rB}}(t',\eta)=&\sigma_p\exp\left(-\mathrm{j}2\pi f_0\frac{2|\boldsymbol{r}_{p0}|}{c}\right)\cdot\\&\exp\left(\mathrm{j}\frac{N_{\mathrm{p}}-1}{2}\Phi_1\right)\frac{\sin\left(\frac{N_{\mathrm{p}}}{2}\Phi_1\right)}{\sin\left(\frac{1}{2}\Phi_1\right)}\end{aligned}\tag{10}$$

当 $\Phi_1=2\kappa\pi$($\kappa=0,\pm1,\pm2,\cdots$)时,$|s_{\mathrm{rB}}(t',\eta)|$取得极大值,此时有

$$\begin{aligned}t'_{p0}=&\tau_{p,\mathrm{fixed}}+\tau_{p,\mathrm{offset}}=\\&\frac{2|\boldsymbol{r}_{p0}|}{c}-\frac{2f_0}{c\Delta f}|\boldsymbol{v}|T_{\mathrm{r}}\cos\varphi_{\eta p}\sin\theta_{\eta p}\end{aligned}\tag{11}$$

式中:$\tau_{p,\mathrm{fixed}}=2|\boldsymbol{r}_{p0}|/c$ 为参考脉冲位置与目标之间的双程距离回波延时;$\tau_{p,\mathrm{offset}}=-2f_0|\boldsymbol{v}|T_{\mathrm{r}}\cdot\cos\varphi_{\eta p}\sin\theta_{\eta p}/(c\Delta f)$ 为目标回波的额外延时,它与雷达平台速度 $|\boldsymbol{v}|$、子脉冲发射时间间隔 T_{r}、参考子脉冲位置以及目标 p 之间的空间几何关系、参考子脉冲发射信号载频 f_0 和频率步进间隔 Δf 有关。

由式(10)可知,$s_{\mathrm{rB}}(t',\eta)$具有周期性,其峰值两边第 1 零点之间的时间间隔反映了距离向分辨率。由于第 1 零点 $t_{p\pm1}$ 满足条件 $N_{\mathrm{p}}\Phi_1/2=\pm\pi$,于是有

$$t_{p\pm1}=\pm\frac{1}{N_{\mathrm{p}}\Delta f}+\frac{2|\boldsymbol{r}_{p0}|}{c}-\frac{2f_0}{c\Delta f}|\boldsymbol{v}|T_{\mathrm{r}}\cos\varphi_{\eta p}\sin\theta_{\eta p}\tag{12}$$

因此,距离分辨率 ρ_{r} 可表示为

$$\rho_{\mathrm{r}}=\left(\frac{1}{2}\cdot\frac{2}{N_{\mathrm{p}}\Delta f}\right)\frac{c}{2}=\frac{c}{2N_{\mathrm{p}}\Delta f}=\frac{c}{2B}\tag{13}$$

式中:系数 1/2 为半功率峰值处的时间间隔;B 为频率步进信号带宽。上述结果与频率步进信号距离向分辨率的理论值一致。

可见,通过将频率步进 SAR 的各组脉冲串等效为沿航迹排列的虚拟阵列信号,并对同一组脉冲串内的所有子脉冲回波进行累加,可实现频率步进 SAR 信号的距离像合成。相比经典成像方法,基于虚拟阵列模型的频率步进 SAR 高分辨距离像合成方法在得到目标正确的距离向成像结果基础上,进一步简化了信号处理的流程。

2 方位能量累积

2.1 后向投影原理

因目标对应的额外延时 $\tau_{p,\mathrm{offset}}$ 具有空变性,不便于统一进行距离校正,因此,后向投影方法成为方位成像的首选。通过逐方位向对目标的回波能量进行累积,后向投影方法可获得任意成像几何条件下的目标二维高分辨成像结果。设最终成像结果 $\boldsymbol{I}$ 的网格点数为 $K\times L$(距离向 × 方位向),各网格点相对于场景中心点 O 的位置矢量为 $\boldsymbol{r}_{kl}$,其中 $k=1,2,\cdots,K$,$l=1,2,\cdots,L$。根据后向投影方法原理,网格点 kl 处对应的二维成像结果可表示为

$$\boldsymbol{I}(k,l)=\int_{-T_{\mathrm{syn}}/2}^{T_{\mathrm{syn}}/2}s_{\mathrm{rB}}(t'_{kl0},\eta)\exp(\mathrm{j}\Phi_{\mathrm{c}})\mathrm{d}\eta\tag{14}$$

式中:$t'_{kl0}=\tau_{kl,\mathrm{fixed}}+\tau_{kl,\mathrm{offset}}$,$\tau_{kl,\mathrm{offset}}$为网络点 kl 对应的时间空变偏移量,$\tau_{kl,\mathrm{fixed}}$为网格点 kl 与参考子脉冲位置 O'之间对应的回波双程时延,它可表示为

$$\tau_{kl,\mathrm{fixed}}=\frac{2\hat{\boldsymbol{r}}_{kl0}\cdot\boldsymbol{r}_{kl0}}{c}=\frac{2|\boldsymbol{r}_{kl0}|}{c}\tag{15}$$

其中:$\hat{\boldsymbol{r}}_{kl0}$为参考子脉冲位置 O'到网格点的单位距离矢量。与 $\boldsymbol{r}_{kl0}$对应的相位补偿因子 $\exp(\mathrm{j}\Phi_{\mathrm{c}})$为

$$\exp(\mathrm{j}\Phi_{\mathrm{c}})=\exp(\mathrm{j}2\pi f_0\cdot2|\boldsymbol{r}_{kl0}|/c)\tag{16}$$

由式(15)和式(16)可知,网格点 kl 对应的

回波真实传播时延 $\tau_{kl,\text{fixed}}$ 和相位补偿因子 $\exp(\mathrm{j}\Phi_{\mathrm{c}})$，与原始后向投影方法目标的脉冲压缩峰值点时刻和相位补偿项相等。

2.2　距离偏移校正

由式(14)中 $t'_{kl0}=\tau_{kl,\text{fixed}}+\tau_{kl,\text{offset}}$ 可知，相比原始后向投影方法，基于虚拟阵列模型的频率步进 SAR 距离像中还引入了一个额外时间偏移量 $\tau_{kl,\text{offset}}$，其表达式为

$$\tau_{kl,\text{offset}}=-\frac{2f_0}{c\Delta f}|\boldsymbol{v}|T_{\mathrm{r}}\cos\varphi_{\eta kl}\sin\theta_{\eta kl} \tag{17}$$

式中：$\varphi_{\eta kl}$ 和 $\theta_{\eta kl}$ 分别为 $\boldsymbol{r}_{kl0}$ 的擦地角和方位角，其随方位慢时间 η 变化；$\tau_{kl,\text{offset}}$ 导致空变距离偏移量 $\Delta R=-(f_0|\boldsymbol{v}|T_{\mathrm{r}}\cos\varphi_{\eta kl}\sin\theta_{\eta kl})/\Delta f$。为获得正确的成像结果，必须在后向投影成像时对其进行补偿。由式(17)可知，$\tau_{kl,\text{offset}}$ 与每个图像网格点 kl 有关，可在反投影过程中同时实现。

2.3　二次相位补偿

当式(8)中二次相位项的取值大于 $\pi/4$ 时，必须对该项进行补偿，否则图像将散焦。当成像场景较小时，由于远场条件下 $\hat{\boldsymbol{r}}_{p0}\approx\hat{\boldsymbol{r}}_0$（$\hat{\boldsymbol{r}}_{p0}$ 和 $\hat{\boldsymbol{r}}_0$ 分别为距离矢量 $\boldsymbol{r}_{p0}$ 和 $\boldsymbol{r}_0$ 的单位向量）。此时，可利用场景中心点 O 处对应的二次相位对整个场景进行统一补偿，有

$$\exp(\mathrm{j}\Phi(n))=\exp\left(-\mathrm{j}n^2\left(\frac{4\pi\Delta f}{c}|\boldsymbol{v}|T_{\mathrm{r}}\cos\varphi_{\eta 0}\sin\theta_{\eta 0}\right)\right) \tag{18}$$

式中：$n=0,1,\cdots,N_{\mathrm{p}}-1$；因子 $\varphi_{\eta 0}$ 和 $\theta_{\eta 0}$ 与 η、参考子脉冲位置 O' 与场景中心点 O 的相对空间几何关系有关。

3　仿真实验

为验证本文方法正确性，现进行 3 组成像实验，实验仿真参数见表 1。

3.1　高分辨距离像合成

假定场景中存在 3 个位于无模糊距离范围内不同距离单元的点目标，分别利用经典方法和本文方法进行距离像合成，图 2 给出了2 种方法对应的距离像对比。

由图 2 可见，2 种方法均能得到目标的正确距离像，但是经典方法成像结果仅表示无模糊距离 $cT_{\mathrm{p}}/2$ 内目标的相对距离，而本文方法距离向成像结果表示的是整个回波信号时间长度内目标的绝对距离。由于设定目标均位于无模糊距离范围内，2 种方法成像结果一致。需要说明的是，实验为得到所有方位时刻完整的目标回波，总的回波时间长度为实际场景延迟前后各增加 $T_{\mathrm{p}}/2$ 的时间长度，故图中有距离为负的部分。

表 1　仿真参数

Table 1　Simulation parameters

参　数	数　值
参考载频/GHz	10
子脉冲数目	1 025
天线波束宽度/(°)	1.30
平台高度/km	8.00
总信号带宽/MHz	700
脉冲串重复频率/Hz	333.14
场景大小(距离向×方位向)/(m×m)	200×200
频率步进量/MHz	0.683
子脉冲宽度/μs	1.46
平台速度/(m·s^{-1})	58.50
参考斜距/km	11.31
占空比	0.50
子脉冲重复频率/kHz	341.46
地距分辨率(距离向×方位向)/(m×m)	0.30×0.36

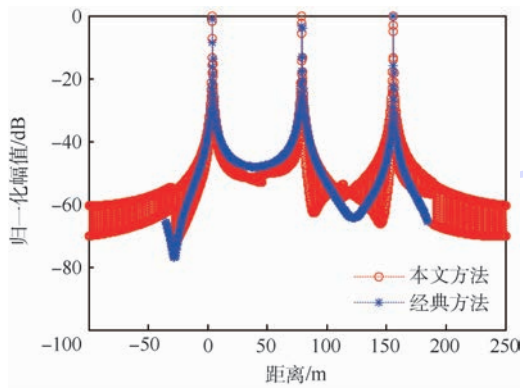


图 2　经典方法与本文方法目标距离像对比

Fig. 2　Comparison of target range profile between classic method and proposed method

3.2　频率步进 SAR 虚拟阵列模型方法成像

为验证本文方法成像性能，在 3.1 节实验 3 个点目标的基础上，方位向上再各自增加 3 个点目标，组成一个 3×3 的点目标阵，其位置分别为(-80，-80，0) m、(-80，0，0) m、(-80，80，0) m、(0，-80，0) m、(0，0，0) m、(0，80，0) m、(80，-80，0) m、(80，0，0) m 和(80，80，0) m。

图 3 给出了本文方法的目标成像结果，9 个点目标聚焦良好。为进一步分析成像性能，表 2 给出了图 3 中所有目标对应的点散布函数统计结果，表 3 给出了所有目标经典方法对应的点散布函数统计结果。其中，PSLR 为峰值旁瓣比，ISLR 为积分旁瓣比。对比表 2 和表 3 的结果可知，本文方法目标点散布函数性能的取值与经典方法非常接近，验证了本文方法的有效性。

北航学报 赠阅

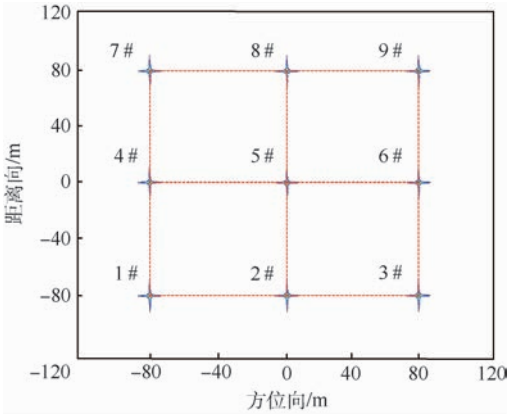


图 3 本文方法成像结果

Fig. 3 Imaging result of proposed method

表 2 本文方法目标点散布函数结果

Table 2 Target point spread function results of proposed method

目标序号	距离向		方位向	
	PSLR/dB	ISLR/dB	PSLR/dB	ISLR/dB
1	-13.3820	-9.8576	-13.2776	-10.2258
2	-13.3835	-9.8487	-13.3035	-10.2092
3	-13.3820	-9.8576	-13.2776	-10.2258
4	-13.4625	-9.8955	-13.2834	-10.2257
5	-13.4701	-9.8851	-13.3262	-10.2283
6	-13.4625	-9.8955	-13.2834	-10.2257
7	-13.6196	-9.9601	-13.2921	-10.2213
8	-13.6249	-9.9522	-13.3064	-10.2165
9	-13.6196	-9.9601	-13.2921	-10.2213

表 3 经典方法目标点散布函数结果

Table 3 Target point spread function results of classic method

目标序号	距离向		方位向	
	PSLR/dB	ISLR/dB	PSLR/dB	ISLR/dB
1	-13.3103	-9.9574	-13.2031	-10.2182
2	-13.3129	-9.8400	-13.2035	-10.2189
3	-13.3140	-9.7712	-13.2743	-10.2215
4	-13.3003	-9.8955	-13.2130	-10.2172
5	-13.3876	-9.9664	-13.2212	-10.2208
6	-13.3200	-10.0090	-13.2303	-10.2175
7	-13.3103	-9.9619	-13.2134	-10.2213
8	-13.3903	-9.8569	-13.2021	-10.2206
9	-13.3196	-9.9854	-13.1984	-10.2191

3.3 抗距离模糊及宽测绘带成像效果

当场景距离向宽度超过频率步进信号的最大无模糊距离范围时，经典成像方法的频率步进 SAR 成像结果将出现距离向模糊。为对比经典方法与本文方法抗距离模糊的成像效果，对场景距离范围为 3 倍频率步进信号无模糊距离的 17 个点目标进行成像实验。实验中，在 3.2 节仿真实验 9 个点目标的基础上，增加了8 个位于无模糊距离范围之外的点目标，新增加的 8 个点目标位置分别为(-296，-296，0)m、(-296，0，0)m、(-296，296，0)m、(0，-296，0)m、(0，296，0)m、(296，-296，0)m、(296，0，0)m 和(296，296，0)m。图 4(a)、(b)给出了经典方法与本文方法的成像结果。

对比图 4(a)和(b)可知，当场景大小超过频率步进信号参数决定的无模糊成像范围时，采用频率步进 SAR 经典成像方法的成像结果中存在严重的距离向模糊问题，有些模糊能量甚至遮盖了正确的目标成像结果；而采用本文方法的成像结果中不存在距离向模糊，精确地实现了对 17 个点目标的正确成像。实验结果证明，本文方法可以有效克服经典成像方法测绘带宽度受距离向最大无模糊距离限制的问题，这对于频率步进 SAR 在高分辨-宽测绘带领域的应用具有重大的现实意义。

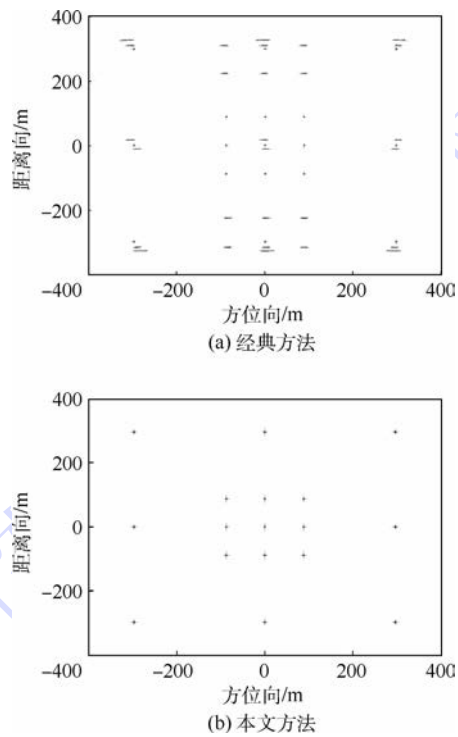


图 4 经典方法和本文方法后向投影成像结果对比

Fig. 4 Comparison of back-projection imaging results between classic method and proposed method

4 结 论

1）频率步进 SAR 虚拟阵列模型成像方法可解决经典方法距离向无模糊测绘带宽度受限的问题，克服频率步进 SAR 在宽测绘带成像时面临的距离模糊问题，有利于频率步进技术在高分辨-宽测绘带 SAR 成像领域的推广应用。

2）虚拟阵列模型高分辨距离像合成方法通

过对各子脉冲回波进行相位补偿和累加,实现对频率步进信号的距离向能量累积,简化了成像处理流程,易于工程实现。

3) 频率步进 SAR 虚拟阵列模型成像方法可实现对目标的二维精确成像,方法复杂度与原始后向投影方法处于同一量级。

后续工作中,将进一步研究方法的快速实现问题,考虑运动误差对方法成像性能的影响并研究相应的运动补偿方法。

参考文献 (References)

[1] 颜志升. 步进频波形优化设计及处理技术[J]. 现代雷达,2015,37(9):40-43.
YAN Z S. Optimized design of the stepped-frequency signal waveform and its processing technology[J]. Modern Radar,2015,37(9):40-43(in Chinese).

[2] 彭岁阳,张军,沈振康. 随机频率步进雷达成像分析[J]. 国防科技大学学报,2011,33(1):59-64.
PENG S Y,ZHANG J,SHEN Z K. Imaging analysis of random step frequency radar[J]. Journal of National University of Defense Technology,2011,33(1):59-64(in Chinese).

[3] 顾福飞,张群,娄昊,等. 一种孔径和频率二维稀疏的步进频 SAR 成像方法[J]. 航空学报,2015,36(4):1221-1229.
GU F F,ZHANG Q,LOU H,et al. A SAR imaging method with 2D sparse aperture and frequency using stepped-frequency waveform[J]. Acta Aeronautica et Astronautica Sinica,2015,36(4):1221-1229(in Chinese).

[4] 吕明久,李少东,杨军,等. 频率步进信号距离-方位联合超分辨成像方法[J]. 电波科学学报,2016,31(3):537-545.
LYU M J,LI S D,YANG J,et al. Range-azimuth joint ISAR super-resolution imaging method based on frequency-stepped signal[J]. Chinese Journal of Radio Science,2016,31(3):537-545(in Chinese).

[5] 李俊慧,王洪,汪学刚,等. 步进频、脉冲和连续波 SAR 的对比研究[J]. 雷达科学与技术,2016,14(1):45-53.
LI J H,WANG H,WANG X G,et al. Comparative study on stepped frequency,pulsed and continuous wave SARs[J]. Radar Science and Technology,2016,14(1):45-53(in Chinese).

[6] 陈一畅,张群,陈校平,等. 多重测量矢量模型下的稀疏步进频率 SAR 成像算法[J]. 电子与信息学报,2014,36(12):2986-2993.
CHEN Y C,ZHANG Q,CHEN X P,et al. An imaging algorithm of sparse stepped frequency SAR based on multiple measurement vectors model[J]. Journal of Electronics & Information Technology,2014,36(12):2986-2993(in Chinese).

[7] 李海英,杨汝良. 频率步进信号的合成孔径雷达处理[J]. 电子学报,2003,31(3):349-352.
LI H Y,YANG R L. Stepped-frequency synthetic aperture radar imaging[J]. Acta Electronica Sinica,2003,31(3):349-352(in Chinese).

[8] 吕明久,李少东,杨军,等. 基于全局最小熵的随机稀疏调频步进信号运动补偿方法[J]. 系统工程与电子技术,2016,38(8):1744-1751.
LYU M J,LI S D,YANG J,et al. Motion-compensation method based on global minimum entropy for random sparse stepped-frequency chirp signal[J]. Systems Engineering and Electronics,2016,38(8):1744-1751(in Chinese).

[9] 梁福来. 低空无人机载 UWB SAR 增强成像技术研究[D]. 长沙:国防科学技术大学,2013:27-35.
LIANG F L. Research on enhanced imaging techniques of low-altitude UAV-mounted UWB SAR[D]. Changsha: National University of Defense Technology,2013:27-35(in Chinese).

[10] 牛涛,陈卫东. 脉冲步进频率雷达的一种运动补偿新方法[J]. 中国科学技术大学学报,2005,35(2):161-166.
NIU T,CHEN W D. A new method of motion compensation for pulse stepped-frequency radars[J]. Journal of University of Science and Technology of China,2005,35(2):161-166(in Chinese).

[11] 李俊慧,王洪,汪学刚,等. 步进频 SAR 的波数域成像算法研究[J]. 现代雷达,2016,38(7):25-31.
LI J H,WANG H,WANG X G,et al. Stepped frequency SAR imaging using wavenumber domain algorithm[J]. Modern Radar,2016,38(7):25-31(in Chinese).

[12] 周芳. 高分辨 SAR/ISAR 成像信号补偿新技术研究[D]. 西安:西安电子科技大学,2014:18-26.
ZHOU F. New signal compensation techniques for high resolution SAR/ISAR imaging[D]. Xi'an: Xidian University,2014:18-26(in Chinese).

[13] VILLANO M,KRIEGER G,MOREIRA A. A novel processing strategy for staggered SAR[J]. IEEE Geoscience and Remote Sensing Letters,2014,11(11):1891-1895.

[14] SCHOLNIK D P. Range-ambiguous clutter suppression with pulse-diverse waveforms[C]// 2011 IEEE Radar Conference. Piscataway,NJ:IEEE Press,2011:336-341.

[15] HALE T B,TEMPLE M A,RAQUET J F,et al. Localized three-dimensional adaptive spatial-temporal processing for airborne radar[J]. IEE Proceedings—Radar, Sonar and Navigation,2003,150(1):18-22.

[16] BAIZERT P,HALE T B,TEMPLE M A,et al. Forward-looking radar GMTI benefits using a linear frequency diverse array[J]. Electronics Letters,2006,42(22):1311-1312.

[17] 王伟伟,吴孙勇,徐京伟,等. 基于频率分集阵列的机载雷达距离模糊杂波抑制方法[J]. 电子与信息学报,2015,37(10):2321-2327.
WANG W W,WU S Y,XU J W,et al. Range ambiguity clutter suppression for airborne radar based on frequency diverse array[J]. Journal of Electronics & Information Technology,2015,37(10):2321-2327(in Chinese).

作者简介:

冉达　男,博士,讲师。主要研究方向:合成孔径雷达成像、雷达信号处理。

尹灿斌　男,博士,讲师。主要研究方向:合成孔径雷达成像及对抗。

贾鑫　男,教授,博士生导师。主要研究方向:合成孔径雷达成像及对抗。

北航学报 赠阅

Imaging approach for frequency-stepped SAR with virtual array model

RAN Da[1,2], YIN Canbin[2,*], JIA Xin[2]

(1. School of Space Information, Space Engineering University, Beijing 101416, China;
2. School of Space Command, Space Engineering University, Beijing 101416, China)

Abstract: Aimed at the problem that frequency-stepped synthetic aperture radar (SAR) images obtained by classic inverse Fourier transform method have a limit on unambiguous range, a new model which takes the pulse sequence of frequency-stepped SAR as an along-track virtual array radar signal and its corresponding imaging method for frequency-stepped SAR were proposed. Meanwhile, unambiguous imaging using modified back-projection method is realized. The virtual array model for frequency-stepped SAR signal was established and the synthesis method of high resolution range profile based on this model was presented. By embedding range migration correction and secondary phase compensation into original back-projection algorithm, a precise two-dimensional image of the target was also obtained. All the results show that the virtual array model based imaging method for frequency-stepped SAR is not restricted by the theoretical limit of frequency-stepped signal's unambiguous range and can get images without range ambiguity for wide swath imaging rapidly.

Keywords: frequency-stepped synthetic aperture radar; virtual array; back-projection; fast Fourier transform (FFT); space time adaptive processing (STAP)

北京航空航天大学学报

Received: 2016-12-15; **Accepted**: 2017-04-24; **Published online**: 2017-08-31 14:18
URL: kns.cnki.net/kcms/detail/11.2625.V.20170831.1418.002.html
* **Corresponding author**. E-mail: ExpressESP@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0014

飞翼无人机机动飞行非线性鲁棒控制方法

李继广[1], 陈欣[1,*], 李亚娟[2], 张榕[1]
(1. 南京航空航天大学 自动化学院, 南京 211106; 2. 南京晓庄学院 信息工程学院, 南京 211171)

摘 要: 针对飞翼布局无人机操纵能力不足的特点,提出了结合流体矢量(FTV-E)控制技术控制策略。设计了内环补偿器以消除系统不利的耦合项,外环控制器采用了反步跟踪算法,并采用粒子群优化(PSO)补偿器补偿各种扰动和不可建模的耦合项的控制方案,证明了控制结构的稳定性。在传统反步控制方法的基础上,增加了内环补偿器。该内环补偿器保留了对飞行有利的气动阻尼项,降低外环控制器的保守性,方便工程实现。仿真结果显示,该控制方案是有效的。

关 键 词: FTV-E 控制; 机动飞行; 控制结构; 输入线性化; 反步控制; PSO 补偿器
中图分类号: V249
文献标识码: A **文章编号**: 1001-5965(2018)01-0089-10

机动飞行是指作战飞机根据任务需求、战场环境等作出的快速改变飞行姿态的飞行行为[1-2]。随着战场复杂化,无人机向高机动性发展已成为不可避免的趋势。一般而言,机动动作是三维空间内的复杂飞行状态,为了研究方便,在控制器设计研究中,通常用二维空间的机动动作组合表示三维空间的机动飞行动作[3-5]。具有机动飞行能力的无人机可以执行更多的作战任务、提高战场生存力,如规避地面防空打击、执行快速移动目标跟踪、拦截巡航导弹、舰上着陆等。

飞翼无人机的机动飞行主要面临两大困难:操纵能力不足和机动飞行条件下的控制律设计。近年来,矢量控制技术和控制设计方法的发展为以上两大困难的解决提供了可能。在提高操纵能力方面,矢量控制技术是最有效的解决方式。除了传统的机械矢量控制方式,新兴起的流体矢量控制技术对于中小型无人机来说更具优势。根据试验结果,和机械喷管相比,流体矢量喷管的优点包括:矢量装置重量减少 24% ~80%;发动机推质比提高 7% ~12%;成本和维护费用降低 37% ~53%;由于流体矢量喷管形状固定没有转动机构,减小了雷达散射截面[6-10]。工程应用方面,英国“恶魔”(Demon)无人技术验证机于 2010 年9 月 17 日成功首飞。该无人机仅依靠喷射气流完成了升降和转向控制。这架无人机成为英国民航管理局(CAA)首次批准并且正式认定“无襟翼飞行”(flapless flight)飞行器。此次试飞成功,对于射流控制技术的实用化具有重大意义。

在控制方面,飞翼无人机具有复杂的耦合性、非线性、非定常、迟滞效应、不确定扰动等问题,使得飞翼无人机的机动飞行控制系统设计面临困难[11-13]。机动飞行飞行状态的快速改变和迟滞效应等使其并不存在典型的平衡状态,这在理论上动摇了传统线性控制方法的基础[13-15]。近年来,非线性控制理论的发展为无人机机动飞行特定环境下的控制律设计的发展开辟了道路。文献[16-21]研究了粒子群优化(PSO)算法,为非线性优化设计提供了可能。文献[22-28]研究非线

收稿日期: 2017-01-12; **录用日期**: 2017-04-13; **网络出版时间**: 2017-05-31 10:28
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170531.1028.002.html
基金项目: 航空科学基金(20160152001); 中央高校基本科研业务费专项资金(NS2015038)
* **通信作者**. E-mail: chenxin@nuaa.edu.cn

引用格式: 李继广, 陈欣, 李亚娟, 等. 飞翼无人机机动飞行非线性鲁棒控制方法[J]. 北京航空航天大学学报, 2018, 44(1): 89-98. LI J G, CHEN X, LI Y J, et al. Nonlinear robust control method for maneuver flight of flying wing UAV[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(1): 89-98 (in Chinese).

北航学报 赠阅

性控制方法在工程实践的应用问题。从上述文献可知,应用动态逆方法为代表的非线性控制理论,解决飞行器机动飞行控制问题的方案已得到广大学者和工程设计人员的高度重视并得到了充分发展和实践。

根据本文采用流体矢量动力的飞翼无人机特定的研究对象,在大迎角机动飞行时除同样存在非线性、非定常及强耦合等问题外,还存在发动机喷流和气动耦合、不可建模部分和扰动的影响,使得无人机的外流畅更加复杂。因此,需要针对该特定的研究对象研究具有强鲁棒性的系统设计方法,以保证系统在各种飞行状态和扰动情况下的鲁棒稳定性和期望的飞行品质。因此,本文采用内外回路分别设计以实现不同控制目的的设计思路。

本文研究重点是飞翼无人机的控制器设计方法。因此,对流体矢量发动机的特性只作简要的介绍,并在该发动机满足控制要求的前提下开展设计。在本文的设计中,为了消除飞翼无人机各种已知的不利非线性耦合项,首先采用气动补偿和线性化解耦方法对内环姿态回路进行补偿;外环跟踪回路采用反步跟踪方法对指定理想指令进行跟踪;为了克服各种不确定性干扰影响,保证跟踪精度采用 PSO 补偿器对跟踪误差进行补偿。本文证明了该控制结构的稳定性。同传统反步控制方法相比,本控制器增加了内环解耦结构。不同于传统的动态逆解耦控制方法,本文在控制结构中保留气动阻尼项,使得线性化后的系统为弱非线性系统。该结构不仅可以降低外环控制器设计的保守性,而且便于工程实现。

1 飞翼无人机气动建模

样例飞翼无人机如图 1 所示。该无人机具有位于翼面尾部的升降副翼舵、机翼上的阻力式方向舵。其中,升降副翼舵可以通过分配算法进行制动,联动为升降舵,差动为副翼。

该无人机的发动机具有流体矢量装置,不仅提供推力,还可以提供矢量控制力矩。该发动机的喷流可以在垂直方向产生 20°的矢量角,以产生俯仰控制力矩。该流体矢量发动机结构如图 2 所示。

该流体矢量发动机具有复杂的控制特性。其具体响应特性如图 3 所示,p_i、Δp 分别为控制量和改变范围。

从图 3 可知,FTV-E 的控制动态响应类似于无人机大迎角条件下的迟滞效应。因此,对该装置的控制难点有两点:非线性控制方法和控制舵机的响应速度。该装置的控制问题已通过合理的控制律设计方法和脉宽调制(Pulse Width Modulator,PWM)控制舵机得到了良好的解决。本文在该装置满足控制要求的前提下分析该装置与无人机外流场的相互影响对无人机控制系统设计的带来的挑战。

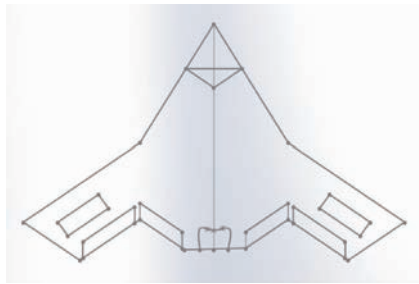


图 1 样例无人机结构图

Fig. 1 Configuration of a sample UAV

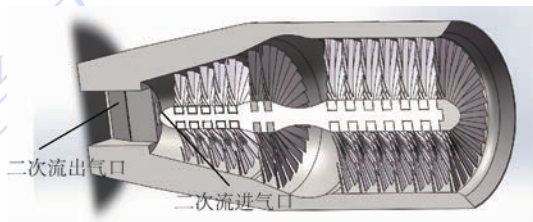


图 2 流体矢量涡轮增压发动机 SolidWorks 模型

Fig. 2 Model of fluidic thrust vectoring-turbocharged engine by SolidWorks

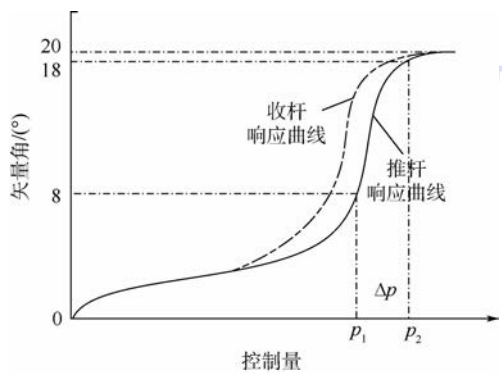


图 3 FTV-E 控制过程的动态响应

Fig. 3 Dynamic response of control process by FTV-E

在流体矢量发动机的控制满足要求的假设下,该该装置产生的力与力矩与传统的矢量发动机相似,因此,本文借鉴具有矢量发动机的无人机建模方式进行建模。而本文假设发动机矢量流与无人机外流场之间的耦合效应为不可建模部分,以干扰的形式来体现。

根据具体的研究对象和后文控制器设计表述方便,对飞翼无人机建立姿态方程如下模型:

$$\dot{\boldsymbol{\omega}} = -\boldsymbol{I}^{-1}\boldsymbol{S}(\boldsymbol{\omega})\boldsymbol{I}\boldsymbol{\omega} + \boldsymbol{I}^{-1}(\boldsymbol{M}_{\mathrm{T}} + \boldsymbol{M}_{\mathrm{A}}) \tag{1}$$

式中：$\boldsymbol{\omega} = [p, q, r]^{\mathrm{T}}$ 为飞翼无人机姿态角速度矩阵，p、q、r 为无人机的姿态角速率；$\boldsymbol{M}_{\mathrm{A}}$ 为气动力矩；$\boldsymbol{M}_{\mathrm{T}}$ 为矢量力矩；$\boldsymbol{I}$ 和 $\boldsymbol{S}(\boldsymbol{\omega})$ 为惯性矩阵和 $\boldsymbol{\omega}$ 的反对称矩阵，具体的形式如下：

$$\boldsymbol{I} = \begin{bmatrix} I_x & 0 & -I_{xz} \\ 0 & I_y & 0 \\ -I_{xz} & 0 & I_z \end{bmatrix}$$

$$\boldsymbol{S}(\boldsymbol{\omega}) = \begin{bmatrix} 0 & -r & q \\ r & 0 & -p \\ -q & p & 0 \end{bmatrix}$$

其中：I_x、I_y、I_z 和 I_{xz} 为无人机的转动惯量。

飞翼无人机的姿态方程、气流坐标系下的速度和气流角方程如下：

$$\begin{bmatrix} \dot{\varphi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} 1 & \sin\varphi\tan\theta & \cos\varphi\tan\theta \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi/\cos\theta & \cos\varphi/\cos\theta \end{bmatrix} \begin{bmatrix} p \\ q \\ r \end{bmatrix} \tag{2}$$

$$\dot{V} = \frac{\cos\alpha\cos\beta}{m}(T_x + F_x) + \frac{\sin\beta}{m}F_y + \frac{\sin\alpha\cos\beta}{m}(T_z + F_z) + g(-\cos\alpha\cos\beta\sin\theta + \sin\varphi\sin\beta\cos\theta + \sin\alpha\cos\beta\cos\varphi\cos\theta) \tag{3}$$

$$\dot{\alpha} = -p\cos\alpha\tan\beta + q - r\sin\alpha\tan\beta - \frac{\sin\alpha}{mV\cos\beta}(T_x + F_x) + \frac{\cos\alpha}{mV\cos\beta}(T_z + F_z) + \frac{g}{V\cos\beta}(\sin\alpha\sin\theta + \cos\alpha\cos\varphi\cos\theta) \tag{4}$$

$$\dot{\beta} = p\sin\alpha - r\cos\alpha - \frac{\cos\alpha\sin\beta}{mV}(T_x + F_x) + \frac{\cos\beta}{mV}F_y - \frac{\sin\alpha\sin\beta}{mV}(T_z + F_z) + \frac{g}{V}(\cos\alpha\sin\beta\sin\theta + \cos\theta\cos\beta\sin\varphi - \sin\alpha\sin\beta\cos\theta) \tag{5}$$

式中：α、β 和 θ、φ 为无人机的气流角和姿态角；其余参量的符号表示均参考惯用表达习惯，具体可参见文献[2-4]。

2　控制器结构设计

对于控制而言，理想的控制对象为线性系统。但是，实际中的对象多为非线性系统，特别是机动飞行的飞翼无人机，其非线性特性非常强烈。为了协调这对矛盾，于是诞生了非线性系统的线性化方法并得到了快速发展。

传统的动态逆解耦线性化控制方法是通过系统的状态反馈补偿掉系统所有的非线性项。具体到无人机则包括气动力/力矩非线性及耦合项、惯性耦合项和重力耦合项等。而在实际中，惯性耦合项、重力耦合项和部分气动耦合项等对控制效果影响较大，设计过程中需要对这些项进行线性化才能达到理想的控制效果。然而，气动非线性项中的阻尼项对无人机的动稳定性保持是有利的，保留下来可以减少操作舵面的负担。保留气动阻尼项的系统是包含时变参数的弱非线性系统，但是，在设计外环控制器时，具有降低系统的保守性，而且便于工程实现。本文控制器结构如图 4 所示。

在图 4 所示的控制结构中，非线性解耦模块、惯性力矩耦合补偿模块和非线性非定常气动力矩补偿模块构成系统的内环控制器。经过内环控制器的补偿、解耦之后，系统为保留气动阻尼项的弱非线性系统。外环控制器采用反步跟踪算法进行设计以满足系统对希望指令的跟踪要求。PSO 补偿器用来消除系统的跟踪误差，提高系统的鲁棒性。

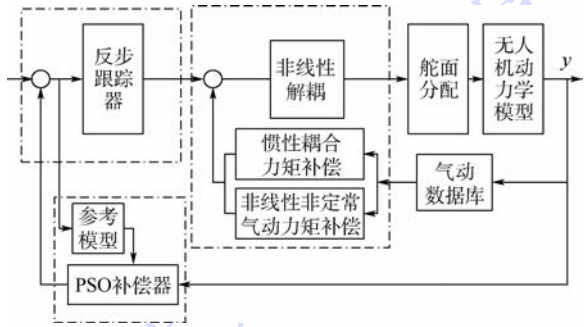


图 4　控制器结构图

Fig. 4　Controller structure diagram

3　姿态方程的线性化解耦

3.1　保留气动阻尼项的力矩补偿

为飞翼无人机进行各种气动补偿且保留气动阻尼项，本文设计了如图 5 所示的气动补偿模块[29]。

把式(1)中气动力矩表示为如下形式：

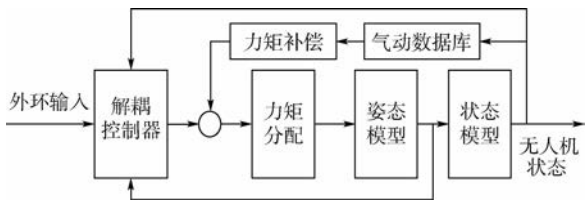


图 5　控制器内环结构[29]

Fig. 5　Structure of inner loop nonlinear controller[29]

北航学报 赠阅

$$\boldsymbol{M}_{\mathrm{A}}=\boldsymbol{\Theta}_1(c_x)\boldsymbol{\varphi}_1+\boldsymbol{\Theta}_2(c_x)\boldsymbol{\delta}_{\mathrm{A}} \tag{6}$$

式中：$\boldsymbol{\delta}_{\mathrm{A}}=[\delta_{\mathrm{a}},\delta_{\mathrm{e}},\delta_{\mathrm{r}}]^{\mathrm{T}}$ 为飞翼无人机气动舵面组成的矩阵，δ_{a}、δ_{e}、δ_{r} 分别为副翼、升降舵阻力式方向舵；$\boldsymbol{\varphi}_1=[1,\alpha,\beta,p,q,r]^{\mathrm{T}}$。

$$\boldsymbol{\Theta}_1(c_x)=Qs\begin{bmatrix}b&0&0\\0&\bar{c}&0\\0&0&b\end{bmatrix}\cdot\begin{bmatrix}C_{l0}&0&C_l^{\beta}&C_l^{p}&0&C_l^{r}\\C_{m0}&C_m^{\alpha}&0&0&C_m^{q}&0\\C_{n0}&0&C_n^{\beta}&C_n^{p}&0&C_n^{r}\end{bmatrix}$$

其中：Q、s 分别为动压和机翼面积；b、$\bar{c}$ 分别为机翼展长和平均气动弦长；其他参数 C_{l0}、C_l^{β}、C_l^{r}、C_l^{p}、C_l^{r}、C_{m0}、C_m^{α}、C_m^{q}、C_n^{β}、C_n^{p}、C_{n0}、C_n^{r} 为无人机的气动参数。

$$\boldsymbol{\Theta}_2(c_x)=Qs\begin{bmatrix}-b&0&0\\0&\bar{c}&0\\0&0&b\end{bmatrix}\begin{bmatrix}-C_l^{\delta_{\mathrm{a}}}&0&C_l^{\delta_{\mathrm{r}}}\\0&C_m^{\delta_{\mathrm{e}}}&0\\C_n^{\delta_{\mathrm{a}}}&0&C_n^{\delta_{\mathrm{e}}}\end{bmatrix}$$

其中：$C_l^{\delta_{\mathrm{a}}}$、$C_l^{\delta_{\mathrm{r}}}$、$C_m^{\delta_{\mathrm{e}}}$、$C_n^{\delta_{\mathrm{e}}}$、$C_n^{\delta_{\mathrm{a}}}$ 为飞翼无人机的气动参数。

推力矢量力矩可以通过油门开度 δ_{th} 和射流偏角 δ_{T} 控制。令推力矢量作用点到质心的距离沿机体坐标系到 X 轴和 Z 轴的投影距离分别为 X_{T} 和 Z_{T}，推力产生的力矩可表示为

$$\boldsymbol{M}_{\mathrm{T}}=\boldsymbol{T}(\delta_{\mathrm{th}})\begin{bmatrix}0\\-Z_{\mathrm{T}}\cos\delta_{\mathrm{T}}-X_{\mathrm{T}}\sin\delta_{\mathrm{T}}\\0\end{bmatrix}\approx$$

$$\boldsymbol{T}(\delta_{\mathrm{th}})\begin{bmatrix}0\\-Z_{\mathrm{T}}-X_{\mathrm{T}}\delta_{\mathrm{T}}\\0\end{bmatrix}=\begin{bmatrix}0\\-\boldsymbol{T}(\delta_{\mathrm{th}})Z_{\mathrm{T}}\\0\end{bmatrix}+$$

$$\begin{bmatrix}0\\-\boldsymbol{T}(\delta_{\mathrm{th}})X_{\mathrm{T}}\\0\end{bmatrix}\delta_{\mathrm{T}}=\boldsymbol{M}_{\mathrm{T0}}+\boldsymbol{\Theta}_3\delta_{\mathrm{T}} \tag{7}$$

式中：$\boldsymbol{T}(\delta_{\mathrm{th}})$ 为发动机推力函数。令气动力矩 $\boldsymbol{\Theta}_i$ 和发动机推力 $\boldsymbol{M}_{\mathrm{T0}}$ 为含摄动函数的非线性函数

$$\begin{cases}\boldsymbol{\Theta}_i=\bar{\boldsymbol{\Theta}}_i+\Delta\boldsymbol{\Theta}_i & i=1,2,3\\ \boldsymbol{M}_{\mathrm{T0}}=\bar{\boldsymbol{M}}_{\mathrm{T0}}+\Delta\boldsymbol{M}_{\mathrm{T0}}\end{cases}$$

其中：$\bar{\boldsymbol{\Theta}}_i$ 和 $\bar{\boldsymbol{M}}_{\mathrm{T0}}$ 分别为气动力矩和发动机推力的标称值，这两组值为慢变量；$\Delta\boldsymbol{\Theta}_i$ 和 $\Delta\boldsymbol{M}_{\mathrm{T0}}$ 分别为气动力矩和发动机推力的摄动值，是随时间而变化的快变量。

将式(6)、式(7)代入式(1)得

$$\dot{\boldsymbol{\omega}}=\bar{\boldsymbol{f}}+\Delta\bar{\boldsymbol{f}}+(\bar{\boldsymbol{g}}+\Delta\boldsymbol{g})\boldsymbol{\delta} \tag{8}$$

式中：$\bar{\boldsymbol{f}}=-\boldsymbol{I}^{-1}\boldsymbol{S}(\boldsymbol{\omega})\boldsymbol{I}\boldsymbol{\omega}+\boldsymbol{I}^{-1}\bar{\boldsymbol{\Theta}}_1\boldsymbol{\varphi}_1+\boldsymbol{I}^{-1}\bar{\boldsymbol{M}}_{\mathrm{T0}}$；$\Delta\bar{\boldsymbol{f}}=\boldsymbol{I}^{-1}\Delta\bar{\boldsymbol{\Theta}}_1\boldsymbol{\varphi}_1+\boldsymbol{I}^{-1}\Delta\bar{\boldsymbol{M}}_{\mathrm{T0}}$；$\bar{\boldsymbol{g}}=\boldsymbol{I}^{-1}[\bar{\boldsymbol{\Theta}}_2,\bar{\boldsymbol{\Theta}}_3]$；$\Delta\bar{\boldsymbol{g}}=\boldsymbol{I}^{-1}[\Delta\bar{\boldsymbol{\Theta}}_2,\Delta\bar{\boldsymbol{\Theta}}_3]$；$\boldsymbol{\delta}=[\boldsymbol{\delta}_{\mathrm{A}}^{\mathrm{T}},\delta_{\mathrm{T}}]^{\mathrm{T}}$。

系统标称模型为

$$\dot{\bar{\boldsymbol{\omega}}}=-\boldsymbol{I}^{-1}\boldsymbol{S}(\boldsymbol{\omega})\boldsymbol{I}\boldsymbol{\omega}+\boldsymbol{I}^{-1}\bar{\boldsymbol{\Theta}}_1\boldsymbol{\varphi}_1-\bar{\boldsymbol{K}}\boldsymbol{\omega}+\boldsymbol{I}^{-1}\bar{\boldsymbol{M}}_{\mathrm{T0}}+\bar{\boldsymbol{K}}\boldsymbol{\omega}+\boldsymbol{u}_{\mathrm{cm}}+\boldsymbol{u}_{\mathrm{lin}}=\boldsymbol{f}_0+\bar{\boldsymbol{K}}\boldsymbol{\omega}+\boldsymbol{u}_{\mathrm{cm}}+\boldsymbol{u}_{\mathrm{lin}} \tag{9}$$

式中：$\boldsymbol{f}_0=\bar{\boldsymbol{f}}-\bar{\boldsymbol{K}}\boldsymbol{\omega}$ 为对飞行不利，需要消除掉的耦合力矩，包括不利的气动耦合力矩、推力耦合力矩、惯性耦合力矩等；$\bar{\boldsymbol{K}}=\mathrm{diag}(\bar{k}_1,\bar{k}_2,\bar{k}_3)$ 为各控制通道的阻尼力矩的标称值，$\bar{k}_1=h_{11}\bar{l}_{\mathrm{p}}+h_{13}\bar{n}_{\mathrm{p}}$，$\bar{k}_2=h_{22}\bar{m}_{\mathrm{p}}$，$\bar{k}_3=h_{31}\bar{l}_{\mathrm{r}}+h_{33}\bar{n}_{\mathrm{r}}$，$h_{ij}(i,j=1,2,3)$ 为惯性矩阵 $\boldsymbol{I}^{-1}$ 的第 i 行和 j 列分量，$\bar{l}_{\mathrm{p}}$、$\bar{n}_{\mathrm{p}}$、$\bar{m}_{\mathrm{p}}$、$\bar{l}_{\mathrm{r}}$、$\bar{n}_{\mathrm{r}}$ 为各通道的阻尼力矩气动参数；$\boldsymbol{u}_{\mathrm{cm}}$、$\boldsymbol{u}_{\mathrm{lin}}$ 为气动补偿控制律、外环控制律；$\bar{\boldsymbol{\Theta}}_1$、$\bar{\boldsymbol{M}}_{\mathrm{T0}}$ 和 $\bar{\boldsymbol{K}}$ 可以通过气动数据库以及推力数据库和飞翼无人机当前飞行状态数据进行插值计算得到。

该气动补偿模块存在的意义是在保留阻尼特性的情况下消除其他非线性特性。因此，补偿力矩为

$$\boldsymbol{u}_{\mathrm{cm}}=\boldsymbol{I}^{-1}(\boldsymbol{S}(\boldsymbol{\omega})\boldsymbol{I}\boldsymbol{\omega}-\bar{\boldsymbol{\Theta}}_1\boldsymbol{\varphi}_1-\bar{\boldsymbol{M}}_{\mathrm{T0}})+\bar{\boldsymbol{K}}\boldsymbol{\omega} \tag{10}$$

通过式(10)所示的气动补偿器可以消除各种影响飞行品质的耦合力矩。

将式(10)代入式(9)可得

$$\dot{\boldsymbol{\omega}}=\bar{\boldsymbol{K}}\boldsymbol{\omega}+\boldsymbol{u}_{\mathrm{lin}} \tag{11}$$

因此，以 $\boldsymbol{u}_{\mathrm{lin}}$ 为控制量进行外环控制器设计时可以保证系统具有较好的闭环性能。

3.2 姿态方程的线性化

为了消除系统状态间的非线性特性，下面介绍采用基于李导数的输入/输出线性化方法。步骤如下：

1) 选择在 x° 邻域内的坐标变换函数集。

$$\xi_k^i(k)=\varphi_k^i(x)=L_f^{k-1}h_i(x)$$
$$1\leqslant k\leqslant r_i,1\leqslant i\leqslant m \tag{12}$$

2) 把系统方程变换为 m 组方程并表示为

$$\begin{cases}\dot{\xi}_1^i=\xi_2^i\\ \dot{\xi}_2^i=\xi_3^i\\ \vdots\\ \dot{\xi}_{r_i-1}^i=\xi_{r_i}^i\\ \dot{\xi}_{r_i}^i=b_i(\xi)+\sum_{j=1}^{m}a_{ij}(\xi)u_j\end{cases}\quad 1\leqslant i\leqslant m \tag{13}$$

式中：

$$\begin{cases}b_i(\xi)=L_f^{r_i}h_i(\boldsymbol{\Phi}^{-1}(\xi))\\ a_{ij}(\xi)=L_{g_j}L_f^{r_i-1}h_i(\boldsymbol{\Phi}^{-1}(\xi))\\ \boldsymbol{\Phi}(x)=[\phi_1^1(x),\phi_2^1(x),\cdots,\phi_{r_1}^1(x),\cdots,\\ \qquad\phi_1^m(x),\phi_2^m(x),\cdots,\phi_{r_m}^m(x)]^{\mathrm{T}}\end{cases} \tag{14}$$

3) 把 $\dot{\xi}_{ri}^i=b_i(\xi)+\sum_{j=1}^{m}a_{ij}(\xi)u_j(1\leqslant i\leqslant m)$ 写成

矩阵形式

$$\begin{bmatrix}\dot{\xi}_{r_1}^1\\ \dot{\xi}_{r_2}^2\\ \vdots\\ \dot{\xi}_{r_m}^m\end{bmatrix}=\boldsymbol{b}(\xi)+\boldsymbol{a}(\xi)\boldsymbol{u}=\begin{bmatrix}b_1(\xi)\\ b_2(\xi)\\ \vdots\\ b_m(\xi)\end{bmatrix}+\begin{bmatrix}a_{11}(\xi) & a_{12}(\xi) & \cdots & a_{1m}(\xi)\\ a_{21}(\xi) & a_{22}(\xi) & \cdots & a_{2m}(\xi)\\ \vdots & \vdots & & \vdots\\ a_{m1}(\xi) & a_{m2}(\xi) & \cdots & a_{mm}(\xi)\end{bmatrix}\begin{bmatrix}u_1\\ u_2\\ \vdots\\ u_m\end{bmatrix} \tag{15}$$

令 $\boldsymbol{b}(\xi)+\boldsymbol{a}(\xi)\boldsymbol{u}=\boldsymbol{v}=[v_1,v_2,\cdots,v_m]^{\mathrm{T}}$，则精确线性化的反馈控制律为

$$\boldsymbol{u}_{\mathrm{lin}}=\boldsymbol{a}^{-1}(\xi)(-\boldsymbol{b}(\xi)+\boldsymbol{v}) \tag{16}$$

经过上述坐标变换和状态反馈，原非线性系统可以变换为 Brunovsky 标准形：

$$\begin{cases}\dot{\xi}_1^i=\xi_2^i\\ \dot{\xi}_2^i=\xi_3^i\\ \vdots\\ \dot{\xi}_{r_i-1}^i=\xi_{r_i}^i\\ \dot{\xi}_{r_i}^i=v_i\end{cases}\qquad 1\leqslant i\leqslant m \tag{17}$$

对于飞翼无人机具体的方程形式，由式(14)可得系统的坐标变换函数为

$$\boldsymbol{\Phi}(x)=\begin{bmatrix}\phi\\ p+q\sin\phi\tan\theta+r\cos\phi\tan\theta\\ \theta\\ q\cos\phi-r\sin\phi\\ \psi\\ q\sin\phi\sec\theta+r\cos\phi\sec\theta\end{bmatrix} \tag{18}$$

由式(13)可得

$$\boldsymbol{a}=\begin{bmatrix}L_{g_1}L_f\lambda_1 & L_{g_2}L_f\lambda_1 & L_{g_3}L_f\lambda_1\\ L_{g_1}L_f\lambda_2 & L_{g_2}L_f\lambda_2 & L_{g_3}L_f\lambda_2\\ L_{g_1}L_f\lambda_3 & L_{g_2}L_f\lambda_3 & L_{g_3}L_f\lambda_3\end{bmatrix} \tag{19}$$

$$\boldsymbol{b}=[L_f^2\lambda_1,L_f^2\lambda_2,L_f^2\lambda_3]^{\mathrm{T}} \tag{20}$$

将飞翼无人机解耦后的姿态方程和轨迹方程分别表示为

$$\dot{\boldsymbol{x}}_1=\boldsymbol{f}_1(\boldsymbol{x}_1)+\boldsymbol{g}_{s1}(\boldsymbol{x}_4)\boldsymbol{x}_2+\boldsymbol{g}_{s2}(\boldsymbol{x}_4)\boldsymbol{u}$$

$$\dot{\boldsymbol{x}}_2=\boldsymbol{f}_{\mathrm{f}}(\boldsymbol{x}_4)+\boldsymbol{g}_{\mathrm{f}}(\boldsymbol{x}_4)\boldsymbol{u}$$

其中：状态向量 $\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\boldsymbol{u}\in\mathbf{R}^3$，且 $\boldsymbol{x}_1=[\alpha,\beta,\varphi]^{\mathrm{T}}$，$\boldsymbol{x}_2=[p,q,r]^{\mathrm{T}}$，$\boldsymbol{x}_3=[\varphi,\theta,\psi]^{\mathrm{T}}$，$\boldsymbol{u}=[\delta_{\mathrm{e}},\delta_{\mathrm{a}},\delta_{\mathrm{r}},\delta_{\mathrm{T}},\delta_{\mathrm{th}}]^{\mathrm{T}}$，$\boldsymbol{x}_4=[x_1,x_2]^{\mathrm{T}}$。

4　粒子群优化算法

PSO 补偿器的目的是提高系统的鲁棒性，减小各种干扰对系统响应的影响。

PSO 算法对于任意非线性函数 $\boldsymbol{\Delta}:\Omega\rightarrow\mathbf{R}^p$，PSO 算法得出的结果存在常数 $\varepsilon>0$ 使得 $\|\boldsymbol{\Delta}-\boldsymbol{W}^{*\mathrm{T}}\boldsymbol{\phi}_1\|_{\mathrm{F}}^2\leqslant\varepsilon$，其中，$\|*\|_{\mathrm{F}}^2$ 为向量的 2 范数，$\boldsymbol{W}^*\in\mathbf{R}^{n\times p}$ 为 PSO 算法输出的理想权值矩阵，ε 为算法的收敛误差。

为提高 PSO 算法的效率，本文采用式(21)所示的混沌函数对算法粒子初置进行初始化，式(22)确定权重系数。

$$y(n+1)=4y^3(n)-3y(n)\qquad -1\leqslant y(n)\leqslant 1 \tag{21}$$

$$\omega(n)=\omega_{\mathrm{start}}-(\omega_{\mathrm{start}}-\omega_{\mathrm{end}})(n/T_{\max})^2 \tag{22}$$

式中：n 为粒子群粒子个数；ω_{start} 和 ω_{end} 分别为初始权重系数和迭代最大次数时的权重系数；$T_{\max}$ 为迭代最大次数。

5　外环控制器设计

外环控制器的作用是根据飞行器的当前飞行状态和输入指令输出自适应律实现对气流角的输入指令进行精确跟踪。使得系统的输出 $\boldsymbol{x}_1$ 渐进跟踪期望的参考输入 $\boldsymbol{x}_{1\mathrm{d}}$。

定义系统的误差为

$$\begin{cases}\tilde{\boldsymbol{\vartheta}}=\boldsymbol{\vartheta}^*-\hat{\boldsymbol{\vartheta}}\\ \tilde{\boldsymbol{\zeta}}=\boldsymbol{\zeta}^*-\hat{\boldsymbol{\zeta}}\end{cases} \tag{23}$$

式中：$\hat{\boldsymbol{\vartheta}}$、$\hat{\boldsymbol{\zeta}}$为理想权值；$\boldsymbol{\vartheta}^*$、$\boldsymbol{\zeta}^*$为 PSO 算法自适应律输出的权值。

系统状态误差可定义为

$$\boldsymbol{z}_1=\boldsymbol{x}_1-\boldsymbol{x}_{1\mathrm{d}}=[z_{11},z_{21},z_{31}]^{\mathrm{T}} \tag{24}$$

$$\boldsymbol{z}_2=\boldsymbol{x}_2-\boldsymbol{x}_{2\mathrm{d}}=[z_{12},z_{22},z_{32}]^{\mathrm{T}} \tag{25}$$

式中：$\boldsymbol{x}_{2\mathrm{d}}$为参考输入。

对式(24)、式(25)求导，并代入闭环误差状态方程可得

$$\dot{\boldsymbol{z}}_1=\boldsymbol{f}_1+\boldsymbol{g}_{s1}\boldsymbol{x}_2+\boldsymbol{\vartheta}^{*\mathrm{T}}\boldsymbol{\phi}_1-\dot{\boldsymbol{x}}_{1\mathrm{d}} \tag{26}$$

$$\dot{\boldsymbol{z}}_2=\boldsymbol{f}_{\mathrm{f}}+\boldsymbol{g}_{\mathrm{f}}\boldsymbol{u}+\boldsymbol{\zeta}^{*\mathrm{T}}\boldsymbol{\phi}_2-\dot{\boldsymbol{x}}_{2\mathrm{d}} \tag{27}$$

式中：$\boldsymbol{\phi}_1$、$\boldsymbol{\phi}_2$ 为防止由于控制器权值过大而破坏系统的稳定性而引入的参数向量。

由式(26)可得

$$\dot{\boldsymbol{z}}_1=\boldsymbol{f}_1+\boldsymbol{g}_{s1}\boldsymbol{x}_{2\mathrm{d}}+\boldsymbol{g}_{s1}\boldsymbol{z}_2+\boldsymbol{\vartheta}^{*\mathrm{T}}\boldsymbol{\phi}_1-\dot{\boldsymbol{x}}_{1\mathrm{d}} \tag{28}$$

定义 $\boldsymbol{z}_1$ 系统的理想控制量为

$$\ddot{\boldsymbol{x}}_{2\mathrm{d}}^*=\boldsymbol{g}_{s1}^{-1}(-\boldsymbol{k}_1\boldsymbol{z}_1-\boldsymbol{f}_1-\boldsymbol{\vartheta}^{*\mathrm{T}}\boldsymbol{\phi}_1+\dot{\boldsymbol{x}}_{1\mathrm{d}}) \tag{29}$$

式中：$\boldsymbol{k}_1>0$ 为需要设计参数向量。因为 $\boldsymbol{\vartheta}^*$ 是未知量，这里用估计值$\hat{\boldsymbol{\vartheta}}$代替。系统的虚拟控制量可表示为

$$\boldsymbol{x}_{2\mathrm{d}}=\boldsymbol{g}_{s1}^{-1}(-\boldsymbol{k}_1\boldsymbol{z}_1-\boldsymbol{f}_1-\hat{\boldsymbol{\vartheta}}^{\mathrm{T}}\boldsymbol{\phi}_1+\dot{\boldsymbol{x}}_{1\mathrm{d}}) \tag{30}$$

将式(30)代入式(28)可得

$$\dot{\boldsymbol{z}}_1=-\boldsymbol{k}_1\boldsymbol{z}_1+\boldsymbol{g}_{s1}\boldsymbol{z}_2+\tilde{\boldsymbol{\vartheta}}^{\mathrm{T}}\boldsymbol{\phi}_1 \tag{31}$$

根据式(27),定义系统的理想控制输入为

$$\boldsymbol{u}^{*}=\boldsymbol{g}_{\mathrm{f}}^{-1}(-\boldsymbol{k}_{2}z_{2}-\boldsymbol{f}_{\mathrm{f}}-\boldsymbol{g}_{\mathrm{s1}}^{\mathrm{T}}z_{1}-\boldsymbol{\zeta}^{*\mathrm{T}}\boldsymbol{\phi}_{2}+\dot{\boldsymbol{x}}_{2\mathrm{d}}) \tag{32}$$

式中:$\boldsymbol{k}_{2}$ 为需要设计的参数向量。

同上

$$\boldsymbol{u}=\boldsymbol{g}_{\mathrm{f}}^{-1}(-\boldsymbol{k}_{2}z_{2}-\boldsymbol{f}_{\mathrm{f}}-\boldsymbol{g}_{\mathrm{s1}}^{\mathrm{T}}z_{1}-\tilde{\boldsymbol{\zeta}}^{\mathrm{T}}\boldsymbol{\phi}_{2}+\dot{\boldsymbol{x}}_{2\mathrm{d}}) \tag{33}$$

代入式(27)可得

$$\dot{z}_{2}=-\boldsymbol{k}_{2}-\boldsymbol{g}_{\mathrm{s1}}^{\mathrm{T}}z_{1}+\tilde{\boldsymbol{\zeta}}^{\mathrm{T}}\boldsymbol{\phi}_{2} \tag{34}$$

6 控制器稳定性分析

分析和证明系统稳定性需要选取 Lyapunov 函数,则

$$\begin{aligned}\boldsymbol{V}(t)&=\frac{1}{2}z_{1}^{\mathrm{T}}z_{1}+\frac{1}{2}z_{2}^{\mathrm{T}}z_{2}+\frac{1}{2\lambda_{1}}\mathrm{tr}(\tilde{\boldsymbol{\vartheta}}^{\mathrm{T}}\tilde{\boldsymbol{\vartheta}})+\\&\frac{1}{2\lambda_{2}}\mathrm{tr}(\tilde{\boldsymbol{\zeta}}^{\mathrm{T}}\tilde{\boldsymbol{\zeta}})=\frac{1}{2}z_{1}^{\mathrm{T}}z_{1}+\frac{1}{2}z_{2}^{\mathrm{T}}z_{2}+\\&\frac{1}{2\lambda_{1}}\|\tilde{\boldsymbol{\vartheta}}\|_{\mathrm{F}}^{2}+\frac{1}{2\lambda_{2}}\|\tilde{\boldsymbol{\zeta}}\|_{\mathrm{F}}^{2}\end{aligned} \tag{35}$$

将 $\tilde{\boldsymbol{\vartheta}}$ 和 $\tilde{\boldsymbol{\zeta}}$ 写成分块矩阵的形式,则式(35)可写为

$$\boldsymbol{V}(t)=\frac{1}{2}z_{1}^{\mathrm{T}}z_{1}+\frac{1}{2}z_{2}^{\mathrm{T}}z_{2}+\frac{1}{2\lambda_{1}}\sum_{i=1}^{3}\tilde{\boldsymbol{\vartheta}}_{i}^{\mathrm{T}}\tilde{\boldsymbol{\vartheta}}_{i}+\frac{1}{2\lambda_{2}}\sum_{i=1}^{3}\tilde{\boldsymbol{\zeta}}_{i}^{\mathrm{T}}\tilde{\boldsymbol{\zeta}}_{i} \tag{36}$$

对 $\boldsymbol{V}(t)$ 对时间 t 求导可得

$$\begin{aligned}\dot{\boldsymbol{V}}(t)&=-k_{1}z_{1}^{\mathrm{T}}z_{1}-k_{2}z_{2}^{\mathrm{T}}z_{2}+z_{1}^{\mathrm{T}}\tilde{\boldsymbol{\vartheta}}^{\mathrm{T}}\boldsymbol{\phi}_{1}+z_{2}^{\mathrm{T}}\tilde{\boldsymbol{\zeta}}^{\mathrm{T}}\phi_{2}+\\&\frac{1}{\lambda_{1}}\sum_{i=1}^{3}\tilde{\boldsymbol{\vartheta}}_{i}^{\mathrm{T}}\dot{\tilde{\boldsymbol{\vartheta}}}_{i}+\frac{1}{\lambda_{2}}\sum_{i=1}^{3}\tilde{\boldsymbol{\zeta}}_{i}^{\mathrm{T}}\dot{\tilde{\boldsymbol{\zeta}}}_{i}\end{aligned} \tag{37}$$

令 PSO 的权值自适应律为

$$\begin{cases}\dot{\hat{\boldsymbol{\vartheta}}}_{i}=\tilde{\lambda}_{1}(z_{i1}\boldsymbol{\phi}_{1}-\alpha_{i}\hat{\boldsymbol{\vartheta}}_{i})\\\dot{\hat{\boldsymbol{\zeta}}}_{i}=\tilde{\lambda}_{2}(z_{i2}\boldsymbol{\phi}_{2}-\beta_{i}\hat{\boldsymbol{\zeta}}_{i})\end{cases} \tag{38}$$

式中:$\tilde{\lambda}_{1}$、$\tilde{\lambda}_{2}$ 为自适应补偿器需要设计的大于零的实数;α_{i}、$\hat{\boldsymbol{\vartheta}}_{i}$、$\beta_{i}$、$\hat{\boldsymbol{\zeta}}_{i}$ 为鲁棒控制项,其作用是防止由于 PSO 权值过大而破坏系统稳定性。

将以上自适应律代入式(37)可得

$$\begin{aligned}\dot{\boldsymbol{V}}(t)&=-k_{1}\|z_{1}\|_{\mathrm{F}}^{2}-k_{2}\|z_{2}\|_{\mathrm{F}}^{2}+\sum_{i=1}^{3}\alpha_{i}\tilde{\boldsymbol{\vartheta}}_{i}^{\mathrm{T}}\tilde{\boldsymbol{\vartheta}}_{i}+\\&\sum_{i=1}^{3}\beta_{i}\tilde{\boldsymbol{\zeta}}_{i}^{\mathrm{T}}\tilde{\boldsymbol{\zeta}}_{i}=-(k_{10}+k_{11})\|z_{1}\|_{\mathrm{F}}^{2}-(k_{20}+\\&k_{21})\|z_{2}\|_{\mathrm{F}}^{2}+\sum_{i=1}^{3}(\alpha_{i}\tilde{\boldsymbol{\vartheta}}_{i}^{\mathrm{T}}\tilde{\boldsymbol{\vartheta}}_{i}+\beta_{i}\tilde{\boldsymbol{\zeta}}_{i}^{\mathrm{T}}\tilde{\boldsymbol{\zeta}})\end{aligned} \tag{39}$$

式中:k_{10}、k_{11}、k_{20}、k_{21} 为需要设计的参数。

由于不等式(40)成立:

$$\begin{cases}\tilde{\boldsymbol{\vartheta}}_{i}^{\mathrm{T}}\tilde{\boldsymbol{\vartheta}}_{i}\leqslant 0.5(\|\boldsymbol{\vartheta}_{i}^{*}\|_{\mathrm{F}}^{2}-\|\tilde{\boldsymbol{\vartheta}}_{i}\|_{\mathrm{F}}^{2})\\\tilde{\boldsymbol{\zeta}}_{i}^{\mathrm{T}}\tilde{\boldsymbol{\zeta}}\leqslant 0.5(\|\boldsymbol{\zeta}_{i}^{*}\|_{\mathrm{F}}^{2}-\|\tilde{\boldsymbol{\zeta}}_{i}\|_{\mathrm{F}}^{2})\end{cases} \tag{40}$$

将式(40)代入式(39)可得

$$\begin{aligned}\dot{\boldsymbol{V}}(t)\leqslant&-k_{10}\|z_{1}\|_{\mathrm{F}}^{2}-k_{20}\|z_{2}\|_{\mathrm{F}}^{2}-\\&0.5\sum_{i=1}^{3}\alpha_{i}\|\tilde{\boldsymbol{\vartheta}}_{i}\|_{\mathrm{F}}^{2}-0.5\sum_{i=1}^{3}\beta_{i}\|\tilde{\boldsymbol{\zeta}}_{i}\|_{\mathrm{F}}^{2}+\tau\end{aligned} \tag{41}$$

式中:$\tau=\sum_{i=1}^{3}\left(\frac{\alpha_{i}}{2}\|\boldsymbol{\vartheta}_{i}^{*}\|^{2}+\frac{\beta_{i}}{2}\|\boldsymbol{\zeta}_{i}^{*}\|^{2}\right)$。

令 $\rho=\min\{2k_{10},2k_{20},\alpha_{i},\beta_{i}\}$ 则式(41)可表示为

$$\dot{\boldsymbol{V}}(t)\leqslant-\rho\boldsymbol{V}(t)+\tau \tag{42}$$

给出如下定理。

定理 1 对于解耦后飞翼无人机,在给定已知的输入参考信号 $\boldsymbol{x}_{1\mathrm{d}}$,采用控制律式(33)和 PSO 补偿器给出的自适应律式(38),对于任意给定的初始紧集:

$$\Omega_{0}=\{\boldsymbol{x}_{1\mathrm{d}},\boldsymbol{x}_{2\mathrm{d}},\hat{\boldsymbol{\vartheta}},\hat{\boldsymbol{\zeta}}\,|\,x_{1\mathrm{d}}(0),x_{2\mathrm{d}}(0),\hat{\vartheta}(0),\hat{\zeta}(0)\}$$

有如下结论:

1) 闭环系统状态 $\boldsymbol{x}(t)=[\boldsymbol{x}_{1}^{\mathrm{T}},\boldsymbol{x}_{2}^{\mathrm{T}}]^{\mathrm{T}}$ 和 PSO 自适应权值 $\hat{\boldsymbol{\vartheta}}$、$\hat{\boldsymbol{\zeta}}$ 有界,即:

$$\|\hat{\boldsymbol{\vartheta}}\|_{\mathrm{F}}\leqslant c_{\tilde{\vartheta}\max}+\|\boldsymbol{\vartheta}^{*}\|_{\mathrm{F}} \tag{43}$$

$$\hat{\boldsymbol{\zeta}}_{\mathrm{F}}\leqslant c_{\tilde{\zeta}\max}+\|\boldsymbol{\zeta}^{*}\|_{\mathrm{F}} \tag{44}$$

$$\|\boldsymbol{x}(t)\|\leqslant c_{x\max}+\max_{i\in[0,t]}\{\|\boldsymbol{x}(\tau)\|_{\mathrm{F}}\} \tag{45}$$

式中:

$$c_{\tilde{\vartheta}\max}=\sqrt{2\lambda_{1}(V(0)+\tau/\rho)}$$

$$c_{\tilde{\zeta}\max}=\sqrt{2\lambda_{2}(V(0)+\tau/\rho)}$$

$$c_{x\max}=\sqrt{2(V(0)+\tau/\rho)}$$

2) 系统的状态及 PSO 补偿器权值指数收敛于一个有界紧集 Ω_{s} 内:

$$\Omega_{s}=\{x,\tilde{\vartheta},\tilde{\zeta},\lim_{t\to\infty}\|\tilde{\boldsymbol{\vartheta}}\|_{\mathrm{F}}=\mu_{\vartheta}^{*},\lim_{t\to\infty}\|\tilde{\boldsymbol{\zeta}}\|_{\mathrm{F}}=\mu_{\zeta}^{*}\} \tag{46}$$

式中:$\mu_{\vartheta}^{*}=\sqrt{2\lambda_{1}\tau/\rho}$;$\mu_{\zeta}^{*}=\sqrt{2\lambda_{2}\tau/\rho}$。

对结论 1)证明如下:

对式(42)两边从[0,t]积分可得

$$0\leqslant|\boldsymbol{V}(t)|\leqslant(V(0)-\tau/\rho)\mathrm{e}^{-\rho t}+\tau/\rho \tag{47}$$

即

$$0\leqslant|\boldsymbol{V}(t)|\leqslant V(0)+\tau/\rho \tag{48}$$

根据式(48)和式(36)可得

$$\|\tilde{\boldsymbol{\vartheta}}\|_{\mathrm{F}}\leqslant\sqrt{2\lambda_{1}\boldsymbol{V}(t)}\leqslant c_{\tilde{\vartheta}\max}\|\tilde{\boldsymbol{\zeta}}\|_{\mathrm{F}}\leqslant\sqrt{2\lambda_{2}\boldsymbol{V}(t)}\leqslant c_{\tilde{\zeta}\max}$$

由于 $\tilde{\boldsymbol{\vartheta}}=\boldsymbol{\vartheta}^{*}-\hat{\boldsymbol{\vartheta}}$,$\tilde{\boldsymbol{\zeta}}=\boldsymbol{\zeta}^{*}-\hat{\boldsymbol{\zeta}}$,则:

$$\|\hat{\boldsymbol{\vartheta}}\|_{\mathrm{F}}-\|\boldsymbol{\vartheta}^{*}\|_{\mathrm{F}}\leqslant\|\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}^{*}\|_{\mathrm{F}}\leqslant c_{\tilde{\vartheta}\max} \tag{49}$$

$$\|\hat{\boldsymbol{\zeta}}\|_{\mathrm{F}}-\|\boldsymbol{\zeta}^{*}\|_{\mathrm{F}}\leqslant\|\hat{\boldsymbol{\zeta}}-\boldsymbol{\zeta}^{*}\|_{\mathrm{F}}\leqslant c_{\tilde{\zeta}\max} \tag{50}$$

由式(49)和式(50)可得式(43)、式(44)。

由于 z_{1}、z_{2}、$\boldsymbol{x}_{1\mathrm{d}}$ 有界,又因 $\boldsymbol{e}(t)=\boldsymbol{x}_{1}(t)-\boldsymbol{x}_{\mathrm{d}}(t)$,可得

北航学报 赠阅

$$\|\boldsymbol{x}(t)\| - \|\boldsymbol{x}_{\mathrm{d}}(t)\| \leqslant \|\boldsymbol{x}(t) - \boldsymbol{x}_{\mathrm{d}}(t)\| \leqslant c_{x\max} \tag{51}$$

即式(45)成立，结论 1)得证。

对结论 2)证明如下：

由式(49)可得

$$\|\boldsymbol{e}(t)\| \leqslant \sqrt{2(V(0) - \tau/\rho)\mathrm{e}^{-\rho t} + 2\tau/\rho} \tag{52}$$

$$\|\tilde{\boldsymbol{\vartheta}}\|_{\mathrm{F}} \leqslant \sqrt{2\lambda_1(V(0) - \tau/\rho)\mathrm{e}^{-\rho t} + 2\tau/\rho} \tag{53}$$

$$\|\tilde{\boldsymbol{\zeta}}\|_{\mathrm{F}} \leqslant \sqrt{2\lambda_2(V(0) - \tau/\rho)\mathrm{e}^{-\rho t} + 2\tau/\rho} \tag{54}$$

若 $V(0) = \tau/\rho$，则 $\forall t \geqslant 0$，$\|\boldsymbol{e}(t)\| \leqslant \mu_{\mathrm{e}}^{*}$。若 $V(0) \neq \tau/\rho$，则存在 T_{r}，使得当 $t > T_{\mathrm{r}}$ 时 $\|\boldsymbol{e}(t)\| \leqslant \mu_{\mathrm{e}}^{*}$。

令 $\mu_{\mathrm{e}} = \sqrt{2(V(0) - \tau/\rho)\mathrm{e}^{-\rho t} + 2\tau/\rho}$，取 $T_{\mathrm{r}} = -\dfrac{1}{\rho}\ln\left(\dfrac{\mu_{\mathrm{e}}^{2} - 2\tau/\rho}{2(V(0) - \tau/\rho)}\right)$，可得

$$\lim_{t\to\infty}\|\boldsymbol{e}(t)\| = \mu_{\mathrm{e}}^{*} \tag{55}$$

$$\lim_{t\to\infty}\|\tilde{\boldsymbol{\vartheta}}(t)\|_{\mathrm{F}} = \mu_{\tilde{\vartheta}}^{*} \tag{56}$$

$$\lim_{t\to\infty}\|\tilde{\boldsymbol{\zeta}}(t)\|_{\mathrm{F}} = \mu_{\tilde{\zeta}}^{*} \tag{57}$$

结论 2)得证。

由式(46)可知，通过调整 k_{10}、k_{20}、λ_1、α_i、λ_2、β_i 的大小，可以调整系统的收敛速度和收敛域的大小。

7　仿真验证

为了验证本文控制器的性能，本文选取耦合严重、抗侧滑要求高的蛇形机动进行仿真。蛇形机动的飞行过程为：在海拔高度 500 m，无人机以 100 m/s 的速度进入机动动作。在该机动动作中，无人机通过快速滚转实现速度方向顺时针转动 180°。然后，通过反向滚转使得速度方向从 180° 转到 0°。运动中 2 个转向过程连续衔接。

为了验证控制器的鲁棒性，在整个机动过程中选取如下气动系数发生不同程度的偏移，不确定性变化范围如表 1 所示，ΔC_{β}^{L}、ΔC_{β}^{N}、C_{p}^{L}、C_{r}^{N} 分别为侧滑角和角加速度引起的气动力矩系数，ΔL 为重心漂移。

蛇形机动仿真结果如图 6 所示，红色实线为无拉偏情况，黑色虚线是气动参数存在扰动情况。

从机动轨迹看，当无人机建立大角度滚转姿态时，飞行高度有所降低。带有气动拉偏情况下的轨迹线在第1个转弯处出现了很小的偏差。

表 1　气动参数偏移幅度

Table 1　Aerodynamic disturbance coefficients

参　数	ΔC_{β}^{L}/%	ΔC_{β}^{N}/%	C_{p}^{L}/%	C_{r}^{N}/%	ΔL/cm
偏移幅度	15	−10	20	20	1.5

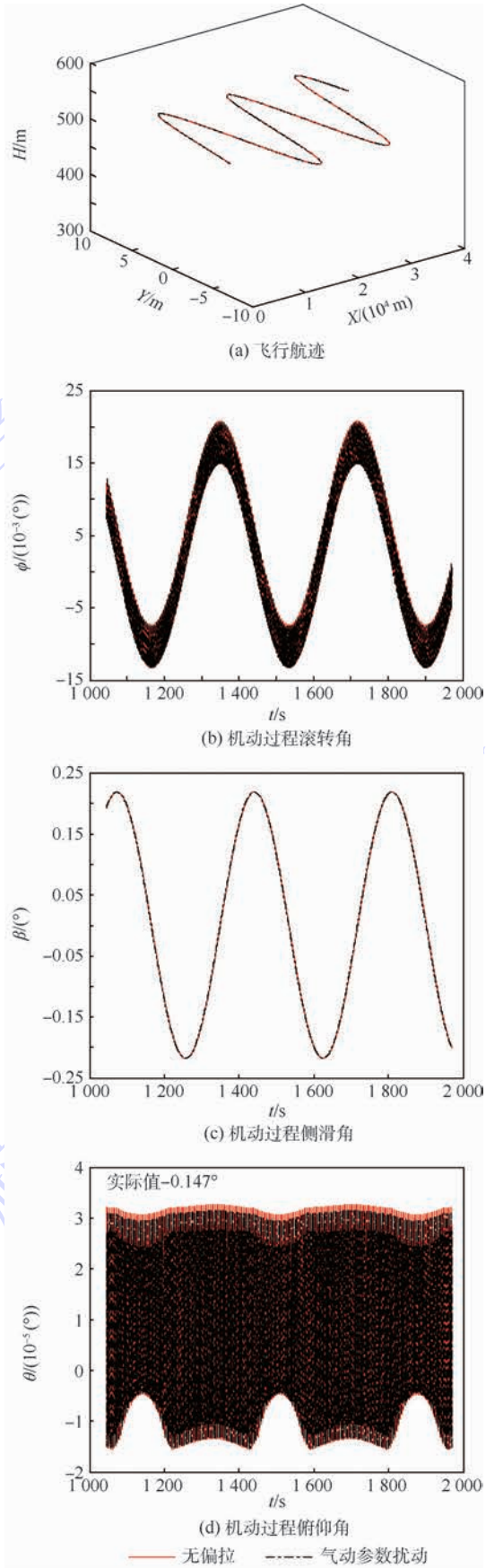


图 6　飞翼无人机蛇形机动仿真结果

Fig. 6　Simulation results of flying wing UAV snake maneuver

但是,飞翼无人机依然完成了机动转弯。从无人机的姿态角和侧滑角的动态图可知,控制器对扰动有很好地适应性。

从图 7 可知:在飞行过程中,飞翼无人机各气动控制舵面出舵量适中,波动量很小。但是,在飞行过程中,由于控制器时刻要对无人机进行气动补偿和解耦,使得各舵面处于快速振荡状态。

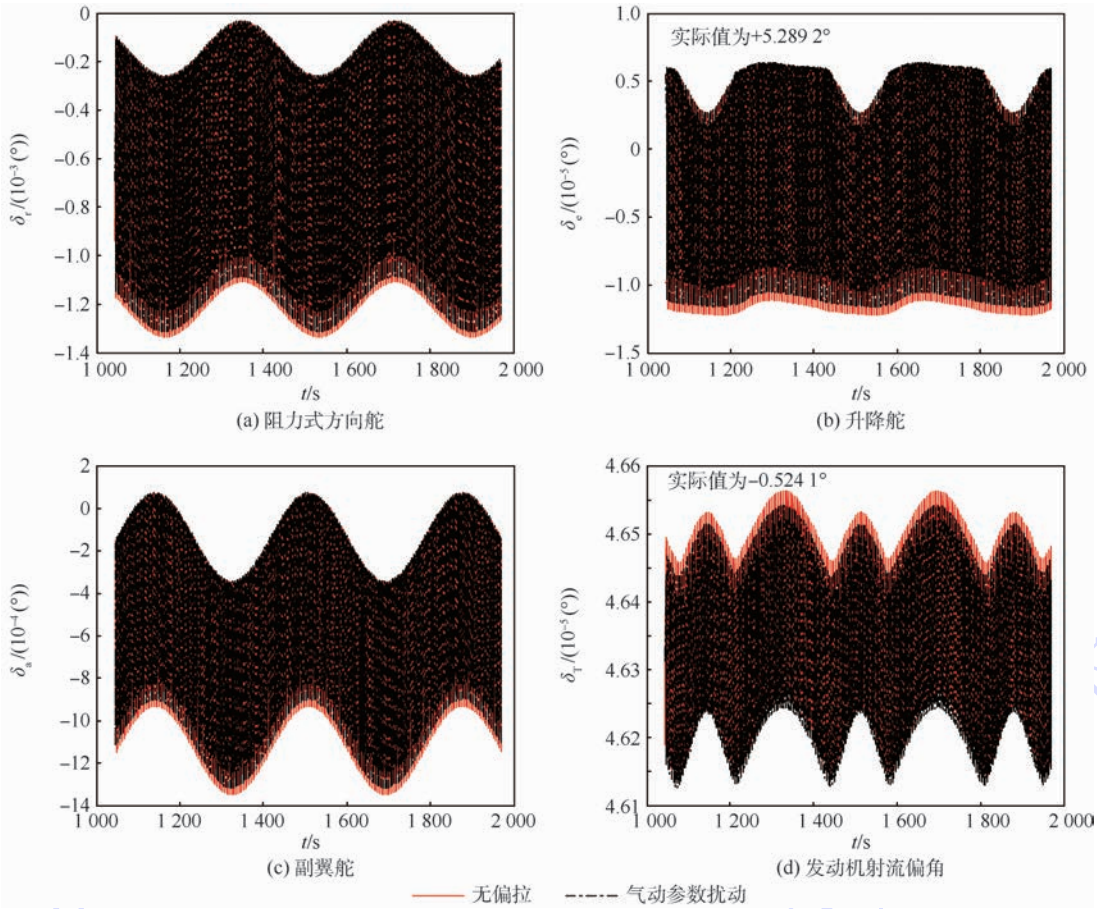


图 7 无人机蛇形机动出舵量

Fig. 7 Rudder angle of UAV snake maneuver

8 结 论

1) 针对飞翼布局无人俯仰操纵能力不足的特点,设计无人机采用流体矢量涡轮增压发动机提供动力和额外俯仰矢量力矩。

2) 针对飞翼无人机的机动飞行控制存在各种耦合和扰动的特点,本文设计了内环线性化解耦补偿器消除各种不利的耦合项,采用反步跟踪算法设计了外环航迹跟踪,并采用 PSO 补偿器补偿各种扰动和不可建模项的控制方案。

3) 本文证明了该控制结构的稳定性。同传统反步控制方法相比,本控制器增加了内环解耦结构,使得线性化后的系统为弱非线性系统。同时保留了对飞行稳定有利的保留气动阻尼项,便于工程实现。

4) 对蛇形机动的仿真结果表明,该设计方案是有效的。

参考文献 (References)

[1] LAN C E, LI J L, YAU W C, et al. Longitudinal and lateral-directional coupling effects on nonlinear unsteady aerodynamic modeling from flight data[C]// AIAA Atmospheric Flight Mechanics Conference and Exhibit. Reston: AIAA, 2013: 394-402.

[2] ALIKHAN M, PEYADA N K, GO T H. Flight dynamics and optimization of three-dimensional perching maneuver[J]. Journal of Guidance, Control, and Dynamics, 2013, 36(6): 1791-1797.

[3] GUO Y, YAO Y, WANG S, et al. Maneuver control strategies to maximize prediction errors in ballistic middle phase[J]. Journal of Guidance, Control, and Dynamics, 2013, 36(4): 1225-1234.

[4] ZHI Q, CAI Y L. Energy-management steering maneuver for thrust vector-controlled interceptors[J]. Journal of Guidance, Control, and Dynamics, 2012, 35(6): 1798-1804.

[5] MUELLER J B, GRIESEMER P R, THOMAS S J. Avoidance maneuver planning incorporating station-keeping constraints and automatic relaxation[J]. Journal of Aerospace Computing Infor-

mation & Communication, 2013, 10(6): 306-322.

[6] NEELY A J, GESTO F N, Young J. Performance studies of shock vector control fluidic thrust vectoring[C]// Proceedings of 43rd AIAA/ASME/SAE/ASEE Joint Propulsion Conference. Reston: AIAA, 2007: 1-14.

[7] KAREN A D. Summary of fluidic thrust vectoring research conducted at NASA langley research center: AIAA-2003-3800 [R]. Reston: AIAA, 2003.

[8] SADIQ M U. Performance analysis and flowfield characterization of secondary injection thrust vector control (SITVC) for a 2DCD nozzle[D]. Los Angeles: University of Southern California, 2007: 85-108.

[9] 王猛杰，额日其太，王强. 激波矢量控制喷管落压比影响矢量性能及分离区控制数值模拟[J]. 航空动力学学报, 2015, 30(3): 527-538.
WANG M J, ERIQITAI, WANG Q. Numerical simulaton of nozzle pressure ratio effect on vector performance and separation control for shock vector control nozzle[J]. Journal of Aerospace Power, 2015, 30(3): 527-538(in Chinese).

[10] KIRANYAZ S, INCE T, GABBOUJ M. Dynamic data clustering using stochastic approximation driven multi-dimensional particle swarm optimization[J]. Lecture Notes in Computer Science, 2010, 22(10): 1448-1462.

[11] YANG Y, CHEN X, LI C. Transient performance improvement in model reference adaptive control using H_∞ optimal method [J]. Journal of the Franklin Institute, 2015, 352(1): 16-32.

[12] 杨艺，陈欣，李春涛. 一种可保证瞬态特性的改进鲁棒模型参考自适应控制[J]. 控制与决策, 2015, 30(8): 1379-1385.
YANG Y, CHEN X, LI C T. A modified robust model reference adaptive controller with guaranteed transient performance[J]. Control and Decision, 2015, 30(8): 1379-1385(in Chinese).

[13] 朱纪洪，张尚敏，周池军，等. 飞机超机动状态动力学特征及对控制系统的挑战[J]. 控制理论与应用, 2014, 31(12): 1650-1662.
ZHU J H, ZHANG S M, ZHOU C J, et al. Dynamic characteristics and challenges for control system of super-maneuverable aircraft[J]. Control Theory & Applications, 2014, 31(12): 1650-1662(in Chinese).

[14] WILSON J R. UAV worldwide roundup 2007[J]. Aerospace America, 2007, 45(5): 30-37.

[15] OSTERHUBER R. FCS requirements for combat aircraft-lessons learned for future designs[C]// Workshop on Stability & Control, 2011: STO-AVT-189.

[16] LI J, CHEN X, LI Z. The attitude decoupling control of the flying wing UAV[C]// Proceedings of IEEE Guidance, Navigation and Control Conference. Piscataway, NJ: IEEE Press, 2017: 357-362.

[17] BANKS A, VINCENT J, ANYAKOHA C. A review of particle swarm optimization. Part Ⅱ: Hybridisation, combinatorial, multicriteria and constrained optimization, and indicative applications[J]. Natural Computing, 2008, 7(1): 109-124.

[18] 范成礼，邢清华，范海雄，等. 带审敛因子的变邻域粒子群算法[J]. 控制与决策, 2014, 29(4): 696-701.
FAN C L, XING Q H, FAN H X, et al. Particle swarm optimization and variable neighborhood search algorithm with convergence criterions[J]. Control and Decision. 2014, 29(4): 696-701(in Chinese).

[19] 贾树晋，杜斌，岳恒. 基于局部搜索与混合多样性策略的多目标粒子群算法[J]. 控制与决策, 2012, 27(6): 813-819.
JIA S J, DU B, YUE H. Local search and hybrid diversity strategy based multi-objective particle swarm optimization algorithm [J]. Control and Decision, 2012, 27(6): 813-819(in Chinese).

[20] 李擎，张超，陈鹏，等. 一种基于粒子群参数优化的改进蚁群算法[J]. 控制与决策, 2013, 28(6): 873-879.
LI Q, ZHANG C, CHEN P, et al. Improved ant colony optimization algorithm based on particle swarm optimization[J]. Control and Decision, 2013, 28(6): 873-879(in Chinese).

[21] POLI R, KENNEDY J, BLACKWELL T. Particle swarm optimization: An overview[J]. Swarm Intelligence, 2007, 1: 33-57.

[22] SOEST W R V, CHU Q P, MULDER J A. Combined feedback linearization and constrained model predictive control for entry flight[J]. Journal of Guidance, Control, and Dynamics, 2006, 29(2): 427-434.

[23] SONNEVELDT L, CHU Q P, MULDER J A. Nonlinear flight control design using constrained adaptive backstepping[J]. Journal of Guidance, Control, and Dynamics, 2007, 30(2): 322-336.

[24] LEE T, KIM Y. Nonlinear adaptive flight control using backstepping and neural networks controller[J]. Journal of Guidance, Control, and Dynamics, 2001, 24(4): 675-682.

[25] SIEBERLING S, CHU Q P, MULDER J A. Robust flight control using incremental nonlinear dynamic inversion and angular acceleration prediction[J]. Journal of Guidance, Control, and Dynamics, 2010, 33(6): 1732-1742.

[26] MACKUNIS W, PATRE P M, KAISER M K, et al. Asymptotic tracking for aircraft via robust and adaptive dynamic inversion methods[J]. IEEE Transactions on Control Systems Technology, 2010, 18(6): 1448-1456.

[27] JOHNSON E N, TURBE M A. Modeling, control, and flight testing of a small-ducted fan aircraft[J]. Journal of Guidance, Control, and Dynamics, 2006, 29(4): 769-779.

[28] XU B, HUANG X, WANG D, et al. Dynamic surface control of constrained hypersonic flight models with parameter estimation and actuator compensation[J]. Asian Journal of Control, 2014, 16(1): 162-174.

[29] 李继广，陈欣，王鑫，等. 飞翼无人机机动飞行非线性鲁棒自适应控制[J]. 系统工程与电子技术, 2017, 39(9): 2058-2067.
LI J G, CHEN X, WANG X, et al. Nonlinear robust adaptive control of flying wing UAV maneuvering flight[J]. Systems Engineering and Electronics, 2017, 39(9): 2058-2067(in Chinese).

作者简介：

李继广　男，博士研究生。主要研究方向：非线性鲁棒控制、无人机控制系统设计开发。

陈欣　男，博士，研究员，博士生导师。主要研究方向：无人机控制系统、三余度飞控计算机。

李亚娟　女，博士，讲师。主要研究方向：激光雷达大气探测。

北航学报 赠阅

Nonlinear robust control method for maneuver flight of flying wing UAV

LI Jiguang[1], CHEN Xin[1,*], LI Yajuan[2], ZHANG Rong[1]

(1. School of Automation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;
2. School of Electronic Engineering, Nanjing Xiaozhuang University, Nanjing 211171, China)

Abstract: As flying wing UAV lacks manipulating ability, a control strategy combined with fluidic thrust vectoring-turbocharged engine (FTV-E) technology is proposed. In this paper, the control scheme is designed: the inner loop compensator is used to eliminate the negative coupling term of system; the outer loop compensator used backstepping tracking algorithm; the particle swarm optimization (PSO) compensator to compensate the disturbance and coupling term that cannot be modeled. The control structure's stability is proved. Based on the traditional backstepping control methods, the proposed controller increases the inner loop compensator. The proposed inner loop compensator retains the aerodynamic damping term which is favorable to flight. This compensator not only can reduce the conservatism of the outer loop controller, but also is convenient for engineering realization. The simulation results show that the proposed control scheme is effective.

Keywords: FTV-E technology; maneuver flight; control structure; input linearization; backstepping control; PSO compensator

北京航空航天大学学报

Received: 2017-01-12; **Accepted**: 2017-04-13; **Published online**: 2017-05-31 10:28

URL: kns.cnki.net/kcms/detail/11.2625.V.20170531.1028.002.html

Foundation items: Aeronautical Science Foundation of China (20160152001); the Fundamental Research Funds for the Central Universities (NS2015038)

* **Corresponding author**. E-mail: chenxin@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0016

液氮贮箱常压停放实验与数值仿真

李佳超，梁国柱*
（北京航空航天大学 宇航学院，北京 100083）

摘 要：为研究低温推进剂的常压停放过程，设计了可视化液氮贮箱实验系统。实验中研究充填率和环境温度对液氮汽化量的影响，并测量贮箱内流体和贮箱外壁面的温度随时间和位置的变化。实验得出贮箱常压停放过程，相变主要在壁面和气液界面产生，并且气枕区存在温度分层，距出口位置越近温度越高；而液体区温度基本一致，处于饱和状态。贮箱外壁面在轴向的温度分布显著不同，处于液体区壁面温度低。运用分子动力学推导出的 Hertz-Knudsen 公式作为气液相变的传热传质源项，并据实验测得温度边界条件，采用混合物模型对贮箱常压停放状态进行 30 min 的数值仿真。仿真得到结果显示体积汽化速率与实验数据的偏差在 5% 以内，液体区的温度仿真与实验的偏差在 0.15 K 左右。

关 键 词：低温贮箱；常压停放；相变；传热传质；混合物模型
中图分类号：V434
文献标识码：A **文章编号**：1001-5965(2018)01-0099-09

新一代大推力液体火箭采用液氢、液氧等低温流体作为芯级推进剂，推进剂在加注完后至发射前一直处于地面停放过程，贮箱壁面漏热使低温推进剂的温度升高并发生汽化，造成推进剂的损失，并影响发射安全。

国内外对低温推进剂的地面停放过程开展过一系列的研究。在实验方面：Lewis 研究中心[1]将 4.89 m^3 的多层绝热液氢贮箱放在真空度为 10^{-4} Pa 环境中，研究低热流密度下液氢贮箱内压力变化和温度分布。Barsi 等[2]用氟化醚(HFE-7000)作为工质研究自增压过程中贮箱内压力随热流密度变化。Seo 和 Jeong[3]的自增压实验用液氮为工质测量热流密度、充填率与压力上升值的关系。Das 等[4]用水作为工质测量 500 W/m^2 热流密度条件下长方体贮箱内气液温度分布。国内 Wang 等[5]研究液氮贮箱有排放和无排放的充填过程，并测量液氮温度分布和气枕区压力变化。聂中山等[6]利用液氢杜瓦瓶盛装液氮工质，进行液氮蒸发实验，测量了压力、充填率与蒸发率相互间的关系。王贵仁[7]采用液氮测量在不同初始充填率下杜瓦瓶内压力随时间变化。乔国发[8]采用液化天然气(LNG)为工质对密闭贮罐内压力、日蒸发气体量及蒸发率进行研究。以上实验多注重于贮箱地面停放的自增压过程，缺乏贮箱地面常压停放过程的实验数据及规律研究。数值计算方面主要有集总参数法和分布参数法，最先使用的是集总参数法，如代予东和赵红轩[9]运用数学方法来模拟箱内的增压、传热和传质过程，仅考虑气枕区向液体区传热引起的传质。Estey 等[10]将气液两相各分成一个分区，并沿界面建立一个分区作为传热传质区。Panzarella 和 Kassemi[11]采用有限元软件 FIDAP 建立一个球型贮箱模型，对气枕区采用集总参数法建立能量和质量模型，并与液体不可压的 Navier-Stokes 方程和

收稿日期：2017-01-12；**录用日期**：2017-05-19；**网络出版时间**：2017-06-19 15:16
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170619.1516.002.html
基金项目：中国运载火箭技术研究院高校联合创新基金资助项目（CALT201302）
* **通信作者**. E-mail：lgz@buaa.edu.cn

引用格式：*李佳超，梁国柱. 液氮贮箱常压停放实验与数值仿真[J]. 北京航空航天大学学报，2018，44(1)：99-107.*
LI J C，LIANG G Z. Experiment and numerical simulation of liquid nitrogen tank atmospheric ground parking [J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(1)：99-107 (in Chinese).

北航学报 赠阅

能量方程耦合求解。从上可以看出集总参数法不能够给出温度、速度等参数的具体分布,也影响到气液界面上的能量和热量交换的计算,本文研究的贮箱内流体传热传质过程剧烈,并且有气体的持续排放,不适合采用集总参数法。分布参数法主要是使用计算流体力学(CFD)方法对贮箱内流体物理过程求解,如刘展等[12]针对低温推进剂箱体射前停放阶段,研究了某低温液氧贮箱在地面停放阶段所经历的开口放置及高温氧气预增压过程。Chen 和 Liang[13]建立了包含液体推进剂和混合气体两相的流体体积法(VOF)计算流体力学模型,并引入了基于热力学平衡假设的推进剂相变模型。

本文的主要目的是研究贮箱地面常压停放过程。通过搭建可视化液氮贮箱实验系统,研究充填率和环境温度对液氮的汽化率、流体温度分布及贮箱外壁面温度的影响规律。从而进一步基于实验现象和实验确定的初边值条件,选择合理的 CFD 方法对液氮贮箱常压停放过程进行数值仿真,并深入分析整个流场的细节与机理,认识实验过程中无法测量的参数变化规律,为低温推进剂的常压停放过程提供合理的预测模型。

1 实验系统

依据实验的目的和性能要求,实验中建立了可视化液氮贮箱常压停放系统,实验系统由液氮供给系统、实验贮箱、测控系统和图像采集系统组成,如图 1 所示。

实验贮箱的材料是高硼硅玻璃,贮箱体积 21.03 L,圆柱段内径 290 mm,外径 300 mm,出口段内径 100 mm,贮箱的总高度 435 mm,平均厚度 5 mm。贮箱内的液位和体积关系通过 4 ℃纯净水标定,如表 1 所示。

实验通过 100 L 杜瓦瓶对贮箱加注液氮,加注过程需要对贮箱进行充分的预冷。多点温度传感器共 12 个测点,用于测量贮箱内流体的温度。壁面温度传感器共 8 个测点,用于测量贮箱固体外壁面的温度。实验主要对低温液体在贮箱内常压停放的物理过程进行研究。实验中记录流体的温度、固体外壁面的温度和贮箱内液体的液位,并通过图像采集设备观察贮箱内液体的运动。实验现场如图 2、图 3 所示。

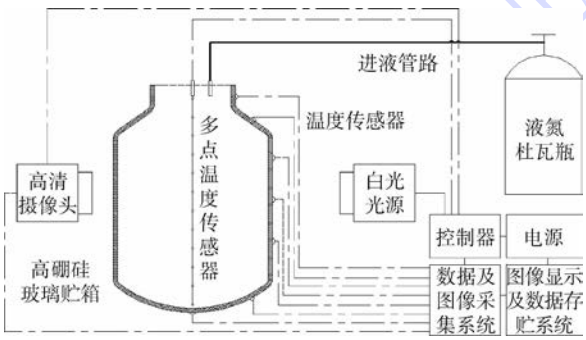


图 1 液氮贮箱常压停放实验系统原理图

Fig. 1 Schematic diagram of experimental system of liquid nitrogen tank atmospheric ground parking

表 1 贮箱液位与体积变化关系

Table 1 Variation of volume of tank with liquid level

液位/cm	体积/L	充填率/%	总体积/L
0	1.607	7.6	21.03
5	4.907	23.3	
10	8.207	39.0	
15	11.507	54.7	
20	14.807	70.4	
25	18.107	86.1	



图 2 液氮贮箱常压停放实验现场图

Fig. 2 Diagram of liquid nitrogen tank atmospheric ground parking on experimental site



图 3 液氮贮箱常压停放实验现场测控图

Fig. 3 Measurement and control diagram of liquid nitrogen tank atmospheric ground parking on experimental site

2 实验结果与分析

2.1 液氮汽化速率与环境温度和充填率关系

实验对液氮体积充填率为 39.0%、54.7%、70.4% 和 86.1%(即液位分别为 10、15、20 和 25 cm)的汽化情况进行研究。图 4 主要给出体积充填率对液氮汽化的影响,V 为液氮体积,t 为时间。在环境温度 275 K 时,贮箱内液氮的液位刻

北航学报 赠阅

度每下降 5 cm，记录间隔时间，环境温度 285 K 时每隔 15 min 记录贮箱内液氮的液位刻度。从图 4 可以看出汽化相同液氮体积的时间间隔随时间的推移而增加；液氮初始体积充填率越大，液体的汽化速率越大，而随着时间的增加，不同充填率下液体汽化速率差值逐渐减小。这是因为充填率越大，液氮与壁面接触的面积越大，接受的热量越多，导致汽化速率越大，而相同液位刻度下初始段的汽化速率不同主要是液体充填过程中液氮初始温度高于饱和温度，充填率越大导致过热的液体越多，初始段液氮体积下降的越快。图 5 给出环境温度对液氮汽化的影响。在环境温度为 275 K、285 K 和 291 K 的 3 种情况下，对液氮体积充填率为 86.1% 进行实验，实验中贮箱内液氮的液位刻度每下降 5 cm 记录时间数据。从图 5 可以看出，汽化相同体积的液氮，环境温度 291 K 所需时间最少，环境温度 275 K 所需时间最多。这是因为充填率相同时，环境温度越高，贮箱内液氮通过贮箱壁面与外界交换的热量越多，而常压停放贮箱内的液氮饱和温度基本保持不变，液体与外界交换热量的越多，汽化的体积就越大。

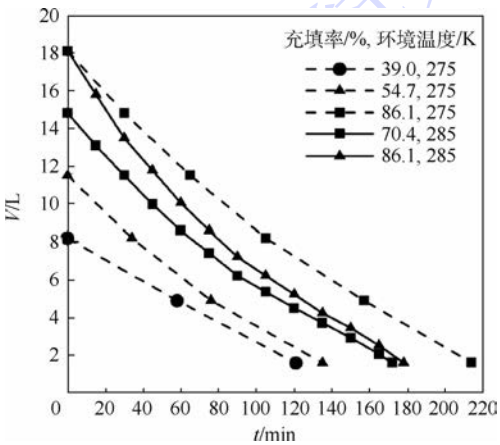


图 4　不同充填率和温度时，液氮体积随时间的变化

Fig. 4　Variation of liquid nitrogen volume with time under different filling rates and temperatures

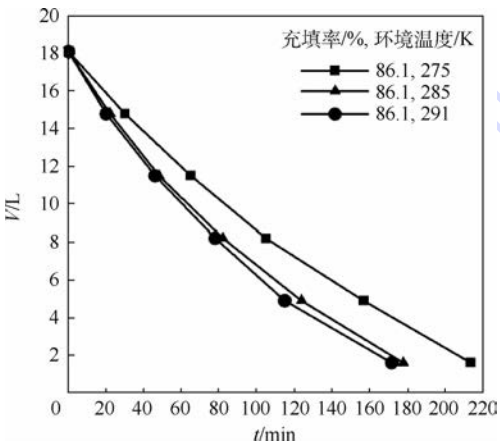


图 5　不同环境温度时，液氮体积随时间的变化

Fig. 5　Variation of liquid nitrogen volume with time under different ambient temperatures

2.2　贮箱内流体温度与时间关系

流体温度测量采用 12 个封装在不锈钢圆管内温度传感器，其中每个测点间隔距离 37 mm。图中每隔 30 min 取测量的温度数据绘制温度与时间、位置曲线图，具体如图 6 所示，Y 为轴向高度，$Y=0$ 为贮箱底部最低点，T 为温度，下标 f 表示流体，a 表示环境，虚线表示不同时刻气液界面位置。从图中可以看出气枕区的温度在轴向存在明显的温度分层，温度随时间增加而增大。液体区在远离气液界面处轴向方向上温度变化不大，且随时间的增加没有明显变化。气液界面附近液体区温度出现的偏差在于沿不锈钢圆管气枕区导热所致。液氮常压停放液体在贮箱内壁面和气液界面处汽化，汽化产生的气泡在浮力的作用

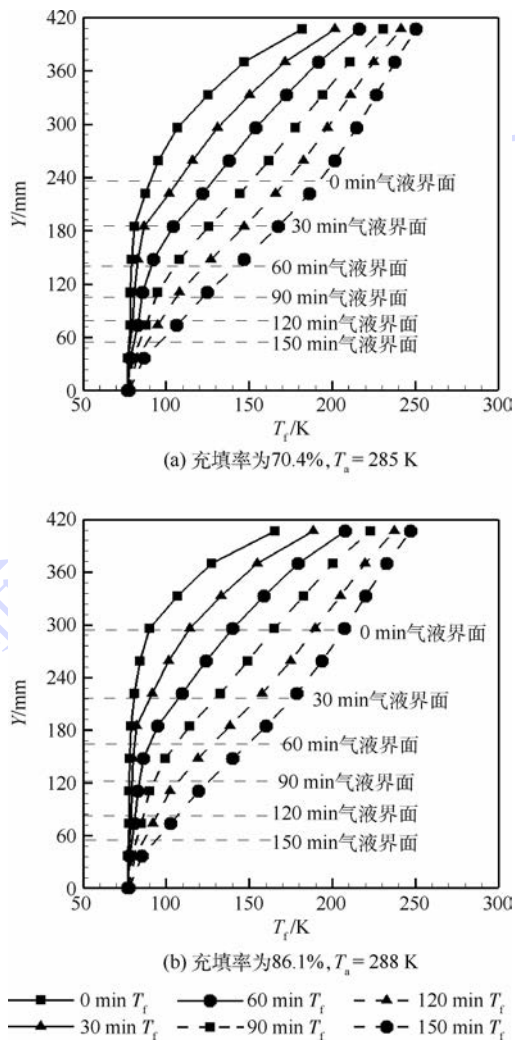


图 6　充填率为 70.4% 和 86.1% 时，流体温度与位置的关系

Fig. 6　Relationship between fluid temperature and location under 70.4% and 86.1% filling rate

北航学报 赠阅

下向气液界面处运动,贮箱内部液体存在剧烈运动和混合,导致液体温度趋于一致。而气枕区温度变化大是因为气体的比热和密度小,受热温度上升快。而轴向的温度分层是因为液体的温度低,生成的气体温度低,气液界面处气体向液体传热。

2.3 贮箱外壁面温度与时间的关系

实验为得到贮箱常压停放实验准确的边界条件,采用贴壁温度传感器测量贮箱外壁面的温度,每隔 30 min 取测量的温度数据绘制温度与时间、位置曲线图,具体如图 7 所示,下标 w 表示固体壁面。从时间上可以看出,贮箱外壁面的温度随时间增加而增大,气枕区外壁面温度在随时间的增速显著大于液体区外壁面温度增速。从位置上可以看出,气枕区的外壁面温度在轴向方向上从出口到液面温度呈下降趋势,而液体外壁面温度在轴向方向上并没有明显的分布趋势,且液体区温度测点的变化值小于气枕区温度测点的变化值。这是因为气枕区内气体的密度小,在比热容差距不大的情况下,气体的温度上升速率快,使得外界传递的热量大部分用于固体区域的升温,导致气枕区贮箱外壁面温度高。而液体温度低,密度和比热容都比气体大,液体与贮箱壁面间的对流换热作用强,使得与贮箱接触的固体区域温度低。高充填率气液界面附近温度偏低的是因为气液界面附近液体对流运动剧烈,并且有液体的汽化,使得液体与固体区热量交换值大,外界用于固体区温度上升的热量变小。贮箱底部测点温度偏低是因为实验时贮箱底部放置在实验台上,壁面与外界的对流换热作用受到限制。

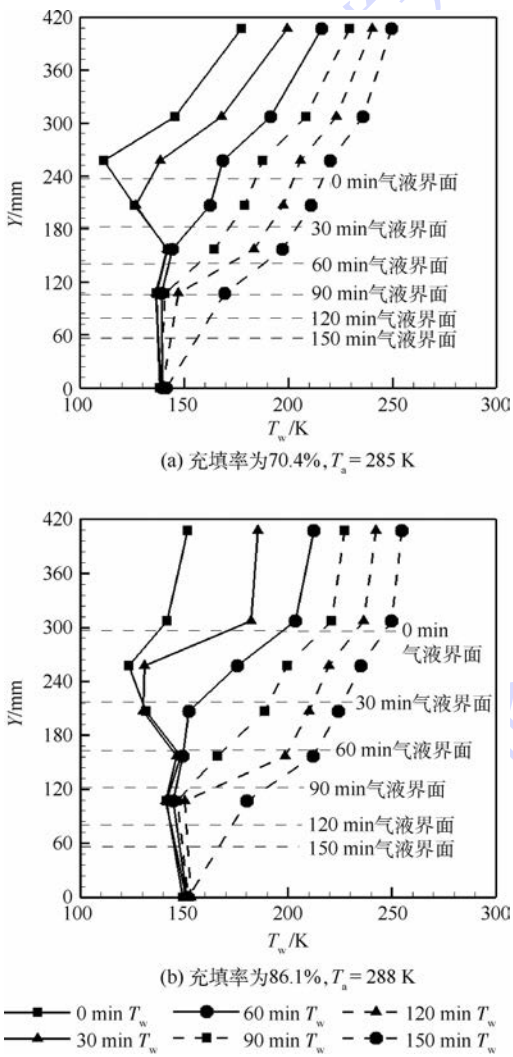


图 7 充填率为 70.4% 和 86.1% 时,壁面温度与位置关系

Fig. 7 Relationship between wall temperature and location under 70.4% and 86.1% filling rate

3 数值计算与分析

3.1 流体控制方程

在贮箱内液氮的常压停放实验过程中,观测到液体(特别是液面)剧烈沸腾和翻滚,液体中有大量的气泡从贮箱内壁面产生,并在浮力的作用下向气液界面处运动,运动过程中气泡逐渐变大,并在液面处破裂。目前低温液体常压停放的数值模拟多采用文献[14-15]所用的 VOF 模型建模,而事实上常压停放中贮箱内部这种复杂的气液两相传热传质流动过程并不适合采用 VOF 模型建模(因为实验观察到只有在贮箱的自增压过程中,气液界面才是准静止的水平面),而更适合采用混合物模型。VOF 模型中缺乏对气液两相间相互作用的计算,并且为获得较为准确气液界面位置付出了过高的计算代价,对计算的收敛性也有不利影响。混合物模型虽然缺乏对准确的气液界面的跟踪和处理,但对于地面常压停放阶段来说气液界面本身就是剧烈变化的,其受到液体的翻滚运动和气泡上升运动的影响。混合物模型不仅可以解决气液混合流动的问题,且能够得到贮箱中气枕和液体推进剂的气液两相分布,能够满足研究需要而且可以减少计算量,因此本文两相流模拟采用混合物模型。而混合物模型中的连续性方程、动量方程和能量方程见文献[16]。连续性方程和能量方程考虑液氮和氮气间传热传质过程中引起的质量源项 S_q 和能量源项 S_h 变化,S_q 为第 q 相质量源项,动量方程中考虑气液两相间的滑移速度、拽力及表面张力的影响。

3.2 相变模型

数值仿真的相变模型主要有相变系数模型、假定界面温度模型、能量平衡模型和温度恢复模型。后 3 种模型计算时不稳定、并且需要初始化

北航学报 赠阅

气液界面[17]。因此,文中采用相变系数模型。Knudsen 和 Partington[18]根据气体动力学理论推导的气液界面质量交换为

$$J=\alpha_i\frac{\sqrt{M}}{\sqrt{2\pi R}}\left(\frac{P}{\sqrt{T_v}}-\frac{P_{sat}(T_l)}{\sqrt{T_l}}\right) \tag{1}$$

式中:P 和 P_{sat}分别为压力和饱和压力。

运用克劳修斯-克拉珀龙方程,并转化为体积源项如下:

$$J=\frac{6\alpha}{D_{Sm}}\cdot\frac{\sqrt{M}}{\sqrt{2\pi RT_{sat}}}\cdot\frac{\Delta H}{1/\rho_v-1/\rho_l}\cdot\frac{T_{sat}-T}{T_{sat}} \tag{2}$$

对式(2)进行简化得到相变系数模型

$$J_{lv}=\beta_l\rho_l\alpha_l\frac{T_{sat}-T_l}{T_{sat}} \tag{3}$$

$$J_{vl}=\beta_v\rho_v\alpha_v\frac{T_{sat}-T_v}{T_{sat}} \tag{4}$$

式中:T_l 和 T_v 分别为液体与气体温度,K;D_{Sm}为气泡的索太尔平均直径,m;M 为气体摩尔质量,kg/mol;R 为摩尔气体常数,J/(mol·K);β 为控制相变强弱的时间松弛系数,s^{-1};a_i 为蒸发/冷凝系数;α、a_l 和 a_v分别为体积分数、液体体积分数和气体体积分数;J_{lv}为液体汽化质量,kg/(m^3·s);J_{vl}气体冷凝质量,kg/(m^3·s);ΔH 为气液的相变潜热,J/kg;ρ_v、ρ_l分别为气体和液体的密度,kg/m^3;T_{sat}为饱和温度,K。

$$\beta_l=\frac{6}{D_{Sm}}\cdot\frac{\sqrt{M}}{\sqrt{2\pi RT_{sat}}}\cdot\frac{\rho_v\Delta H}{\rho_l-\rho_v} \tag{5}$$

$$\beta_v=\frac{6}{D_{Sm}}\cdot\frac{\sqrt{M}}{\sqrt{2\pi RT_{sat}}}\cdot\frac{\rho_l\Delta H}{\rho_l-\rho_v} \tag{6}$$

传质过程引起的能量交换为式(7)、式(8),Q_{lv}为汽化热交换量,J/(m^3·s);Q_{vl}为冷凝热交换量,J/(m^3·s)。

$$Q_{lv}=J_{lv}\Delta H \tag{7}$$

$$Q_{vl}=J_{vl}\Delta H \tag{8}$$

相变系数模型中关键是确定时间松弛系数,而过大的时间松弛系数使计算难以收敛,过小的时间松弛系数会使气液界面温度与饱和温度存在较大偏差[19-20]。Lee[21]采用时间松弛系数 0.1 取得很好计算结果。本文通过实验对比与试算,时间松弛系数取 0.1 s^{-1}。仿真过程中连续性方程中 S_q 通过式(3)、式(4)控制,能量方程中 S_h 通过式(7)、式(8)控制。

3.3　固体壁导热控制方程

贮箱固体壁面与内外流体间存在热交换作用。固体壁面的导热方程如下:

$$\frac{\partial}{\partial t}(\rho_w h_w)=\nabla\cdot(k_w\nabla T_w) \tag{9}$$

式中:ρ_w、h_w、k_w 和 T_w 分别为高硼硅玻璃贮箱壁面的密度、显焓、导热系数和温度,ρ_w = 2 230 kg/m^3,k_w 值随温度的变化较大,其值对固体壁热量传递有十分重要的影响。温度为 150 K 和 80 K 时导热系数值分别为 0.76 和 0.48,导热系数具体值通过文献[22]插值计算得出。

3.4　数值方法与初边值条件

贮箱的仿真模型通过 Fluent 建立,采用二维轴对称结构,分别对气枕区、液氮区和固体区划分结构化网格。网格在靠近贮箱内壁面、气液界面和压力出口处的区域进行适当加密,通过网格验证试算,最终采用了网格数为 25 550 的二维轴对称网格,其中气枕区网格数 10 109,液体区网格数 12 006,固体区网格数 3 435,最小网格边长 0.1 mm,具体如图 8 所示,X 和 Y 分别为径向方向和轴向方向。

考虑到贮箱液面高度不大,忽略重力对饱和温度的影响,而且实验发现液氮处于剧烈沸腾状态,因此近似认为贮箱常压停放过程的初始状态为气液两相饱和的准静止状态,贮箱内部流体(气体和液体两部分)的温度均为对应气体压力(101 325 Pa)下的饱和温度(77.35 K);贮箱外的环境温度为 288 K,液氮的体积充填率 86.1%。仿真出口条件采用恒定压力出口,其值为 101 325 Pa。壁面采用温度边界条件,通过 UDF(User-Define Function)将实验测量的贮箱固体外壁面温度以一阶插值方式分段加入,具体值如图 9所示。仿真中通过采用 4 核 i7 计算 300 h 对液氮常压停放进行 1 800 s 数值仿真。实验中观测到贮箱内有大量气泡的上升运动,所以仿真中考虑了浮力和拽力,计算时氮气和液氮间的拽力函数选择 Schiller-Naumann 模型,滑移速度选择 Maninnen-et-al 模型,浮升力通过采用 Boussinesq 模型[23]。湍流模型采用适合低雷诺数 RNG k-ε,壁面采用增强壁面函数(enhanced wall)。相变模型通过 UDF 以质量源项和能量源项加载到连续性方程和能量方程。贮箱常压停放是一个瞬态过程,仿真中选择基于压力的求解器,压力插值方式选择 body-force-weight,速度压力耦合方式选择 PISO,体积分数插值方式选择 Quick 格式,其余的选择二阶迎风格式。氮气采用理想气体模型,液氮和氮气的物性参数考虑随温度的变化,通过查找文献[24]采用一阶插值计算获得,并将获得的参数通过 UDF 加载到 Fluent 模型中。计算中连续性、动量、能量方程中残差分别为 10^{-3}、10^{-4}、10^{-7}。对模型多次计算时发现,时间步长在

北航学报 赠阅

0.01 ~0.02 s 时,计算中能够保持严格的收敛性,且计算结果基本一致,实际计算时间步长取0.02 s。

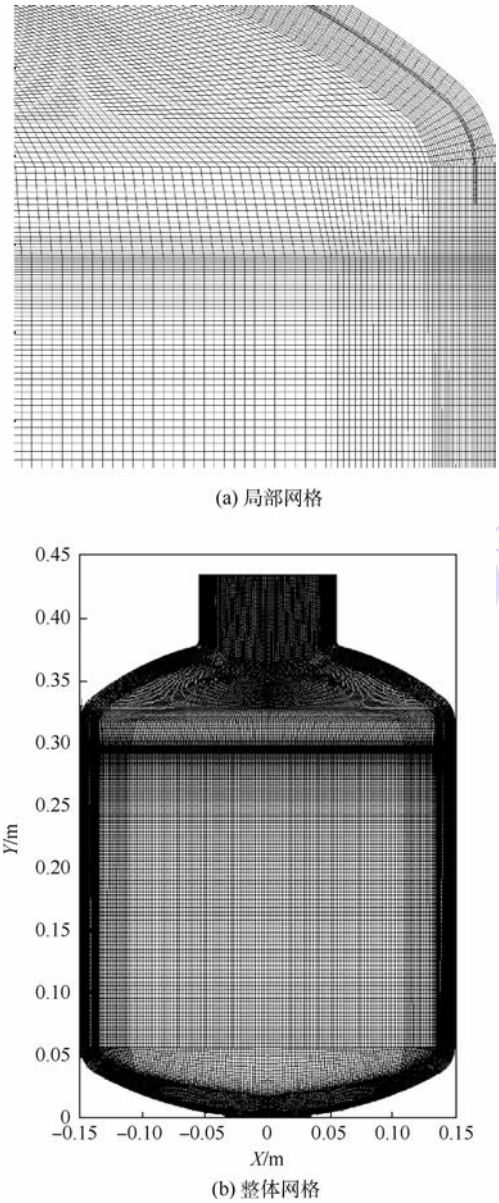


图 8 常压停放贮箱网格图

Fig. 8 Grids of tank during atmospheric ground parking

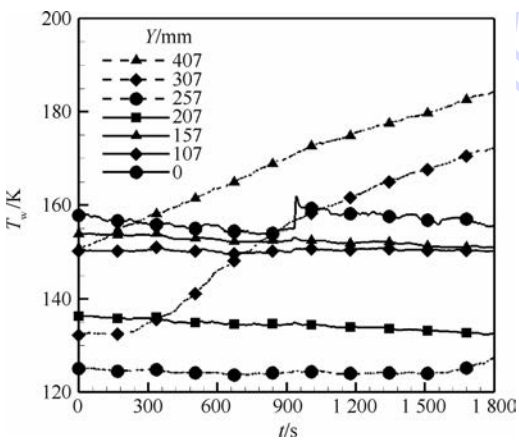


图 9 贮箱外壁面温度随时间变化

Fig. 9 Variation of tank's outer wall temperature with time

3.5 仿真与实验结果比较

图 10 给出实验和仿真得出的液氮体积随时间变化的比较。可以看出,5 min 时仿真的液氮体积比实验中液氮体积多,这是实验过程中加注的液氮温度高于饱和温度,导致初始测量阶段液氮的体积变化快,5 min 后在同一时刻仿真过程的液氮体积比实验过程中的液氮体积小。实验过程液氮体积的平均汽化速率为 0.002 67 L/s,仿真过程液氮体积的平均汽化速率为 0.002 78 L/s。图 11 给出实验和仿真得出液氮温度随时间变化的比较。实验中由于温度传感器存在着轴向的热传导,为了尽量减少液体温度测量的偏差,选用液体底部温度测量值与仿真进行比较。可以看出实验中液氮温度的测量值在 77.4 ~77.6 K 之间,而仿真过程监测点液氮的温度随时间在 77.35 K 左右变化。从图 12 看出贮箱内液氮温度基本一致,气液界面附近的液体处于饱和温度,其余液体主

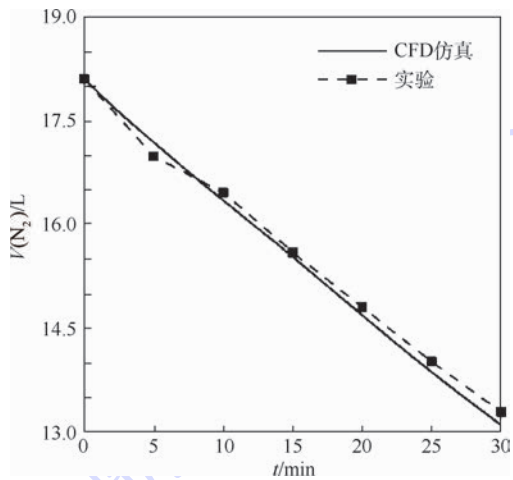


图 10 实验和仿真得出的液氮体积的比较

Fig. 10 Comparison of liquid nitrogen volume between experiment and simulation

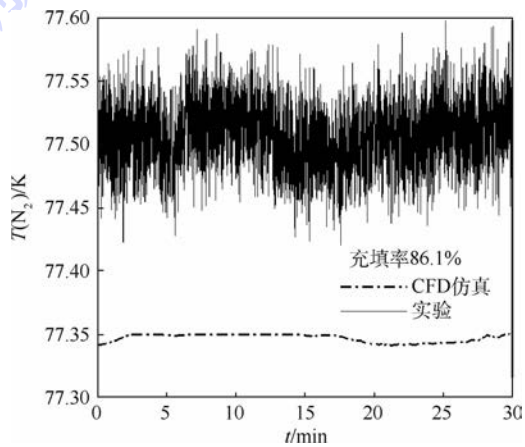


图 11 实验和仿真得出的液氮温度的比较

Fig. 11 Comparison of liquid nitrogen temperature between experiment and simulation

体区的温度略低于饱和温度,液体区等温线表明了液体处于剧烈的运动中。气枕区温度存在温度分层,距出口越近,温度越高,且气液界面附近气枕区温度梯度较小。通过比较仿真过程得出在液体区的温度与实验中测量的液体温度基本保持一致,液体区的温度基本处于饱和温度,并且液体区存在气泡运动引起的剧烈翻滚运动,可以认为贮箱内部的液体处于沸腾状态。气枕区的温度值仿真与实验存在差异,因为实际过程中气枕区的温度测点受到外界导热影响,但是仿真过程气枕区温度的分布趋势与实验基本一致。因此,可以认为仿真对低温液体常压停放过程有准确的预测性。

由于实验没有直接测量贮箱内流体的运动及贮箱内壁面与流体间的热量交换值,本文给出仿真具体数值结果。图 13 给出仿真过程贮箱内流体的速度分布。可以看出贮箱内液体由于底部受热向气液界面处运动,靠近气液界面处液体沿壁面向下运动。气枕区与液体在气液界面处存在着强烈的对流换热,壁面附近气体沿壁面向上运动。而随着时间的推移,贮箱内流体的运动趋于稳定。

图 14 给出流体与固体壁面的换热量 Q。可以看出:1 min前贮箱内所有的内壁面换热量呈上升趋势,1 min 后贮箱底部液体内壁面和气枕区接触内壁面换热量基本保持不变,气枕内壁面换热量值在 60 W 左右,底部液体壁面的换热量在 500 W左右,而贮箱柱状段的换热量和总的换热量随时间增加而减少,柱状段液体换热量随时间从 1 164 W 降至 940 W,总的换热量从 1 736 W 降至 1 520 W。贮箱柱状段内壁面换热量减少是因为液体的沸腾使气体与固体壁面接触面积变大,而气体与固体壁面的换热能力远小于液体与壁面

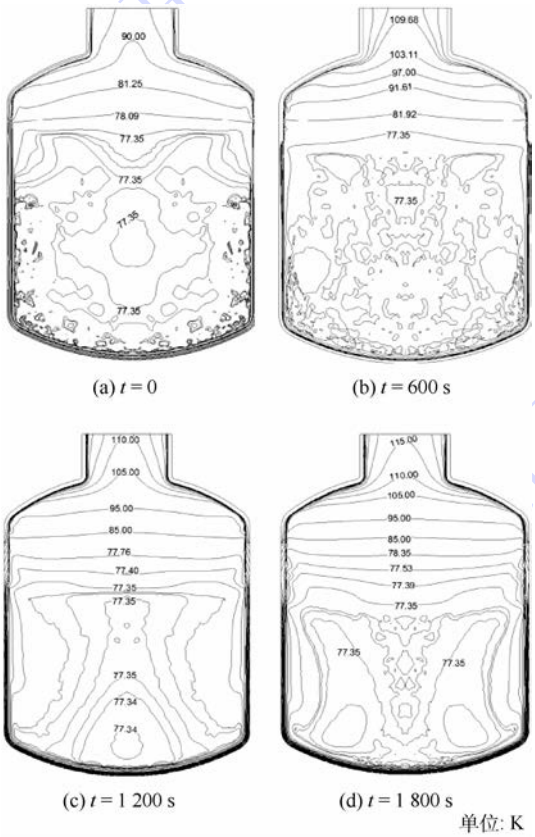


图 12　仿真得到的贮箱内流体温度分布

Fig. 12　Fluid temperature distribution simulated in tank

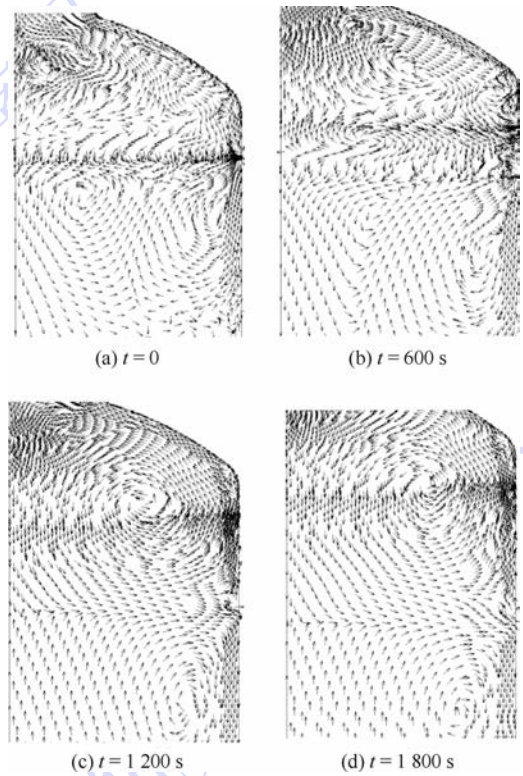


图 13　仿真得到的贮箱内流体速度分布

Fig. 13　Fluid velocity distribution simulated in tank

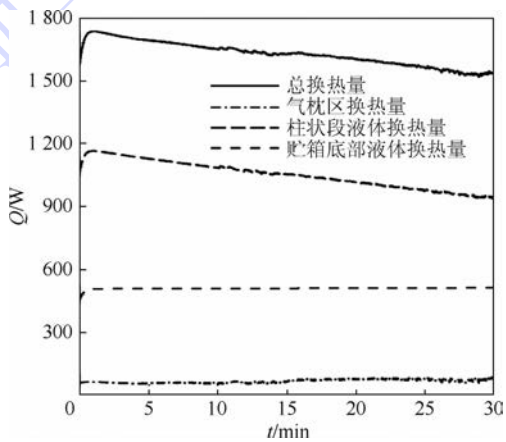


图 14　仿真得到的流体与固体壁面的换热量与时间的关系

Fig. 14　Relationship of heat transition between liquid and solid wall simulated in tank and time

北航学报 赠阅

的换热能力。结合本节液体温度和体积变化关系的分析,可以得出常压停放液体的汽化主要取决于固体壁面对液体的加热。

4 结 论

1)常压停放过程,环境温度越高和贮箱的体积充填率越大,液体的汽化速率越大。

2)常压停放过程,液体区温度基本一致,液体区处于饱和状态;气枕区温度存在沿轴向的分层,气体离贮箱压力出口越近,温度越高。

3)常压停放过程达到稳定后,液体区与贮箱固体壁面换热量远大于气枕区与贮箱固体壁面换热量,液体区的换热量对液体汽化起决定作用。

4)运用混合物模型,并根据相变系数模型能够准确的对低温液体常压停放过程进行预测,并能为低温液体自增压过程的仿真提供准确的初始状态。

参考文献(References)

[1] STOCHL R J, KNOLL R H. Thermal performance of a liquid hydrogen tank multilayer insulation system at warm boundary temperatures of 630, 530, and 152 R[C]// AIAA, SAE, ASME, and ASEE, 27th Joint Propulsion Conference. Reston: AIAA, 1991: 1-20.

[2] BARSI S, KASSEMI M, PANZARELLA C H, et al. A tank self-pressurization experiment using a model fluid in normal gravity [C]// AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2005: 1-13.

[3] SEO M, JEONG S. Analysis of self-pressurization phenomenon of cryogenic fluid storage tank with thermal diffusion model [J]. Cryogenics, 2010, 50(9): 549-555.

[4] DAS S P, CHAKRABORTY S, DUTTA P. Studies on thermal stratification phenomenon in LH_2 storage vessel [J]. Heat Transfer Engineering, 2004, 25(4): 54-66.

[5] WANG C L, LI Y, WANG R S. Performance comparison between no-vent and vented fills in vertical thermal-insulated cryogenic cylinders[J]. Experimental Thermal and Fluid Science, 2011, 35(2): 311-318.

[6] 聂中山,李青,洪国同,等. 车载液氢杜瓦蒸发率理论与试验研究[J]. 低温工程,2004(4):55-58.
NIE Z S, LI Q, HONG G T, et al. Theoretical and experimental investigation of evaporation rate in vehicle liquid hydrogen dewar[J]. Cryogenics, 2004(4): 55-58(in Chinese).

[7] 王贵仁. 低温容器内压力变化规律的研究[D]. 兰州:兰州理工大学,2008:28-40.
WANG G R. The research on variation law of pressure in cryogenic vessel[D]. Lanzhou: Lanzhou University of Technology, 2008: 28-40(in Chinese).

[8] 乔国发. 影响 LNG 储存容器蒸发率因素的研究[D]. 东营:中国石油大学,2007:73-87.
QIAO G F. The study on the influential factors of evaporation rate of the liquefied natural gas tank[D]. Dongying: China University of Petroleum, 2007: 73-87(in Chinese).

[9] 代予东,赵红轩. 运用数学方法模拟推进剂贮箱增压[J]. 火箭推进,2003,29(3):34-40.
DAI Y D, ZHAO H X. Numerical modeling of pressurization of a propellant tank[J]. Journal of Rocket Propulsion, 2003, 29(3): 34-40(in Chinese).

[10] ESTEY P N, LEWIS D H, CONNOR M. Prediction of a propellant tank pressure history using state space methods[J]. Journal of Spacecraft and Rockets, 1983, 20(1): 49-54.

[11] PANZARELLA C H, KASSEMI M. On the validity of purely thermodynamic descriptions of two-phase cryogenic fluid storage [J]. Journal of Fluid Mechanics, 2003, 484: 41-68.

[12] 刘展,孙培杰,李鹏,等. 地面停放低温液氧贮箱热物理过程研究[J]. 西安交通大学学报,2016,50(9):36-42.
LIU Z, SUN P J, LI P, et al. Research on the thermal physical process of cryogenic liquid oxygen tank in ground parking[J]. Journal of Xi'an Jiaotong University, 2016, 50(9): 36-42(in Chinese).

[13] CHEN L, LIANG G Z. Simulation research of vaporization and pressure variation in a cryogenic propellant tank at the launch site[J]. Microgravity Science and Technology, 2013, 25(4): 203-211.

[14] FADHL B, WROBEL L C, JOUHARA H. CFD modelling of a two-phase closed thermosyphon charged with R134a and R404a [J]. Applied Thermal Engineering, 2015, 78: 482-490.

[15] 陈亮,梁国柱,邓新宇,等. 贮箱内低温推进剂汽化过程的 CFD 数值仿真[J]. 北京航空航天大学学报,2013,39(2):264-268.
CHEN L, LIANG G Z, DENG X Y, et al. CFD numerical simulation of cryogenic propellant vaporization in tank[J]. Journal of Beijing University of Aeronautics and Astronautics, 2013, 39(2): 264-268(in Chinese).

[16] Fluent A. ANSYS I: Fluent theory guide[EB/OL]. Canonbury: ANSYS Inc., 2010. https://support.ansys.com/.

[17] 刘泉. 纯蒸气及含不凝气蒸气冷凝的数值研究[D]. 合肥:中国科学技术大学,2015:73-87.
LIU Q. Numerical investigation on condensation with and without non-condensable gas[D]. Hefei: University of Science and Technology of China, 2015: 73-87(in Chinese).

[18] KNUDSEN M, PARTINGTON J R. The kinetic theory of gases: Some modern aspects[J]. The Journal of Physical Chemistry, 1935, 39(2): 307.

[19] DE SCHEPPER S C K, HEYNDERICKX G J, MARIN G B. Modeling the evaporation of a hydrocarbon feedstock in the convection section of a steam cracker[J]. Computers and Chemical Engineering, 2009, 33(1): 122-132.

[20] LIU Z, SUNDEN B, YUAN J. VOF modeling and analysis of filmwise condensation between vertical parallel plates[J]. Heat Transfer Research, 2012, 43(1): 47-68.

[21] LEE W H. A pressure iteration scheme for two-phase flow modeling: LA-UR-79-975[R]. New Mexico: Los Alamos National Laboratory, 1979.

[22] 陈国邦. 低温工程材料[M]. 杭州:浙江大学出版社,1998:195-201.

CHEN G B. Cryogenic engineering materials[M]. Hangzhou: Zhejiang University Press,1998:195-201(in Chinese).

[23] ISHII M, HIBIKI T. Thermo-fluid dynamics of two-phase flow [M]. Berlin: Springer,2011:361-395.

[24] 陈国邦,黄永华,包锐. 低温流体热物理性质[M]. 北京:国防工业出版社,2006:190-200.
CHEN G B, HUANG Y H, BAO R. Cryogenic fluid thermos physical properties [M]. Beijing: National Defense Industry Press,2006:190-200(in Chinese).

作者简介:

李佳超　男,博士研究生。主要研究方向:运载火箭低温贮箱工作过程。

梁国柱　男,博士,教授,博士生导师。主要研究方向:火箭发动机与运载火箭动力系统设计与仿真。

Experiment and numerical simulation of liquid nitrogen tank atmospheric ground parking

LI Jiachao, LIANG Guozhu *

(School of Astronautics, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: In order to study the progress of the cryogenic propellant during atmospheric ground parking, a visualization liquid nitrogen tank experiment system was designed. The experiment researched how the filling rate and ambient temperature affected the evaporation mass of liquid nitrogen, and measured the fluid inside tank and the wall temperature outside tank which changed with the time and location. The experimental results show that during atmospheric ground parking the phase transition mainly happens in the wall and gas-liquid interface, air pillow zone has temperature levels, and the air pillow temperature increases with the decrease of the distance from the exit. The liquid stays in the saturated state with almost consistent temperature, and the outer wall temperature distribution of the tank is significantly different in the axial direction and lower in liquid zone. The heat and mass transfer between liquid and gas is deduced from the Hertz-Knudsen equation based on the molecular dynamics theory. According to the temperature boundary conditions acquired from the experiment, physical process of 30 min in liquid nitrogen tank during atmospheric ground parking was simulated using mixture model. The simulation results show that the deviation of volume vaporization rate between simulation and experiment is within 5%, and the deviation of temperature simulation and experiment in the liquid zone is about 0.15 K.

Keywords: cryogenic tank; atmospheric ground parking; phase transition; heat and mass transfer; mixture model

Received: 2017-01-12; **Accepted**: 2017-05-19; **Published online**: 2017-06-19 15:16

URL: kns.cnki.net/kcms/detail/11.2625.V.20170619.1516.002.html

Foundation item: China Academy of Launch Vehicle Technology-University Joint Innovation Fund Project (CALT201302)

* **Corresponding author**. E-mail: lgz@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0018

星载 GNSS-R 辅助海洋盐度探测方法

王峰，杨东凯*
（北京航空航天大学 电子信息工程学院，北京 100083）

摘 要：全球导航卫星系统反射计(GNSS-R)是近年来兴起的一种被动式遥感手段，可用于提高海洋盐度(SSS)反演精度。首先，在回顾辐射计亮温模型和 GNSS-R 散射功率模型，并建立星载仿真场景的基础上，研究了 GNSS-R 辅助辐射计探测海洋盐度的性能，使辐射计工作于 GPS L1 频点 1575.42 MHz 时，通过共用天线和射频前端可以降低星载设备的质量和功耗，但对海洋盐度大于 25 psu 的条件下，垂直和水平极化的亮温对海洋盐度的灵敏度分别下降约 0.1 和0.08 K/psu；其次，分析了 GPS L1 反射信号对辐射计的干扰，发现在仿真场景下当辐射亮温变化 1 K 时，GPS L1 反射信号引入了小于 2.5×10^{-4} K 的误差；再者，讨论了不同入射角情况下定义的垂直和水平极化的 GNSS-R 观测量对亮温校正量的灵敏度，结果表明随入射角增大，水平、垂直极化信号的观测量对亮温校正量的灵敏度分别呈现下降和上升趋势；最后，分析了定义的 GNSS-R 观测量对亮温校正量的灵敏度与空间分辨率之间的关系，得出了高灵敏度、高空间分辨率反演算法的研究对星载 GNSS-R 辅助辐射计海洋盐度探测至关重要的结论。

关 键 词：全球导航卫星系统反射计(GNSS-R)；海洋盐度(SSS)；辅助探测；灵敏度；空间分辨率
中图分类号：P731.12；TN959.4
文献标识码：A **文章编号**：1001-5965(2018)01-0108-09

海洋盐度(SSS)是研究全球海洋循环和气候变化的重要参数，对海洋环境监测和海上经济活动的安全保障具有至关重要的作用[1]。从探测机理进行分类，海洋盐度探测分为接触式和遥感式。浮标作为最常用的接触式探测，已得到广泛应用，但空间覆盖率受到限制。遥感式根据搭载平台的不同分为岸基、机载和星载探测。星载海洋盐度探测具有全天候、大范围连续观测的特点，适合全球范围内的长期观测。尽管其探测机理的研究已较为成熟，但由于受到探测精度和空间分辨率的限制，星载探测任务的起步较晚。目前，典型的海洋盐度探测卫星主要有 2 颗，即欧洲航天局(ESA)的 SMOS (Soil Moisture and Ocean Salinity)卫星[2]和美国国家航空航空局(NASA)的 Aquarius/SAC-D 卫星[3]。由于 L 波段被公认为海洋盐度探测的最佳频段(1 400 ~ 1 427 MHz)[4]，因此，2 颗卫星搭载的遥感器均是 L 波段辐射计。由于辐射亮温不仅受海洋盐度影响，而且受海洋温度和粗糙度等影响，海洋盐度探测精度受到一定限制。为了获得精确的海洋盐度结果，一方面，可以通过多角度、多频段观测，同时获取海洋盐度、粗糙度以及温度等信息[5]；另一方面，可以增加海面粗糙度和温度的探测载荷，利用同步获取的粗糙度和温度信息校正海洋盐度探测误差[3]。2 个方案无疑都增加了卫星载荷的复杂性、质量以及功耗。

收稿日期：2017-01-13；**录用日期**：2017-03-31；**网络出版时间**：2017-05-27 17:21
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170527.1721.004.html
基金项目：北京航空航天大学基本科研业务费-博士研究生创新基金
* **通信作者**. E-mail：yangdongkai@sina.com

引用格式：*王峰，杨东凯. 星载 GNSS-R 辅助海洋盐度探测方法[J]. 北京航空航天大学学报，2018，44(1)：108-116.*
WANG F，YANG D K. Sea surface salinity determination method assisted by spaceborne GNSS-R [J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(1)：108-116 (in Chinese).

全球导航卫星系统反射计（GNSS-R）利用导航卫星反射信号进行地表参数的探测，是一门新型的机会源探测技术，具有时空分辨率高，设备质量轻、体积小等特点。目前，已通过岸基、机载以及星载实验验证可用于海面粗糙度[6]、海冰[7]以及土壤湿度[8]等的探测。是否可将该技术用于探测 SSS 已经成为最近的研究课题之一。文献[9]分析了介电常数对海洋盐度的灵敏性，并进行相应的岸基实验，得出全球导航卫星系统（GNSS）反射信号的幅度和相位对海洋盐度灵敏性均较弱的结论。文献[10]从雷达散射截面的角度分析了直接利用 GNSS-R 探测海洋盐度的可行性。加泰罗尼亚理工大学提出的 PAU（Passive Advanced Unit）系统通过红外辐射计和 GNSS-R 接收机辅助 L 波段辐射计提高海洋盐度反演精度[11]。

本文对星载 GNSS-R 辅助海洋盐度探测的技术进行了探索和分析。首先，回顾了亮温模型和 GNSS-R 散射功率模型，建立了星载仿真场景；然后，从一级输出量 DDM（Delay-Doppler Map）的角度出发定义了 GNSS-R 观测量——相关功率和；最后，讨论了导航反射信号对辐射探测的干扰，以及入射角和空间分辨率对 GNSS-R 辅助辐射计探测海洋盐度的影响。

1　理论模型

1.1　亮温模型

由于受海面粗糙度的影响，海表面的辐射亮温 $T_{B,p}$ 可以分为平静海面的亮温和粗糙度引起的亮温校正量 $\Delta T_{B,p}$[5]：

$$T_{B,p}(f,\theta,T,S,U_{10}) = e_p(f,\theta,T,S)T + \Delta T_{B,p}(\theta,U_{10}) \tag{1}$$

式中：e_p 为平静海面的发射率，下标 p = H，V，H 和 V 分别表示水平和垂直极化；f 为信号频率；θ 为信号入射角；T 为海面温度；S 为海洋盐度；U_{10} 为距海面 10 m 处的风速。从式（1）可以看出，当信号频率和接收平台几何配置已知时，辐射亮温与海洋盐度、粗糙度以及温度相关。由于本文主要讨论利用导航卫星反射信号校正海面粗糙度对辐射计盐度探测的影响，因此，海面温度被认为是已知量，在后续分析中，假设为 25 ℃。

1.1.1　平静海面反射率

e_p 为菲涅耳反射率的函数：

$$e_p(f,\theta,T,S) = 1 - \Gamma_p(f,\theta,T,S) \tag{2}$$

式中：反射率 Γ_p 可表示为

$$\Gamma_H(f,\theta,T,S) = \left|\frac{\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\cos\theta + \sqrt{\varepsilon - \sin^2\theta}}\right|^2 \tag{3}$$

$$\Gamma_V(f,\theta,T,S) = \left|\frac{\varepsilon\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\varepsilon\cos\theta + \sqrt{\varepsilon - \sin^2\theta}}\right|^2 \tag{4}$$

式中：ε 为海水介电常数，为频率、温度及海洋盐度的函数。由于 Klein-Swift 海水介电常数模型在微波辐射计反演海洋盐度中被广泛应用[12]，因此，在本文的仿真中，亦采用该模型。

1.1.2　亮温校正量

海面粗糙度引起海面辐射亮温的单调增加。为了精确测量海洋盐度，必须对海面粗糙度引起的附加项进行校正。ESA 在 2000 年和 2001 年进行了 WISE（Wind and Salinity Experiment）[13]，旨在研究海面粗糙度引起的亮温校正量与信号入射角之间的关系，并得出了经验模型，但该模型与最新星载辐射计数据不能很好吻合。Hejazin 等基于 El-Nimri 模型[14]，利用实测数据拟合得出了更适合于 Aquarius 数据修正的风速与亮温校正量的关系[15]。该模型在 0 ~ 20 m/s 的风速范围内，与实测数据保持一致。由于本文在后续分析中采用 Aquarius 卫星轨道高度，因此，海面粗糙度引起的亮温校正量模型采用 Hejazin 等提出的模型[15]。该模型表示如下：

$$\Delta T_{B,p} = (a_{p,0} + a_{p,1}Z_{p,\theta} + a_{p,2}Z_{p,U_{10}} + a_{p,3}Z_{p,\theta}Z_{p,U_{10}})\sqrt[4]{f} \tag{5}$$

式中：$Z_{H,\theta}$、$Z_{H,U_{10}}$、$Z_{V,\theta}$ 以及 $Z_{V,U_{10}}$ 可分别表示为

$$\begin{cases} Z_{H,\theta} = 0.5 + \dfrac{\arctan((\theta - b_{H,1})/b_{H,2})}{\pi} \\ Z_{H,U_{10}} = 0.5 + \dfrac{\arctan((U_{10} - c_{H,1})/c_{H,2})}{\pi} \\ Z_{V,\theta} = \exp\left(-\exp\left(\dfrac{b_{V,1} - \theta}{b_{V,2}}\right)\right) \\ Z_{V,U_{10}} = \exp\left(-\exp\left(\dfrac{c_{V,1} - U_{10}}{c_{V,2}}\right)\right) \end{cases} \tag{6}$$

其中：$a_{p,0}$、$a_{p,1}$、$a_{p,2}$、$a_{p,3}$、$b_{p,1}$、$b_{p,2}$、$c_{p,1}$ 及 $c_{p,2}$ 为拟合参数，具体值可参考文献[14]。

1.2　GNSS-R 散射功率模型

GNSS-R 的基础观测量是接收机输出的一级输出量 DDM，表征了能量在时延-多普勒域的分布情况。Zavorotny 和 Voronovich 根据基尔霍夫光学近似模型（KA-GO）推导了 GNSS-R 散射功率模型，其表达式为[16]

$$\langle |Y(\tau,f)|^2 \rangle = \frac{P_t G_t \lambda^2 T_i^2}{(4\pi)^2 L_{ft} L_{fr}} \int_A \frac{G_r(\boldsymbol{r})}{4\pi R_r^2(\boldsymbol{r}) R_t^2(\boldsymbol{r})} \cdot \Lambda^2(\tau - \tau(\boldsymbol{r}))\,\mathrm{sinc}(f - f(\boldsymbol{r}))\,\sigma_0(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} \tag{7}$$

北航学报 赠阅

式中：τ 为信号时延；P_t 和 G_t 分别为导航卫星发射功率和增益；λ 为导航信号波长；T_i 为相关积分时间；L_{ft} 和 L_{fr} 分别为下行和上行链路的大气损耗；$\boldsymbol{r}$ 为散射单元；A 为积分区域；$G_r(\boldsymbol{r})$ 为散射单元 $\boldsymbol{r}$ 处的接收天线增益；$R_t(\boldsymbol{r})$ 和 $R_r(\boldsymbol{r})$ 分别为导航卫星和接收机到散射单元 $\boldsymbol{r}$ 的几何距离；$\Lambda(\tau)$ 为导航伪随机码的自相关函数；$\mathrm{sinc}(f)$ 为辛格函数；$\sigma_0(\boldsymbol{r})$ 为散射单元 $\boldsymbol{r}$ 处的散射截面，通过基尔霍夫近似方法计算，可以表示为

$$\sigma_0(\boldsymbol{r}) = \pi\Gamma_p \frac{|\boldsymbol{q}|^4}{q_z^4} P_{\mathrm{pdf}}\left(-\frac{\boldsymbol{q}_\perp}{q_z}\right) \tag{8}$$

其中：$\boldsymbol{q}=(\boldsymbol{q}_\perp, q_z)$ 为散射矢量，$\boldsymbol{q}_\perp=(q_x, q_y)$，$q_x$、$q_y$、$q_z$ 分别为散射矢量 x、y、z 轴分量；$P_{\mathrm{pdf}}(x,y)$ 为海面斜度的概率密度函数，假设服从二维高斯分布：

$$P_{\mathrm{pdf}}(x,y) \sim N(0,\sigma_{sx}^2;0,\sigma_{sy}^2;\rho) \tag{9}$$

式中：σ_{sx}^2、σ_{sy}^2 和 ρ 分别为海面斜度 x、y 轴方差和相关系数，可以根据海浪谱模型确定，本文采用 Elfouhaily 模型[17]。从式(8)可知，GNSS-R 接收机输出的 DDM 中，同时包括了菲涅耳反射系数，即海洋盐度信息和海面斜度，即海面粗糙度信息，因此，从理论的角度看，GNSS-R 技术，可以反演海洋粗糙度和海洋盐度。但是，文献[9-10]从不同的角度分析了 GNSS-R 直接探测海洋盐度的可行性，结果表明就目前工业技术水平，该方案可行性较低。

2 星载 GNSS-R 几何关系

星载 GNSS-R 几何关系如图 1 所示，原点位于地心，YOZ 平面与信号入射面共面，Z 轴指向镜面反射点切面法向方向。图中：S 为镜面反射点；R_e 为地球半径，采用地球圆模型，为 6 731 km；h_t 为导航卫星轨道高度，假设为 20 200 km；h_r 为低轨(LEO)卫星轨道高度，以 Aquarius/SAC-D 卫星高度 657 km 为例；$X_rY_rZ_r$ 和 $X_tY_tZ_t$ 分别为 LEO 和导航卫星载体坐标系；φ 为天线波束角，假设 3 dB波束宽度为 25°，最大增益为 12 dB，方向增益图为高斯型，即

$$G_r(\boldsymbol{r}) = 15.85\exp\left(-\left(\frac{\phi(\boldsymbol{r})}{30.03°}\right)^2\right) \tag{10}$$

式中：$\phi(\boldsymbol{r})$ 为 LEO 卫星和散射单元 $\boldsymbol{r}$ 连线与天线指向角之间的夹角，假设天线指向角与 θ 相等。

导航卫星、LEO 卫星及散射单元 $\boldsymbol{r}$ 的位置分别为

$$\boldsymbol{T}_p = (0, D\sin\theta, R_e + D\cos\theta) \tag{11}$$

$$\boldsymbol{R}_p = (0, -d\sin\theta, R_e + d\cos\theta) \tag{12}$$

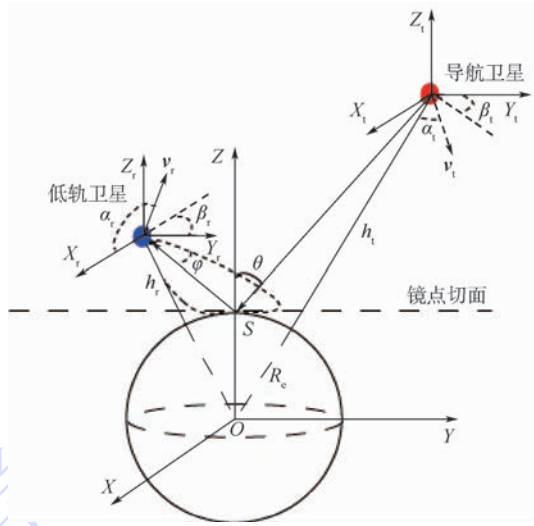


图 1 GNSS-R 几何关系

Fig. 1 GNSS-R geometry

$$\boldsymbol{S}_{x,y} = (\delta_x, \delta_y, 1)R_e \tag{13}$$

式中：δ_x 和 δ_y 分别为散射单元 $\boldsymbol{r}$ 与地心连线在 XOZ 和 YOZ 平面内的投影与 Z 轴之间的夹角；D 和 d 分别为导航卫星和 LEO 卫星到镜面反射点的几何距离，即

$$D = -R_e\cos\theta + \sqrt{(h_t+R_e)^2 - R_e^2\sin^2\theta} \tag{14}$$

$$d = -R_e\cos\theta + \sqrt{(h_r+R_e)^2 - R_e^2\sin^2\theta} \tag{15}$$

导航卫星、LEO 卫星的速度分别为

$$\boldsymbol{v}_t = v_t(\cos\alpha_t, \sin\alpha_t\cos\beta_t, \sin\alpha_t\sin\beta_t) \tag{16}$$

$$\boldsymbol{v}_r = v_r(\cos\alpha_r, \sin\alpha_r\cos\beta_r, \sin\alpha_r\sin\beta_r) \tag{17}$$

式中：v_t 和 v_r 分别为导航卫星和 LEO 卫星的速率，分别为 2.8 和 7.5 km/s；α_t 和 α_r 分别为导航卫星和 LEO 卫星速度与 X 轴之间的夹角；β_t 和 β_r 分别为导航卫星和 LEO 卫星与地心连线的垂直方向与 Y 轴之间的夹角，可表示为

$$\beta_t = \arcsin\left(\frac{D\cos\theta}{h_t+R_e}\right) \tag{18}$$

$$\beta_r = \arcsin\left(\frac{d\cos\theta}{h_r+R_e}\right) \tag{19}$$

3 GNSS-R 观测量

由 1.2 节描述的 GNSS-R 散射功率模型可知，DDM 中包含了海面粗糙度信息，为时延、多普勒的函数。利用 DDM 反演海面粗糙度最直接有效的方法是建立海面粗糙度和 DDM 特征参数(观测量)之间的联系。本文定义特定区域内的相关功率和 GNSS-R 观测量 M_{obs} 作为相关功率和 DDM 特征参数。

$$M_{\mathrm{obs}} = \int_{f_{\min}}^{f_{\max}}\int_{\tau_{\min}}^{\tau_{\max}} \langle |Y(\tau,f)|^2 \rangle \mathrm{d}\tau\mathrm{d}f \tag{20}$$

北航学报 赠阅

式中：$f_{\max}$、$f_{\min}$分别为多普勒积分范围的上、下限；$\tau_{\max}$、$\tau_{\min}$分别为时延积分范围的上、下限。

图 2 给出了式(20)积分区域的示意图。由于 DDM 分布的起始位置与海面信息无关，均为 −1，因此，本文 $\tau_{\min}$ 固定为 −1。如图 3 所示，海面的不同散射单元具有不同的时延和多普勒，因此 DDM 不仅反映了相关功率随时延-多普勒的分布情况，同时也反映了功率在反射面上的分布，即式(20)中的积分区域将决定观测量的空间分辨率。假设 $f_{\max}=|f_{\min}|$，定义有效空间分辨率为[18]

$$R_{\mathrm{S}}=\sqrt{\int\prod\left(\frac{\Delta\tau(x,y)-\frac{\tau_{\max}}{2}}{\tau_{\max}}\right)\prod\left(\frac{\Delta f(x,y)}{2f_{\max}}\right)\mathrm{d}x\mathrm{d}y} \tag{21}$$

式中：$\Delta\tau(x,y)$和$\Delta f(x,y)$分别为散射单元 $\boldsymbol{r}$ 与镜面反射点之间的时延和多普勒差。散射单元 $\boldsymbol{r}$ 的时延和多普勒可分别表示为

$$\tau(x,y)=\frac{|\boldsymbol{T}_{\mathrm{p}}-\boldsymbol{S}_{x,y}|+|\boldsymbol{R}_{\mathrm{p}}-\boldsymbol{S}_{x,y}|}{c} \tag{22}$$

$$f(x,y)=\frac{\boldsymbol{v}_{\mathrm{t}}\cdot\boldsymbol{m}_{x,y}-\boldsymbol{v}_{\mathrm{r}}\cdot\boldsymbol{n}_{x,y}}{\lambda} \tag{23}$$

式中：c为光速；$\boldsymbol{m}_{x,y}$和$\boldsymbol{n}_{x,y}$分别为入射信号和散射单元 $\boldsymbol{r}$ 对应散射分量的单位矢量。

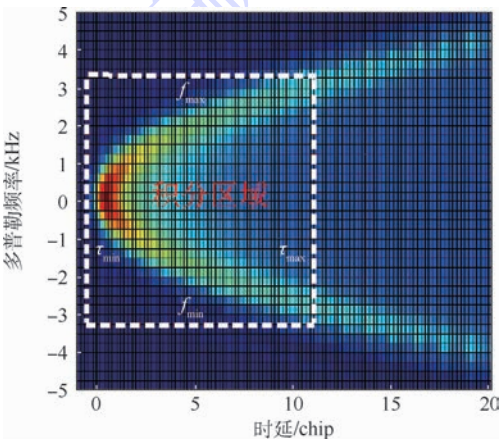


图 2　观测量积分区域示意图

Fig. 2　Schematic of integration region of observable

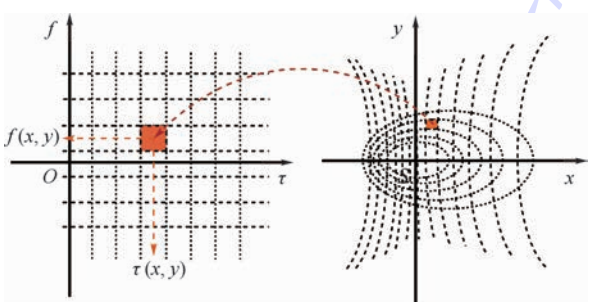


图 3　空间域到时延-多普勒域映射示意图

Fig. 3　Schematic of mapping from spatial domain to delay-Doppler domain

图 4 给出了高度角为 30°和 45°时，不同多普勒范围条件下，有效空间分辨率随最大时延的变化曲线。从图中可以看出：①随着最大时延和多普勒增大，有效空间分辨率下降；②当多普勒范围足够大时，决定有效空间分辨率的是最大时延；③入射角越小，有效空间分辨率越高。

不同的入射角和观测量积分区域对 GNSS-R 辅助辐射计修正亮温校正量的性能有影响。本文定义观测量对亮温校正量的灵敏度为 $S_{\Delta T_{\mathrm{B},p}}$，进行上述影响的分析。

$$S_{\Delta T_{\mathrm{B},p}}=\frac{\partial M_{\mathrm{obs}}}{\partial\Delta T_{\mathrm{B},p}}=\frac{\partial M_{\mathrm{obs}}}{\partial U_{10}}\cdot\frac{\partial U_{10}}{\partial\Delta T_{\mathrm{B},p}} \tag{24}$$

由于对式(5)进行微分的解析求解较困难，因此，在后续分析中，采用数值仿真的方法求解式(24)所定义的灵敏度。

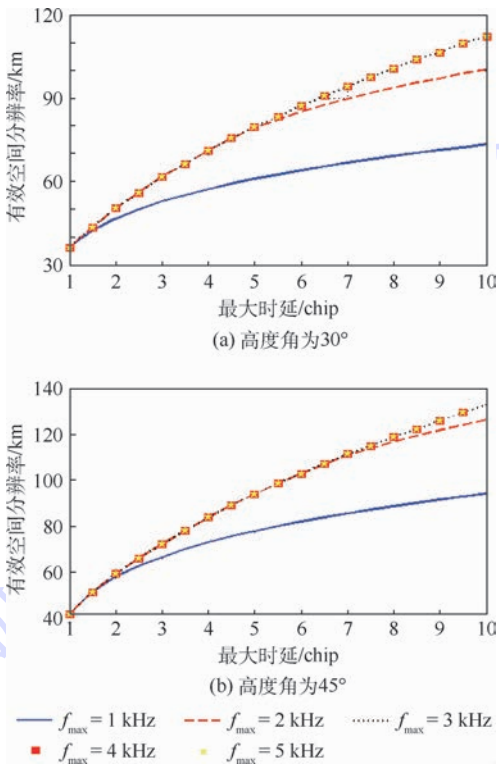


图 4　当高度角为 30°和 45°时，有效空间分辨率随最大时延的变化

Fig. 4　Variation of effective spatial resolution with maximal delay at elevation angle of 30° and 45°

4　GNSS-R 辅助探测方法

4.1　信号处理

利用 GNSS-R 作为海洋盐度探测的辅助设备具有如下优势：①降雨对 L 波段信号的衰减小，可以降低降雨对海面粗糙度反演的影响；②L 频

段占用率较低,可以减小其他信号对海面粗糙度校正的干扰;③无源探测技术,可以降低成本和设备质量,且通过与 L 波段辐射计共用天线、射频前端,可进一步降低设备质量和功耗,更利于星载平台的发展,尤其是未来微纳卫星的发展。图 5 给出了 GNSS-R 辅助辐射计探测海洋盐度的框图。该方案首先由加泰罗尼亚理工大学提出[11]。接收天线包括右旋、垂直以及水平极化天线,其中对天指向的右旋天线接收 GNSS 直射信号;对地的垂直、水平极化天线用于接收海面辐射信号以及经海面反射的 GNSS 信号。直射信号通过捕获、跟踪、定位解算等步骤得到 LEO 卫星位置和速度信息。卫星选择旨在根据直射信号定位结果以及选星准则选择合适的卫星并控制波束形成单元使得垂直、水平极化天线指向所选择卫星的镜面反射点。补偿估计是根据 LEO 和导航卫星的位置、速度信息估计反射信号相对于直射信号的码延时和多普勒频率。本地信号发生器产生多路不同频率本地载波和不同码延时的本地码信号。反射信号接收机将多路本地载波、码信号和经海洋反射的 GNSS 信号进行互相关,并进行非相干累加产生 DDM。辐射计接收机将辐射信号进行平方积分检波得到辐射信号的功率信息。参数综合估计单元利用 DDM 提取的观测量,并根据亮温校正量和海洋盐度反演模型得到海洋盐度信息。

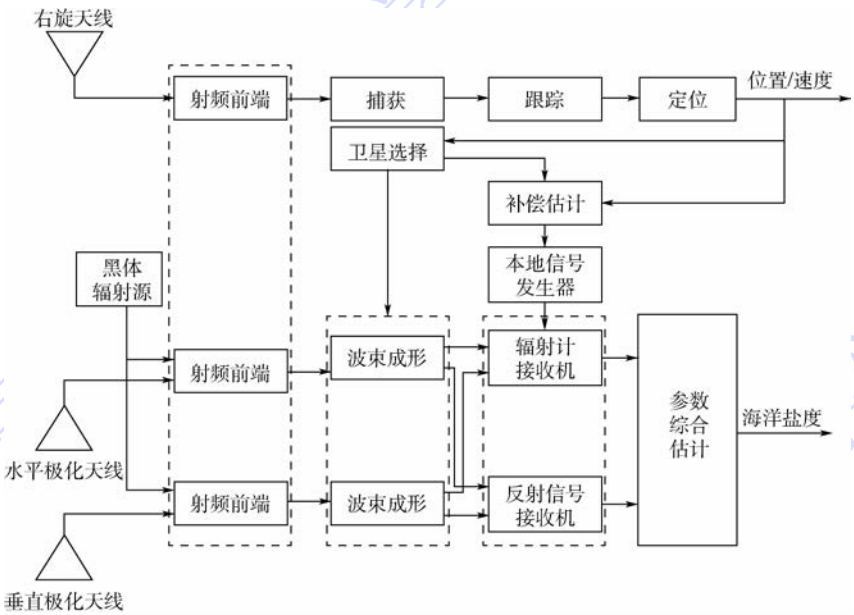


图 5 GNSS-R 辅助辐射计海洋盐度探测框架

Fig. 5 SSS determination architecture of radiometer assisted by GNSS-R

4.2 灵敏度分析

尽管卫星导航信号位于 L 波段,但是 L 波段辐射计的最佳频段为 1 400 ~ 1 427 MHz,例如 SMOS 和 Aquarius/SAC-D 采用的射频均为 1 413 MHz,而 GPS L1 为 1 575.42 MHz。当 L 波段辐射计工作在导航信号频段时,亮温对海洋盐度的灵敏度将有所下降。本文以 GPS L1 为例进行仿真分析。图 6 为不同频率的海面亮温对海洋盐度的灵敏度随盐度的变化曲线。从图中可以看出,当辐射计工作在 L1 波段时,海面亮温对海洋盐度的灵敏度较最佳波段1 413 MHz下降,当盐度大于 25 psu,垂直极化(VP)的灵敏度下降约 0.1 K/psu,水平极化(HP)下降约 0.08 K/psu。因此,虽然使辐射计工作于 GPS L1 频点,并和 GNSS-R 共用天线和射

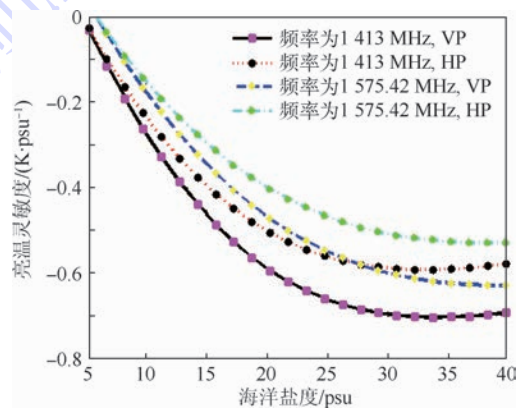


图 6 海面温度为 25 ℃,频率为 1 413 和 1 575.42 MHz 时,海面亮温对海洋盐度的灵敏度随海洋盐度的变化曲线

Fig. 6 Sensitivity curves of sea surface brightness temperature to SSS varying with SSS when sea surface temperature is 25 ℃ and frequency is 1 413 and 1 575.42 MHz

频前端可以降低设备质量和功耗，但也导致了盐度探测精度下降的不利因素。这对海面粗糙度的反演精度提出了更高的要求。

4.3　导航信号干扰

尽管导航信号采用扩频体制，如不进行解扩处理，接收信号功率埋没在噪声里，但是，仍会对辐射计测量产生干扰。为了衡量 GNSS 反射信号对辐射计的干扰程度，定义信干比为

$$R_{\mathrm{SI}}=\frac{T_{\mathrm{B},p}k_{\mathrm{B}}B}{P_{\mathrm{r}}} \tag{25}$$

式中：k_{B} 为玻尔兹曼常数；B 为辐射计带宽，本文假设为 20 MHz；P_{r} 为镜面反射点对应的反射信号功率，通过双基地雷达方程表示为

$$P_{\mathrm{r}}=\frac{G_{\mathrm{t}}P_{\mathrm{t}}\lambda^2G_{\mathrm{r}}\sigma_{\mathrm{S}}}{(4\pi)^3D^2d^2}A_{\mathrm{S}} \tag{26}$$

其中：σ_{S} 为镜面反射点的双基散射系数，可以通过式(8)得到；A_{S} 为镜面反射区域面积，假设为 25 km × 25 km。对于 GPS L1 信号，P_{t} 和 G_{t} 约为 26.8 W 和 12.1 dB[19]；λ 为 0.19 m。图 7 给出信干比随信号入射角的变化趋势。从图中可知，当海面温度为 25 ℃，海洋盐度为 25 psu 时，为对于水平、垂直极化辐射计，信干比均大于 35 dB，即辐射计测量 1 K 的亮温变化量时，GPS L1 反射信号引入了小于 2.5×10^{-4} K 的亮温误差。尽管随着入射角增大，水平极化辐射亮温呈下降趋势[15]，但由于 GPS L1 反射信号衰减比例更大，因此 GPS L1 反射信号对水平极化辐射亮温的干扰仍随入射角的增大而增大。

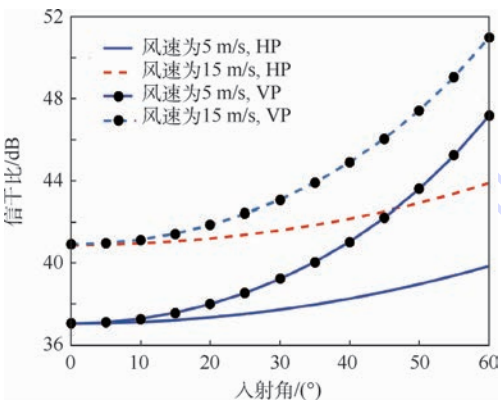


图 7　海面温度为 25 ℃，海洋盐度为 25 psu 时，信干比随信号入射角的变化

Fig. 7　Changing trend of signal-to-interference ratio with incident angle of signal when sea surface temperature is 25 ℃ and SSS is 25 psu

4.4　入射角影响

从式(5)和式(6)可以看出，亮温校正量与信号入射角相关，且 GNSS-R 的时延-多普勒相关功率 DDM 也受入射角影响，因此，GNSS-R 观测量对亮温校正量的灵敏度受入射角影响。基于模型式(5)～式(9)，利用式(24)对 GNSS 反射信号观测量对亮温校正量的灵敏度进行数值仿真。图 8 给出了不同入射角条件下的仿真结果。从图中可以看出：①亮温校正量与式(20)定义的反射信号观测量成反比例关系，主要原因是随着海面粗糙度的增加，海面辐射亮温和 GNSS 反射信号分别呈增大和减弱趋势；②随着亮温校正量的增大，观测量对亮温校正量的灵敏度下降，即随着海面粗糙度增加，GNSS 反射信号辅助辐射计校正亮温的性能下降；③对于水平极化信号，随着入射角的增加，观测量对亮温校正量的灵敏度呈下降趋势；④对于垂直极化信号，随着入射角增加观测量对亮温校正量呈增大趋势。随入射角增大，水平极化亮温对海面粗糙度越不敏感，而垂直极化亮温越敏感是造成上述③和④的主要原因。

4.5　空间分辨率影响

如第 3 节所述，观测量 M_{obs} 的定义与空间分辨率相关，因此，另一个需讨论的问题是观测量对亮温校正量的灵敏度与空间分辨率之间的关系。如图 9 所示，灵敏度与空间分辨率呈现反比例关系，即高的反演精度具有较差的空间分辨率。但是，值得注意的是：①空间分辨率需限制在同一风区内，即最大时延 $\tau_{\max}$ 的上限受制于风区大小，通常假设风区大小为 50 km，从图可知，空间分辨率小于 50 km 时，灵敏度较小，不利于高精度反演；②时延-多普勒采样个数受限于星载平台的实时处理能力，即当空间分辨率对应的最大时延和多普勒范围确定后，时延-多普勒采样步进需满足：

$$N=\left\lfloor\frac{\tau_{\max}-\tau_{\min}}{\tau_{\mathrm{step}}}\right\rfloor+1 \tag{27}$$

$$M=\left\lfloor\frac{2f_{\max}}{f_{\mathrm{step}}}\right\rfloor+1 \tag{28}$$

$$MN\leqslant N_{\max} \tag{29}$$

式中：M 为时延通道数；N 为多普勒通道数；τ_{step} 为时延步进；f_{step} 为多普勒频率步进；$\lfloor\cdot\rfloor$ 为向下取整算子；$N_{\max}$ 为处理单元实时处理容许的最大相关器数目。根据上述讨论，高精度、高空间分辨率反演算法的研究和在轨实时处理能力的提高是星载 GNSS-R 辅助 L 波段辐射计海洋盐度探测的关键技术之一。

北航学报 赠阅

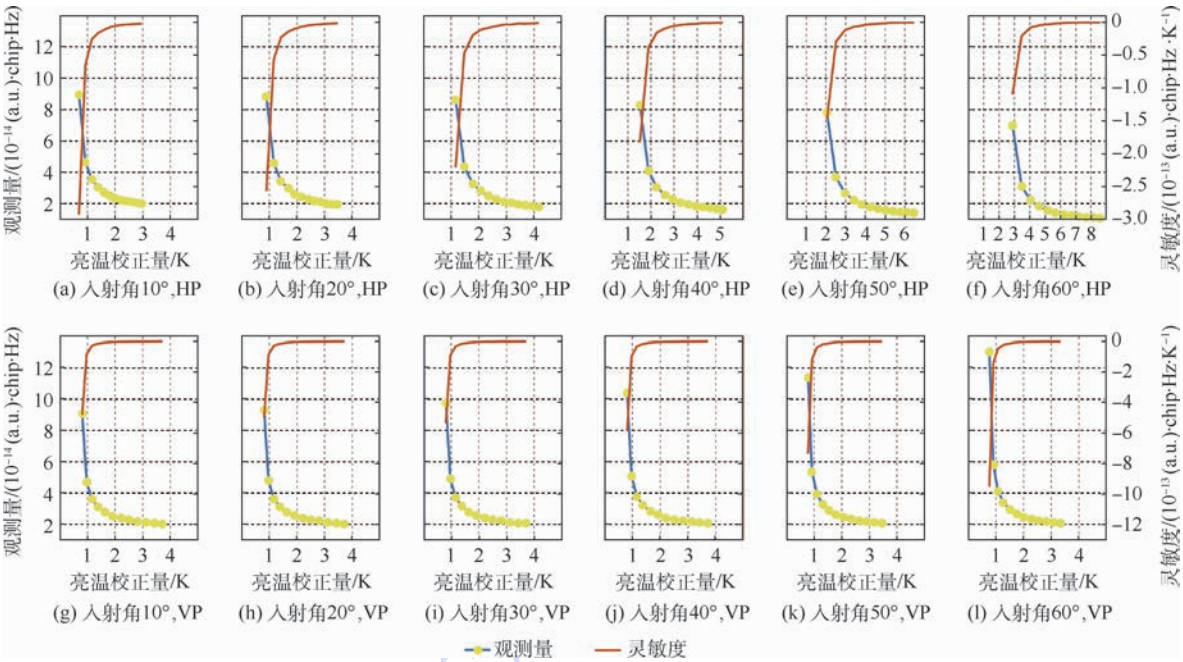


图 8 多普勒频率范围为[-3,3] kHz,最大时延为 5 chips 时,水平极化、垂直极化信号的观测量及其灵敏度与亮温校正量的关系

Fig. 8 Relationship of observable and its sensitivity with brightness temperature variation for horizontal and vertical polarization signal when Doppler frequency range is[-3,3] kHz and maximal delay is 5 chips

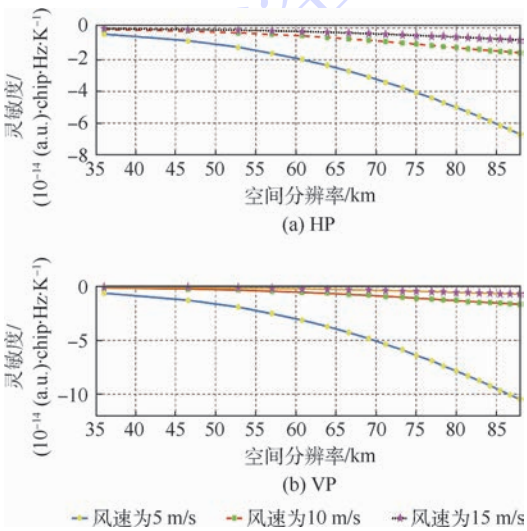


图 9 入射角为 30°,多普勒频率范围为[-1,1] kHz 时,水平极化、垂直极化信号观测量对亮温校正量的灵敏度与空间分辨率的关系

Fig. 9 Relationship between sensitivity of horizontal and vertical polarization signal observable to brightness temperature variation and spatial resolution when incident angle is 30° and Doppler frequency range is[-1,1] kHz

5 结 论

GNSS-R 技术是一种无源探测方法,具有低质量、低功耗等特点,将其用于辅助 L 波段辐射计进行海洋盐度探测可以降低星载设备的质量和功耗。本文讨论了 GNSS-R 辅助辐射计探测海洋盐度的方案,结论如下:

1) 使辐射计工作于 GPS L 频点 1 575.42 MHz 时,虽然通过与 GNSS-R 共用天线和射频前端可减小设备质量和功耗,但相比于辐射计最佳工作频段 1 413 MHz,海面亮温对盐度的灵敏度下降,当盐度大于 25 psu,垂直、水平极化的灵敏度分别下降约 0.1 和 0.08 K/psu。

2) 当辐射计工作在 GPS L1 频点时,辐射亮稳改变 1 K,GPS L1 反射信号对辐射计引入小于 2.5×10^{-4} K 的误差。

3) 对于水平极化信号,观测量对亮温校正量的灵敏度随入射角增大而减小。

4) 对于垂直极化信号,随着入射角增加相关功率和对亮温校正量呈增大趋势。

5) 观测量对亮温校正量的灵敏度与空间分辨率成反比例关系,由于空间分辨率受到风区大小的限制,高灵敏度、高分辨率反演算法的研究是星载 GNSS-R 辅助辐射计探测盐度的关键技术之一。

参考文献 (References)

[1] 刘良明. 卫星海洋遥感导航[M]. 武汉:武汉大学出版社,2010:265-266.

LIU L M. An introduction to satellite oceanic remote sensing [M]. Wuhan:Wuhan University Press,2010:265-266(in Chi-

北航学报 赠阅

nese).

[2] FONT J, LAGERLOEF G S E, VINE D M L, et al. The determination of surface salinity with the European SMOS space mission[J]. IEEE Transactions on Geoscience & Remote Sensing, 2004, 42(10): 2196-2205.

[3] VINE D M L, LAGERLOEF G S E, COLOMB F R, et al. Aquarius: An instrument to monitor sea surface salinity from space[J]. IEEE Transactions on Geoscience & Remote Sensing, 2007, 45(7): 2040-2050.

[4] LAGERLOEF G S E, SWIFT C T, LEVINE D M. Sea surface salinity: The next remote sensing challenge[J]. Oceanography, 1995, 8(2): 40-50.

[5] VALLLLOSSERA M. Determination of sea surface salinity and wind speed by L-band microwave radiometry from a fixed platform[J]. International Journal of Remote Sensing, 2004, 25(1): 111-128.

[6] MARCHAN-HERNANDEZ J F, VALENCIA E, RODRIGUEZ-ALVAREZ N, et al. Sea-state determination using GNSS-R data [J]. IEEE Geoscience & Remote Sensing Letters, 2010, 7(4): 621-625.

[7] FABRA F, CARDELLACH E, RIUS A, et al. Phase altimetry with dual polarization GNSS-R over sea ice[J]. IEEE Transactions on Geoscience & Remote Sensing, 2012, 50(6): 2112-2121.

[8] RODRIGUEZALVAREZ N, BOSCHLLUIS X, CAMPS A, et al. Soil moisture retrieval using GNSS-R techniques: Experimental results over a bare soil field[J]. IEEE Transactions on Geoscience & Remote Sensing, 2009, 47(11): 3616-3624.

[9] SOLAT F. Sea surface remote sensing with GNSS and sunlight reflections[D]. Catalunya: Universitat Politécnica de Catalunya, 2003: 149-164.

[10] 佟晓林. 提高卫星微波遥感海面盐度反演精度的方法研究[D]. 武汉:华中科技大学, 2015: 121-127.
TONG X L. Research on improving accuracy methods of ocean surface salinity in satellite microwave remote sensing[D]. Wuhan: Huazhong University of Science and Technology, 2015: 121-127(in Chinese).

[11] XAVIER B L, ADRIANO C, ISAAC R P, et al. PAU/RAD: Design and preliminary calibration results of a new L-band pseudo-correlation radiometer concept[J]. Sensors, 2008, 8(7): 4392.

[12] KLEIN L, SWIFT C. An improved model for the dielectric constant of sea water at microwave frequencies[J]. IEEE Transactions on Antennas & Propagation, 2003, 25(1): 104-111.

[13] CAMPS A, FONT J, VALL-LLOSSERA M, et al. The WISE 2000 and 2001 field experiments in support of the SMOS mission: Sea surface L-band brightness temperature observations and their application to sea surface salinity retrieval[J]. IEEE Transactions on Geoscience & Remote Sensing, 2004, 42(4): 804-823.

[14] EL-NIMRI S F. Development of an improved microwave ocean surface emissivity radiative transfer model[D]. Orlando: University of Centeral Florida, 2010: 64-65.

[15] HEJAZIN Y, JONES W L, SANTOS-GARCIA A, et al. A roughness correction for aquarius sea surface salinity using the CONAE microwave radiometer[J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2016, 8(12): 5500-5510.

[16] ZAVOROTNY V U, VORONOVICH A G. Scattering of GPS signals from the ocean with wind remote sensing application[J]. IEEE Transactions on Geoscience & Remote Sensing, 2002, 38(2): 951-964.

[17] ELFOUHAILY T, CHAPRON B, KATSAROS K, et al. A unified directional spectrum for long and short wind-driven waves [J]. Journal of Geophysical Research: Oceans, 1997, 102(C7): 15781-15796.

[18] CLARIZIA M P, RUF C S. On the spatial resolution of GNSS reflectometry[J]. IEEE Geoscience & Remote Sensing Letters, 2016, 13(8): 1064-1068.

[19] 谢刚. GPS 原理与接收机设计[M]. 北京:电子工业出版社, 2015: 242-244.
XIE G. Principles of GPS and receiver design[M]. Beijing: Publishing House of Electronics Industry, 2015: 242-244 (in Chinese).

作者简介:

王峰　男,博士研究生。主要研究方向:导航卫星信号处理及其遥感应用。

杨东凯　男,博士,教授,博士生导师。主要研究方向:GNSS 遥感应用及室内定位技术。

北京航空航天大学学报

北航学报 赠阅

Sea surface salinity determination method assisted by spaceborne GNSS-R

WANG Feng, YANG Dongkai*

(School of Electronic and Information Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: Global navigation satellite system-reflectometry (GNSS-R) is a new remote sensing way which is passive radar and could be used to improve the retrieval precision of the sea surface salinity (SSS). The model of brightness temperature and the scattering power model of GNSS-R were reviewed, and the spaceborne simulation scenario was developed in this paper. Based on those, the performance of GNSS-R aiding radiometer to determine SSS was first explored. Although it is possible to decrease the mass and power consumption of spaceborne equipment by sharing the antenna and radio frequency font-end between GNSS-R and radiometer working on the frequency of GPS L1 1 575.42 MHz, when SSS is larger than 25 psu, the sensitivity of brightness temperature to SSS reduces by about 0.1 and 0.08 K/psu for the vertical and horizontal polarization signal respectively. The distortion of reflected GPS L1 signals on the measurement of the brightness temperature was analyzed. It is found that under the condition of the simulation scenario for 1 K variation of brightness temperature, reflected GPS L1 signals introduce error less than 2.5×10^{-4} K. Subsequently, the sensitivity of the GNSS-R observable to the brightness temperature variation for the vertical and horizontal polarization signal was explored. The results show that when incidence angle increases, the sensitivity of the horizontal and vertical polarization signal show falling and rising tendency respectively. Finally, the relationship between the sensitivity of the observable to the brightness temperature variation and the spatial resolution was analyzed. The conclusion is that the study of retrieval algorithm having high accuracy and spatial resolution is crucial for spaceborne GNSS-R aiding radiometer to determine SSS.

Keywords: global navigation satellite system-reflectometry (GNSS-R); sea surface salinity (SSS); assisted determination; sensitivity; spatial resolution

Received: 2017-01-13; **Accepted**: 2017-03-31; **Published online**: 2017-05-27 17:21

URL: kns.cnki.net/kcms/detail/11.2625.V.20170527.1721.004.html

Foundation item: the Innovation Foundation of BUAA for PhD Graduates

* **Corresponding author**. E-mail: yangdongkai@sina.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0019

一种增量并行式动态图异常检测算法

韩涛，兰雨晴*，肖利民，刘艳芳

（北京航空航天大学 计算机学院，北京 100083）

摘 要：图结构异常检测可以发现金融欺诈行为、网络入侵和可疑的社交行为。针对当前检测图异常算法的计算复杂度高、不能处理大规模动态图的缺点，研究并提出了一种增量并行式的算法以便更有效地发现和检测大规模动态图中的异常。该算法使用时间滑动窗口对图进行划分，在初始化阶段选取 N 个子图，使用最小描述长度（MDL）原理并行检测正常模式和异常模式，并行迭代地检测其他子图中的正常结构和异常结构。在多个大规模图数据集上的实验结果表明，检测动态图结构异常准确率达到 96%，召回率达到 85%，运行时间减少了一个数量级。同时还讨论了滑动窗口大小和并行数量对算法运行时间的影响。

关 键 词：异常检测；增量；并行；滑动窗口；最小描述长度（MDL）原理
中图分类号：TP391
文献标识码：A **文章编号**：1001-5965(2018)01-0117-08

很多应用领域中数据之间的复杂关系均可以通过图直观地表现出来，例如互联网、社交网络和生物领域。这些真实应用中的图数据通常是大规模的，并且随着时间推进数据量不断增加。例如，在社交媒体（博客、微博和微信）和信息共享平台（YouTube 和 Flicker）中，用户之间持续的社交行为会产生大量的、持续的、相互交互的数据，而这些交互可以自然地使用动态图来表示——结点表示人、物体或其他实体，边表示实体之间的联系[1]。

传统算法很难对上述大规模动态图进行有效分析和挖掘。首先，由于图的规模庞大，导致在图的计算上的时间过长。其次，程序往往不能获取图的全部数据，而只能获取数据的一部分。例如社交网络图，通常只能通过爬虫抓取获得部分数据。另外，即使有的图规模略小一些，但是计算某些图的重要的度量值需要的处理时间非常长（例如生物细胞科学的实验）。因此，需要采用并行处理的技术提高图的处理能力。

目前大部分的研究聚焦在基于图的数据关系结构分析。目前图挖掘算法针对特定的数据集处理一类特定的图，或者把一种具体的图算法应用到不同的领域。但是，这些算法都不能很好地解决图挖掘的扩展性问题，尤其是大规模动态图的异常检测问题。例如 Facebook 拥有 8 亿用户，用户每分钟发表 50 万条评论，超过 29 万条状态更新。这样包含上亿结点、每分钟产生几十万边的图的分析和异常检测问题都没有得到很好地解决。另外，除了图的结构信息，语义信息（如标签、权重和属性等信息）也能辅助检测图异常，提高检测的准确性。

异常行为通常模仿正常的行为模式，所以越接近于正常模式的异常，越难以分辨。因此本文把与正常模式相似的近似模式称为异常模式。在基于图的表示中，异常模式通常在正常模式上进行修改，例如添加边和结点、删除边和结点或修改

收稿日期：2017-01-16；**录用日期**：2017-02-06；**网络出版时间**：2017-03-23 16:47
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170323.1647.004.html
* **通信作者**. E-mail：lanyuqing@buaa.edu.cn

引用格式：*韩涛，兰雨晴，肖利民，等. 一种增量并行式动态图异常检测算法[J]. 北京航空航天大学学报，2018，44(1)：117-124.*
HAN T，LAN Y Q，XIAO L M，et al. Incremental and parallel algorithm for anomaly detection in dynamic graphs[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(1)：117-124 (in Chinese).

北航学报 赠阅

结点属性信息。目前基于图的异常检测(Graph Based Anomaly Detection, GBAD)[2]首先使用一种基于最小描述长度的压缩算法来寻找正常模式,然后分析与正常模式相近的模式,计算其异常值,最后判断这些相近的模式是否为异常模式。虽然这个算法在很多领域被广泛应用[3],但是算法的可扩展性问题,尤其是处理百万级结点的图的算法效率问题没有得到很好地解决。大规模图随着时间不断演化,这也加剧了分析的困难——正常模式随着时间或者事件触发也会发生变化。为解决此问题,本文提出了针对大规模图数据的增量并行式异常检测的 DPADS(Dynamic Parallel Anomaly Detections)算法,该算法可以更有效的处理大规模静态以及动态流图的异常检测。

1 相关工作

根据早期的图异常检测的研究,Noble 和 Cook[4]探讨图结构的异常。他们先基于正常模式对图进行压缩,后把余下的结构定义为异常。Akoglu 等[5]处理了大规模图的异常检测问题,但是他们的目标是检测异常的结点。文献[4-5]都认为图是静态的。

一种解决大规模图的算法是把图看做边的数据流,每次处理图的一个或者多个边。研究者在异常检测领域提出一些不同的算法来处理图的边数据流。其中一种被称为"semi-streaming model"的算法,可以处理不能把所有边存储到内存的大规模图。Feigenbaum 等提出了 semi-streaming 常量近似算法,来处理无权重和带权重图的匹配问题,同时也扩展应用到二分图[6]。其他的研究则把此算法扩展,来解决不同的图问题,例如有向图的最短路径问题、使用中间的临时流来解决特定问题等[7-9]。总之,这些算法分析了可用的内存和访问整个图所需访问硬盘次数这两者之间的关系,并根据实际情况进行策略调整。

基于 k-core 结构的图结构中的模式识别和异常检测[10]通常用来进行层级结构分析,图可视化和图聚类。它探讨了常见的模式与 k-core 之间的 3 种关系并设计相关算法,并证明这些模式有效地处理大规模图数据。随时间变化的图数据多层级异常检测[11]提出一种建模和分析的框架检测标签的异常。它通过概率模型描述子图和通过相关的层级模型同时检测异常。此算法不仅能检测结构异常,也能检测标签上下文中的事件。一种图划分算法[12]有效地检测图数据流规模的异常。文献[13]归纳总结了基于图的异常检测和描述。

有些算法检测图数据流中的异常,其目标是通过分析边数的统计信息识别异常的子图簇,而不是发现图或图数据流中结构的异常。另外一些算法尝试使用基于整个图的算法[2]来发现异常子图[13],但这类算法没有解决可扩展性问题。

另外,大多数异常检测算法使用监督的学习算法,需要数据提前打好标签、做好分类,然后再训练模型,进行预测。而实际情况是大多数时候不能提前知道异常或者正常的相关信息,因此需要一种无监督的算法来检测那些看起来像正常、合法但实际上结构不同的异常模式。

2 DPADS 算法的流程及分析

DPADS 算法把静态图的异常检测算法 GBAD 和并行异常检测(Parallel Anomaly Detection, PLAD)算法扩展到大规模动态图的异常检测中,如图 1 所示,T_{i-1}、T_i、T_{i+1}为时间滑动窗口。本文定义 3 种基本类型图的异常:添加、修改和移除。添加异常是正常模式增加了结点或边。修改异常包含了一个结点或边的意外标签。移除异常的子结构比正常子结构缺少了边或结点。

DPADS 算法检测图的异常基于这样的思想:异常的子结构(或子图)是正常模式的结构变种(正常模式边和节点的增加或者缺失)。假设 $d(G_1, G_2)$表示 2 个图 G_1 和 G_2 之间的结构差异度量,计算把图 G_1 转化为 G_2 的同构图的计算量(添加、删除点与改变标签的变化数量),衡量 G_1 和 G_2 之间的差异。

定义 1 在图 G 中,正常模式 S 可由最小描述长度(Minimum Description Length, MDL)原理来判定,并最小化以下目标函数:

$$L(S,G) = \min(L(G|S) + L(S))$$

式中:G 为整个图;$L(G|S)$为使用 S 压缩 G 之后的描述长度;$L(S)$为正常模式 S 的描述长度;S 为正常模式。MDL[14]原理即要求模型的总描述长

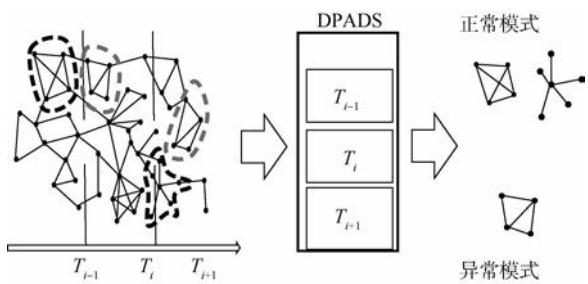


图 1 DPADS 算法并行处理子图

Fig. 1 Parallel processing subgraphs of DPADS algorithm

北航学报 赠阅

度最小。正常模式检测的 MDL 原理具有 2 个重要的性质:

1)当有 2 个模式都能很好地匹配原始图时,MDL 原理选择“最简单”的模式,即选择压缩率更高的描述。这也反映了奥卡姆剃刀对简单理论的优先选择。

2)MDL 原理是一致的。随着数据量的不断增加,中间过程产生的正常模式收敛于真正的正常模式。

定义 2　图 G_1 和 G_2 之间的差异:

$$d(G_1,G_2)=\begin{cases}0 & G_1 \text{ 与 } G_2 \text{ 同构}\\ d(G_1,G_1')+1 & G_2 \text{ 与 } G_1' \text{ 同构},G_1 \text{ 与 } G_1' \text{ 差一条边或一个节点或改变了某一个属性中的一种情况}\end{cases}$$

定义 3　一个子结构 S_A 在图 G 中为异常如果$(0<d(S_A,S)<D)$且$(0<P(S_A\mid S,G)<P)$。式中 S 为 G 中的正常模式,D 定义了异常模式 S_A 到正常模式 S 的最大差异,$P(S_A\mid S,G)$为图 G 在正常模式 S 的压缩下 S_A 为异常的概率,P 限定了 S_A 为异常的最大概率。

定义 4　在图 G 中,基于正常子结构 S 的异常子结构 S_A 的值 $R=d(S_A,S)P(S_A\mid S,G)$。根据实验分析 R 值在(0, 2.6)之间,表示 S_A 为异常的子结构可能性越大。

若找到一个正常模式 S 和与其相似的异常模式 S_A,则可迭代地找到其他的异常。先使用正常模式对原始图进行压缩,例如把正常模式替换成一个带有新标签的结点。再在压缩之后的图上寻找正常模式 S 和相关的异常 S_A,这个过程可以进行多次迭代以找到更多的正常模式和异常模式,直到遍历到整个图,当图进一步压缩时还能够进行不同级别的异常检测。

2.1　DPADS 算法流程图及删除边的处理

本文提出了 DPADS 算法,它以 n 个子图为输入,可以是静态图的切分,也可以是随时间获取的图的一部分。DPADS 主要分为 2 个阶段:初始化和迭代处理。初始化阶段的主要目标是在 n 个子图中找出正常模式 S 及其相关的异常模式。首先,并行处理 n 个子图,分别检测 top-M 个正常模式,一共得到 $n\times M$ 个正常模式。然后,判定正常模式集合 S。最后根据正常模式 S 检测异常的结构。迭代处理阶段的主要目标是迭代分析新获取的数据的结构异常。首先,把活动窗口向后移动一个窗口,让新获取的子图包含在滑动窗口内。其次,在新的子图中检测 top-M 正常模式。然后,从滑动窗口中的所有子图中判定正常模式 S'。如果 $S'=S$,只需检测新子图里的异常;否则,窗口里的每个子图都基于正常模式 S'检测异常结构。接下来,对每个异常子结构计算 R 值,值最小的子结构判定为异常结构。最后,重复迭代过程。DPADS 算法流程图如图 2 所示。

1)初始化

步骤 1　并行处理 n 个子图:

a. 每个子图发现 top-M 正常模式。

b. 每个子图等待所有子图都完成发现它们正常模式。

步骤 2　在 $n\times M$ 个正常模式中判定基准模式 S。

步骤 3　每个子图根据基准模式 S 发现异常的子结构。

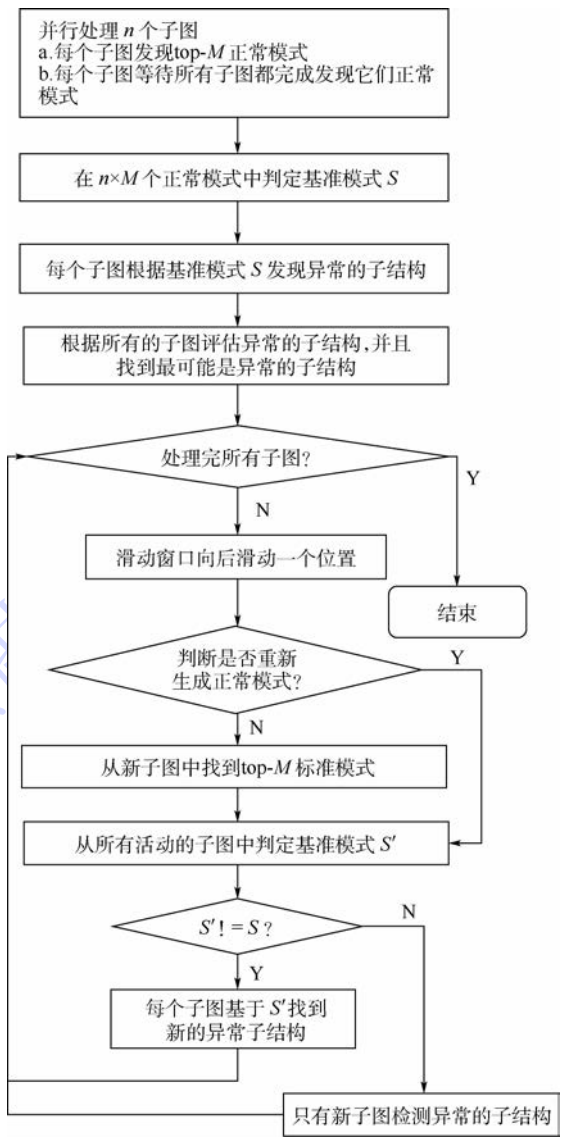


图 2　DPADS 算法流程图

Fig. 2　Flowchart of DPADS algorithm

北航学报 赠阅

步骤 4 根据所有的子图评估异常的子结构,并且找到最可能是异常的子结构。

2)迭代处理

步骤 5 处理新的子图:

a. 如果序号在最前面的子图超过了阈值 T(基于一定的标准,例如可处理的子图数量,或者子图的时间戳),那么把这些子图移除,不进行下一步分析。

b. 判断是否重新生成正常模式,若为 TRUE,则跳转到步骤 d。

c. 否则,从新子图中找到 top-M 标准模式。

d. 从所有活动的子图中判定基准模式 S'。

e. 如果(S'! $=S$),每个子图基于 S'找到新的异常子结构。

f. 否则只有新子图检测异常的子结构。

g. 评估所有子图的异常子结构,并且找到最可能是异常的子结构。

h. 重复上述过程。

PLAD 算法[15]解决了图随着时间不断插入新的结点、边和属性过程中的异常检测问题,但它并没有考虑已有的结点、边和属性随时间或环境变化而移除的过程。所以本文对以上的算法进行改进以适应动态图的情景。算法的主要思想是维护正常模式,因此可以根据正常模式检测异常。但是图随着时间动态更新会导致正常模式发生变化或者重新检测,这一过程耗时较长。当前目标是尽量少地重新检测正常模式[16]。在移除边 e 之后,需要保证 $d(S',S)<\dfrac{\beta}{(1+\varepsilon)^2}$,$\varepsilon$、$\beta$ 为参数,$\varepsilon>0$,$\beta>0$。如果不等式成立,则正常模式 S 不需要重构,返回 FALSE。否则重新检测正常模式。

2.2 算法分析

针对 DPADS 算法,描述检测正常模式的算法所需要的信息量可以分为正常模式描述长度代价 $L(S)$和图 G 的描述长度代价 $L(G|S)$。首先简要介绍用到的符号,S_i 为 S 中独立的正常模式,$n(S_i)$为 S_i 的数量,$n_i(a)$为 S_i 中属性 a 的数量,n_i 和 e_i 分别为 S_i 中的结点数和边数。

1) 正常模式描述代价

正常模式描述代价指的是描述获得正常模式算法所需要的信息量,描述复杂度由以下部分构成:

A. 结点个数 N,属性数量 L,需要的信息量为(lb N + lb L)bit。

B. 记录正常模式的个数 M,需要 lb M bit。

C. 记录在寻找正常模式过程中,对 l 个属性进行检测,需要 lb l bit。

D. 描述每个正常模式中的结点信息,需要

$$nH(p)=-\sum_{i=1}^{M}(n(S_i)\text{lb}\ p_i)$$

式中:$p_i=n(S_i)/n$;$H(p)$为香农熵理论,p 为概率。

E. 描述正常模式之间的连接关系,理论上它们之间的连接越少越好,e_{ij}表示正常模式之间的边个数。$p_{ij}=\dfrac{e_{ij}}{n_i}n_j$ 描述的代价为

$$-[e_{ij}\text{lb}\ p_{ij}+(n_in_j-e_{ij})\text{lb}(1-p_{ij})]$$

F. 描述每个正常模式中所包含的属性信息,需要

$$LH(p)=-\sum_{i=1}^{L}(n_i(a)\text{lb}\ p_i)$$

式中:$p_i=n_i(a)/L$。

2) 图描述代价

图描述代价指的是在给定正常模式情况下,对原始图进行压缩描述所需要的代价。假设正常模式中的结点和边服从 0-1 分布,且节点相互独立,则根据香农熵理论,对正常模式 S 中的元素 x 进行描述需要花费 $n(S)H(p(x))$ bit 的信息量,其中 $H(p(x))=-\sum_{x_i=0}^{e}p(x_i)\text{lb}\ p(x_i)$,$p(x)$ 为元素 x 存在的概率。

在正常模式内部,如果结点及其边和属性联系紧密,则其相应的熵值比较小,所需要的描述信息量也较小。

因此,基于 MDL 描述的图的复杂度由两部分组成:

A. 用期望来描述正常模式内部的联系,表示正常模式中结点之间的关系。正常模式 S 的结点数为 n,边数为 e,因此,$p_i=p(S=1)=e/n^2$,描述所需要的代价为

$$-[e\text{lb}\ p_i+(n^2-e)\text{lb}(1-p_i)]$$

B. 对每个正常模式中的属性信息,设每个正常模式中属性包含 a_i 的结点个数 $n(a_i)$,则 $p(a_i)=n(a_i)/n_i$,则在每个正常模式中的属性描述代价为

$$-\sum_{i=1}^{L}[n(a_i)\text{lb}\ p(a_i)+(n-n(a_i))\cdot \text{lb}(1-p(a_i))]$$

根据以上的分析,总的正常模式描述代价函数为

$$C_{\text{normal}}=C_S+C_{G|S}=\text{lb}\ N+\text{lb}\ L-\sum_{i=1}^{M}(n(S_i)\text{lb}\ p_i)-$$

北航学报 赠阅

$$\sum_{i=1}^{e}(n(a)\mathrm{lb}\ p_i) - \sum_{j=1}^{k}\sum_{i=1}^{k}[e_{ij}\mathrm{lb}\ p_{ij} + (n_i n_j - e_{ij})\mathrm{lb}(1-p_{ij})] - \sum_{i=1}^{k}[l_{ii}\mathrm{lb}\ p_{ii} + (n_i^2 - l_{ii})\mathrm{lb}(1-p_{ii})] - \sum_{i=1}^{L}[n(a_i)\mathrm{lb}\ p(a_i) + (n_i - n(a_i))\mathrm{lb}(1-p(a_i))]$$

式中:C_S 和 $C_{G|S}$ 分别为正常模式和基于最小原理用正常模式 S 描述图 G 的描述代价函数。

那么,$n \times M$ 个正常模式与其异常模式(3 种情况)的描述代价函数为

$$C_{\mathrm{total}} = nMC_{\mathrm{normal}} + 3n(S)C_{\mathrm{abnormal}} \approx (nM + 3n(S))C_{\mathrm{normal}}$$

式中:$n(S)$ 为正常模式 S 的个数;C_{abnormal} 为异常模式描述代价函数。

3　实验和评估

本节在多个真实的数据集上验证第 2 节提出的算法,包括效率、准确性和误报性。从斯坦福大学 SNAP[17] 项目中选取了 3 个数据集:无向图、有向图和有向时序图进行算法验证,数据的详细信息如表 1 所示。

表 1　实验数据集

Table 1　Experimental data sets

数据名称	类　型	结点个数	边 个 数
YouTube	无向图	1 134 890	2 987 624
LiveJ	有向图	484 751	68 993 773
Math	有向时序图	24 818	50 650

3.1　效　率

在 3 个数据集上的运行时间总体上随着数据集的大小而变化。图 3 给出本文提出算法DPADS与对比算法PLAD在运行时间t上的比较:DPADS算法在运行时间上优于现有的算法。DPADS 算法不仅处理图的增量式增加而且还解决现有的节点和边随时间减少的演化流式图的异常检测,通过提前判断正常模式的变化,减少整个异常检测过程的时间。

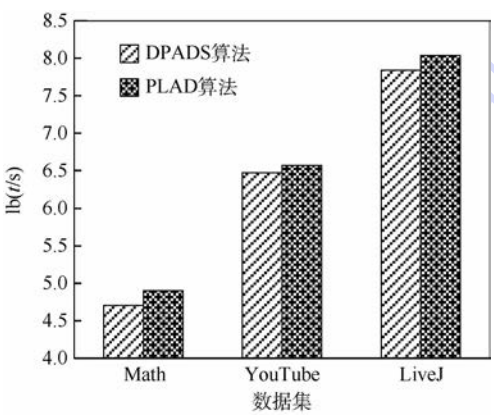


图 3　DPADS 与 PLAD 算法运行时间比较

Fig. 3　Comparison of running time between DPADS and PLAD algorithms

3.2　准确率、召回率和 ROC

异常检测的重要指标是 PR(Precision Recall)曲线和 ROC(Receiver Operating Characteristic)曲线。异常检测的准确率为检测到异常个数与总样本数之比;召回率为检测到的异常个数与总异常个数之比。从图 4 中可以看出,理想结果由空心三角表示,准确率和召回率都是 100%。本文提出的 DPADS 算法检测异常准确率平均达到 96%,即检测到的异常模式中 96% 都是异常模式;召回率的平均值为 85%,即 85% 的异常都可以检测到。由于 DPADS 算法首先对图进行划分,而且划分时只按照时间或获得边的顺序划分,并没有考虑到图的连接性和子图的相关性,所以导致准确率没有达到 100%。另外,由于动态图随时间不断变化,正常模型也随之不断变化,导致召回率没有达到 100%。

ROC 曲线是一种对于灵敏度进行描述的功能图像。ROC 曲线可以通过描述真阳性率(True Positive Rate, TPR)和假阳性率(False Positive Rate, FPR)来实现。最好的检测方式是在左上角的点,即 ROC 空间坐标轴(0,1)点,它代表 100% 灵敏(没有假阴性)和 100% 特异(没有假阳性)。完全随机的检测的任一结果对应的准确度都是 50%。如图 5 所示,虚线表示真阳性率和假阳性率都为 0.5 的随机结果,在此虚线上方的结果都比随机结果的效果好;坐标上(0,1)点的实心三角表示理想的结果,3 个实心圆点分别是 3 个数据集的 ROC 结果,分别为 0.95、0.94和 0.98。这 3 个值都接近(0,1)点,即都接近理想的结果。

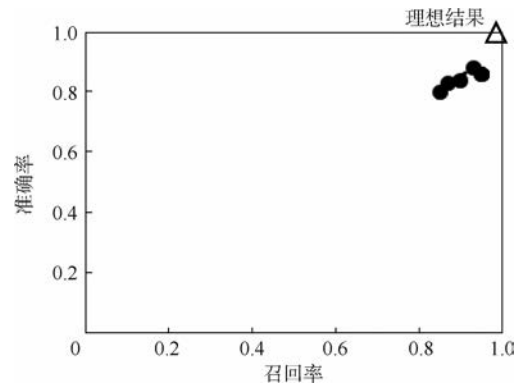


图 4　准确率和召回率

Fig. 4　Accurate rate and recall rate

3.3 M 值对准确率和召回率的影响

M 值表示每个子图中正常模式选取的个数，即 top-M。正常模式根据 $L(S)+L(G|S)$ 计算的值进行从小到大排序，前 M 个作为 top-M 的正常模式。M 是根据经验设定的固定值。如果 M 取值过小，导致每次检测没有包含足够的正常模式，则准确率和召回率下降。如果 M 取值到一定范围，不会影响算法的准确率和召回率，如图 6 所示。

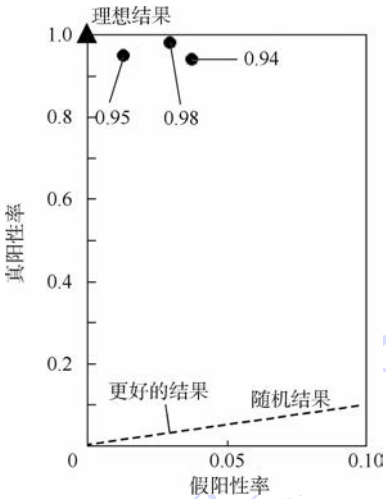


图 5 ROC 曲线

Fig. 5 ROC curve

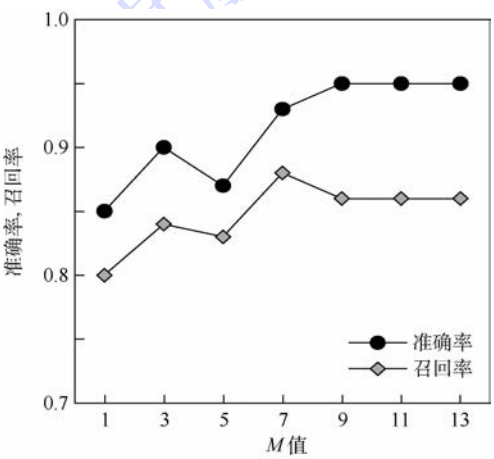


图 6 M 值对准确率和召回率的影响

Fig. 6 Influence of M value on accurate rate and recall rate

4 案例分析

本文使用图生成器产生包含两百万边的稀疏图。虽然这个图的规模没有达到“大数据”典型的上十亿的结点，但是 DPADS 算法能快速处理“大数据”规模的图。

1）静态图的异常检测算法

图包含一个特定的正常子图由 10 个结点 $(v_1,v_2,\cdots,v_{10})$ 和 9 条边 $(e_1,e_1,\cdots,e_9)$ 组成，随机产生多个级别的异常子结构插入到图中。在整个图上运行静态图异常检测算法后，异常子结构包含一个多余结点 V_W 和多余边 E_W，连接到正常模式上，如图 7 所示。但是正常模式和异常子结构花费了 76 356 s(21.21 h) 的运行时间，这在现实环境中运行时间太长，不能接受。

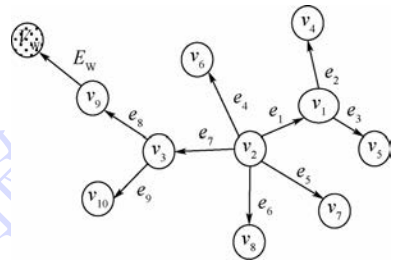


图 7 正常模式和异常模式

Fig. 7 Normal pattern and abnormal pattern

2）DPADS 算法

为了验证本文提出的增量式异常检测算法在图的流数据上的有效性，本文把 DPADS 算法应用到合成数据集上。在此实验中，原始图划分成 100 个小图，每个图包含大约 20 000 条边。初始化阶段随机选取 20(T)个小图。选取 20 个小图的目的是得到初始化的正常模式和异常模式。然后并行处理它们，花费 40 s，每个小图得到 3(M)个正常的模式。

在处理过程中即使小图采用顺序方式处理，总共花费 486 s 完成 20 个小图的正常模式检测。然后，对 60($T\times M$)个正常模式进行分析和整合，确定正常模式集 S。最后，根据正常模式 S 检测异常模式 S_A，这一步骤花费 156 s 完成。在迭代阶段，本文使用 $T=20$ 个作为处理窗口的大小（把 20 个小图都放到内存中）。然后依照算法进行多次迭代。

实验结果表明，本文算法可以应用到大规模图中。该算法在图流数据上检测到的异常模式与在整个图上检测到的异常模式是一致的，而在流数据上运行完成时间是 3 017 s（在整个图上运行静态图算法花费 76 356 s）。每个小图的运行时间如图 8 所示，含有异常模式的小图由菱形点标出，没有检测到异常的小图由圆点表示。

值得指出的是，窗口的大小（保留在内存中小图的个数）也影响了检测异常模式的时间。在检测结束时，异常的子结构被认为是异常模式。但是，如果异常模式在前面的小图出现，而窗口向后滑动，前面的小图不在窗口范围内($T=20$ 个)，那么不能检测到这样的异常结构。这符合处理流数据的“概念漂移”，因此检测到的异常模式在新

的数据到来之前变少。本文还设置不同的窗口大小并进行实验(初始过程子图的个数 T),可以发现:

1) $T=1$ 个时,最新的小图总是检测到当前的异常子模式(如果存在)。

2) $T=5$ 个时,由于插入的异常模式比较少,窗口中仅有一个小图包含一个异常子结构,使得新的子图总是包含当前的最可能为异常的子结构。

3) $T=10$ 个时,当异常子结构在小图 85 发现的时候,它的异常值低于之前的异常子结构的异常值,所以变成了当前的异常模式(当 $T=15$ 个时,结果相同)。

除了准确率之外,运行时间也受到窗口大小 T 的影响,如图 9 所示。由于并行处理的小图增加,结果显示运行的总体时间随着窗口大小 T 的增加而减少。

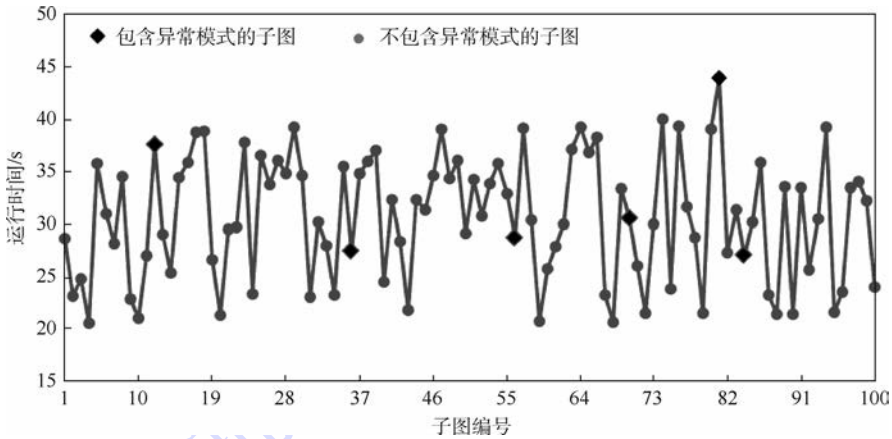


图 8　DPADS 算法在每个子图上的运行时间并标记发现异常的子图

Fig. 8　Running time of DPADS algorithm in each subgraph and marked subgraphs with abnormal pattern

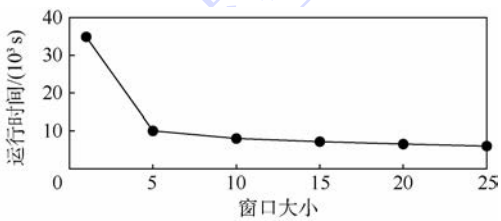


图 9　DPADS 算法在合成数据集上随窗口大小变化的运行时间

Fig. 9　Running time of DPADS algorithm on synthetic datasets with different windows size

5　结　论

本文提出大规模动态图的异常检测算法 DPADS,贡献如下:

1) 算法大大缩短了运行时间,例如处理结点数为一百万,边数量为三百万的动态图的运行时间为 5 h,而其他算法运行时间为 21 h。

2) 算法可实现较为优异的检测性能,例如处理边数量约七千万动态图数据检测异常结构准确率达到 96%,召回率达到 85%。

3) 滑动窗口大小约束了并行处理子图的数量,滑动窗口越大,算法运行时间越少。

4) 动态图不仅由结点和边随时间不断增加,而且还会不断减少。算法先增量并行处理大规模动态图的增加模式,再在此基础上处理边和结点减少的情况。

为使本文提出的算法进一步提高准确率和召回率,仍需要优化正常模式的检测和异常模式的参数。

参考文献 (References)

[1] AHMED N K, NEVILLE J, KOMPELLA R. Network sampling: From static to streaming graphs [J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2014, 8(2): 7: 1-7: 56.

[2] EBERLE W, HOLDER L. Anomaly detection in data represented as graphs[J]. Intelligent Data Analysis, 2007, 11(6): 663-689.

[3] EBERLE W, HOLDER L, GRAVES J. Insider threat detection using a graph-based approach[J]. Journal of Applied Security Research, 2011, 6(1): 32-81.

[4] NOBLE C C, COOK D J. Graph-based anomaly detection[C]// Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2003: 631-636.

[5] AKOGLU L, MCGLHON M, FALOUSTSOS C. OddBall: Spotting anomalies in weighted graphs [C] // Proceedings of the 14th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining. Berlin: Springer-Verlag, 2010, 3: 410-421.

[6] FEIGENBAUM J, KANNAN S, MCGREGOR A, et al. On graph problems in a semi-streaming model[J]. Theoretical Computer Science, 2005, 348(2-3): 207-216.

[7] DEMETRESCU C, FINOCCHI I, RIBICHINI A. Trading off

space for passes in graph streaming problems[J]. ACM Transactions on Algorithms(TALG),2009,6(1):6:1-6:17.

[8] AGGARWAL G,DATAR M,RAJAGOPALAN S,et al. On the streaming model augmented with a sorting primitive[C]// Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science(FOCS). Washington,D. C.:IEEE Computer Society,2004:540-549.

[9] SARMA A,GOLLAPUDI S,PANIGRAHY R. Estimating PageRank on graph streams[C]// Proceedings of the 27th ACM Sigmod-Sigact-Sigart Symposium on Principles of Database Systems. New York:ACM Press,2008:69-78.

[10] SHIN K,ELIASSI-RAD T,FALOUTSOS C. CoreScope:Graph mining using k-core analysis-Patterns,anomalies and algorithms[C]//2016 IEEE 16th International Conference on Data Mining (ICDM). Washington,D. C.:IEEE Computer Society,2017:469-478.

[11] BRIDGES R A,COLLINS J P,FERRAGUT E M,et al. Multi-level anomaly detection on time-varying graph data[C]//2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). New York:ACM Press,2016:579-583.

[12] EBERLE W,HOLDER L. A partitioning approach to scaling anomaly detection in graph streams[C]// 2014 IEEE International Conference on Big Data. Washington,D. C.:IEEE Computer Society,2014:17-24.

[13] AKOGLU L,TONG H,KOUTRA D. Graph based anomaly detection and description:A survey[J]. Data Mining and Knowledge Discovery,2015,29(3):626-688.

[14] 吴烨,钟志农,熊伟,等. 一种高效的属性图聚类算法[J]. 计算机学报,2013,36(8):1704-1713.
WU Y,ZHONG Z N,XIONG W,et al. An efficient method for attributed graph clustering[J]. Chinese Journal of Computers,2013,36(8):1704-1713(in Chinese).

[15] EBERLE W,HOLDER L. Incremental anomaly detection in graphs[C]//2013 IEEE 13th International Conference on Data Mining Workshops. Washington,D. C.:IEEE Computer Society,2013:521-528.

[16] EPASTO A,LATTANZI S,SOZIO M. Efficient densest subgraph computation in evolving graphs[C]// Proceedings of the 24th International Conference on World Wide Web. Geneva:International World Wide Web Conferences Steering Committee,2015:300-310.

[17] YANG J,LESKOVEC J. Defining and evaluating network communities based on ground-truth[C]// Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. New York:ACM Press,2012,3:1-3:8.

作者简介:

韩涛 女,博士研究生。主要研究方向:社交网络、数据挖掘、大数据。

兰雨晴 男,博士,副教授,硕士生导师。主要研究方向:操作系统、大数据、数据安全。

肖利民 男,博士,教授,博士生导师。主要研究方向:高性能计算机系统、大数据。

刘艳芳 女,博士研究生。主要研究方向:可信计算、软件自动化测试、大数据。

Incremental and parallel algorithm for anomaly detection in dynamic graphs

HAN Tao, LAN Yuqing*, XIAO Limin, LIU Yanfang

(School of Computer Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: Financial fraud behavior, network intrusion and suspicious social actions can be detected by structural anomaly detection in graphs. The existing anomaly detection algorithms require high computational complexity and cannot process large-scale dynamic graphs. So an incremental and parallel algorithm is proposed to discover and detect abnormal patterns in dynamic graphs effectively and efficiently. The whole graph was partitioned into subgraphs by time sliding windows. *N* subgraphs in time sliding windows were processed in parallel by minimum description length (MDL) principle to discover both normal and abnormal patterns. Structural outliers can be detected gradually in parallel based on normal patterns. The results of experiments conducted in multiple large-scale graphs show that the precision rate for detecting the abnormal patterns of dynamic graph reaches 96%, recall rate reaches 85%, and running time reduces by an order of magnitude. The impact of the size of sliding windows and the number of parallel on running time of the algorithm is also discussed.

Keywords: anomaly detection; incremental; parallel; sliding window; minimum description length (MDL) principle

Received: 2017-01-16; **Accepted**: 2017-02-06; **Published online**: 2017-03-23 16:47

URL: kns.cnki.net/kcms/detail/11.2625.V.20170323.1647.004.html

* **Corresponding author**. E-mail: lanyuqing@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0023

GNSS 海面反射信号的三维建模方法

祁永强，张波，杨东凯*，张彦仲，张健敏

（北京航空航天大学 电子信息工程学院，北京 100083）

摘 要：全球导航卫星系统反射(GNSS-R)技术应用中需 GNSS-R 信号模拟器来测试反射信号接收机，以降低成本。为此，提出了一种基于双基雷达原理的全球导航卫星系统(GNSS)海面反射信号建模方法。首先，分析了 GNSS-R 双基雷达遥感原理，根据延迟和多普勒频率在海面的分布特点，选择海面的反射点，并计算相应反射单元的面积；然后，对散射系数进行了计算；最后，对多条反射信号的合路信号进行相关的仿真验证。验证结果表明：模拟的海面反射信号的相关功率曲线与 ZV 模型理论曲线的相关系数优于 0.92，能够有效地用于 GNSS 海面反射信号的生成。

关 键 词：全球导航卫星系统(GNSS)；海洋遥感；反射信号；双基雷达；ZV 模型
中图分类号：TN967.1
文献标识码：A **文章编号**：1001-5965(2018)01-0125-07

全球导航卫星系统反射(GNSS-R)已成为一种新兴的海洋遥感技术[1]。该技术以全球导航卫星系统(GNSS)卫星为信号源，采用双基雷达前向散射模式进行海洋信息的探测[2]。具有信号源稳定、覆盖范围广、成本低等优点[3]，近 20 年得到了国内外学者的广泛关注。

美国、欧洲以及中国都投入了很多精力对 GNSS-R 海洋遥感技术进行研究，并开展了多次有关的地基、机载和星载试验，取得了大量的研究成果[4-7]。鉴于 GNSS-R 的许多试验耗费大、可重复性差的特点，需要研制可在实验室测试的 GNSS-R 信号模拟器。

目前对于 GNSS-R 信号模拟器的研制处于起步阶段，可用于反射信号模拟器的信号模型还在探索中。本文根据 GNSS-R 双基雷达遥感原理，分析了海面反射信号的空间域和散射系数的计算，以此建立起反射信号的模型，并对计算的信号参数进行了相关验证。

1 GNSS-R 双基雷达遥感原理

GNSS-R 海洋遥感的原理，是用陆基、机载或星载接收机，接收经海面反射的 GNSS 信号，可以看作为一种收发分置的 L 波段雷达系统。基于微波信号散射理论，尤其是利用双基地雷达方程，可分析海面反射信号与 GNSS 直接信号在频率、相位、强度等参数之间的变化，从而实现海面的微波遥感探测。这也是 GNSS 海面反射信号建模的理论基础。

图 1 为 GNSS-R 海洋遥感原理示意图。GNSS 卫星与海面反射区域、机载接收机形成了收发分置的双基雷达系统。接收机通常需要 2 副天线，一副天线朝上，用来接收直射信号；另一副天线向下，用来接收海面反射信号，以测量 GNSS 反射信号的功率来得到海面特征信息。海面反射区域可以看作大量独立反射元的集合，那么接收机接收到的海面反射信号就可以看作每个反射元的反射

收稿日期：2017-01-16；**录用日期**：2017-02-15；**网络出版时间**：2017-04-28 15:12
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170428.1512.001.html
基金项目：国家自然科学基金（61171070）
* **通信作者**. E-mail：yangdongkai@sina.com

引用格式：祁永强，张波，杨东凯，等. GNSS 海面反射信号的三维建模方法[J]. 北京航空航天大学学报，2018，44(1)：125-131.
QI Y Q，ZHANG B，YANG D K，et al. Three-dimensional modeling method of GNSS sea surface reflection signal[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(1)：125-131 (in Chinese).

北航学报 赠阅

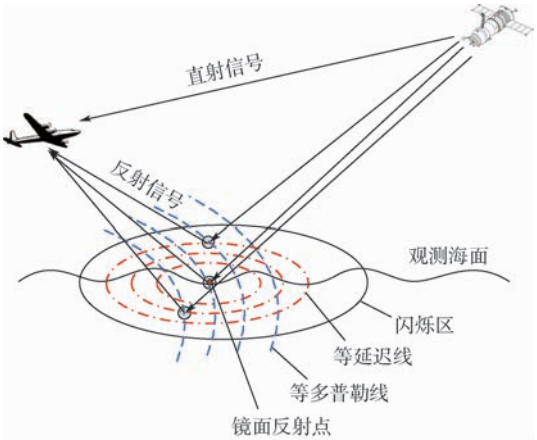


图 1 GNSS-R 海洋遥感示意图

Fig. 1 Schematic diagram of GNSS-R ocean remote sensing

信号的叠加。为了便于分析,接收到的海面反射信号可表达为

$$s_r(t) = \sum_k A_k p(t - \tau_k)\sin(2\pi(f + f_k)t + \varphi_k) \quad (1)$$

式中:A_k 为信号幅值;$p(t)$ 为值为 ±1 的数据码与伪码的异或和;τ_k 为信号时延;f_k 为多普勒频率;t 为时间;f 为载波频率;φ_k 为初始相位。

不考虑信号与海面的相互作用时,信号时延和多普勒频率都可以在海面准确映射[8]。也就是说,信号在海面上的基本时间以及频率特性是较容易预知的,但获知反射后的信号功率是比较难的。

设卫星发射功率为 P_t,发射天线功率增益为 G_t,发射信号波长为 λ,接收天线功率增益为 G_r,卫星到海面反射点的距离为 R_t,接收机到海面反射点的距离为 R_r,双基雷达截面积为 σ,那么海面上某个反射点的信号功率可表示为

$$P_r = \frac{P_t G_t G_r \lambda^2 \sigma}{(4\pi)^3 R_t^2 R_r^2} \quad (2)$$

从式(2)可以看出,关键是雷达截面积的计算。σ 的计算公式为

$$\sigma = \sigma_0 A \quad (3)$$

式中:σ_0 为标准化的双基散射系数;A 为反射元对应的面积。

2 海面反射信号的空间域分析

2.1 延迟和多普勒频率在海面的分布

在研究中,引入镜面反射点概念,即从反射区反射的反射信号中路径最短的理论反射点。接收机接收的海面反射信号主要来自镜面反射点周围的闪烁区。GNSS 卫星的信号经海面反射后将在闪烁区上形成一系列的等延迟线和等多普勒线,如图2所示,X方向为平行于载机的飞行方向,

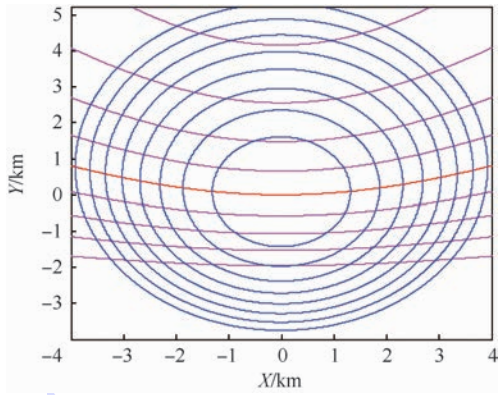


图 2 延迟/等多普勒线

Fig. 2 Iso-delay/Doppler lines

Y 方向为垂直于载机的飞行方向。

根据扩频信号接收原理可知,反射信号不同的传播时间表现为不同的码延迟,等延迟线就是与镜面点反射信号相比具有相同时间延迟的点组成的曲线。其形状为椭圆,相邻 2 个椭圆之间的时间间隔可根据需要而定。等延迟线的形状和大小与接收机的高度、GNSS 卫星的高度角和方位角等因素有关[9]。

等延迟线的椭圆方程形式[10]为

$$\frac{(y - y_0)^2}{a^2} + \frac{x^2}{b^2} = 1 \quad (4)$$

式中:

$$y_0 = \delta\frac{\cos\gamma}{\sin^2\gamma} \quad (5)$$

$$a = \frac{\sqrt{\delta(\delta + 2h\sin\gamma)}}{\sin^2\gamma} \approx \sqrt{\frac{2\delta h}{\sin^2\gamma}} \quad (6)$$

$$b = a\sin\gamma \approx \sqrt{\frac{2\delta h}{\sin\gamma}} \quad (7)$$

其中:δ 为任意散射路径(发射端—任意散射点—接收机)与镜面反射路径(发射端—镜面反射点—接收机)之间的路径差;h 为接收机的高度;γ 为卫星仰角。

通常接收机和卫星存在相对运动,接收机接收到的海面不同反射点的反射信号可能具有不同的多普勒频率,且随着卫星参数和接收平台运动参数的变化而变化。具有相同多普勒频率的海面反射点组成的曲线称为等多普勒线。其形状是二次曲线,对称轴是接收机运动速度在海面上的投影。

等多普勒线的二次方程形式[10]为

$$\begin{aligned}(V_D^2 - V_{rx}^2)x^2 + (V_D^2 - V_{ry}^2)y^2 - 2V_{rx}V_{ry}xy + \\ 2hV_{rx}(V_{rz} - V_{ry}\cot\gamma)x + 2h[V_{ry}V_{rz} + \\ (V_D^2 - V_{ry}^2)\cot\gamma]y = h^2[(V_{ry}\cot\gamma - \\ V_{rz}) - V_D^2\csc^2\gamma]\end{aligned} \quad (8)$$

北航学报 赠阅

式中：V_{rx}、V_{ry}、V_{rz}分别为接收机在 X、Y 和 Z 3 个方向的速度；V_D 定义为

$$V_D = \lambda f_{D_0} + V_{ty}\cos\gamma + V_{tz}\sin\gamma \tag{9}$$

其中：f_{D_0}为反射信号的多普勒频率；V_{ty}、V_{tz}分别为卫星在 Y 和 Z 方向的速度。

以 GPS 卫星 C/A 码信号为例，设卫星高度为 20 200 km，仰角为 60°，速度为（2 728，－1 048，1 131）m/s，接收机高度为 5 km，速度为（0，120，0）m/s，等延迟线和等多普勒线如图 2 所示。其中，镜面反射点是（0，0）点。等延迟线是椭圆线，每个椭圆延迟1/2 chip。等多普勒线是弧线，相邻 2 条弧线间隔为 50 Hz。

2.2　海面反射点的选取

GNSS-R 接收机接收到的海面反射信号是海面一系列散射点反射信号的集合。这就涉及到反射点选取的问题。反射点数越多，计算量就越大，增加实现难度；但是反射点数太少了，就可能不能够正确反映海面特性信息。如何科学合理地选取反射点，是反射信号建模的一个关键。

选取海面反射点可根据信号在海面的分布情况来选取其特征点来完成。镜面反射点由其定义可知是唯一的，也是闪烁区的中心点，如图 3 中的 O 点。镜面反射点通常作为测量和建模应用的主要参考点，其重要性不言而喻。另一类点是等延迟线和等多普勒线的交点，通常有 2 个交叉点，如图 3 中的 A 点和 B 点。通过这些点，才能建立起延迟和多普勒频率之间的拓扑关系。还有一类点是等延迟线和等多普勒线的切点，是椭圆与二次曲线的一个平滑过渡点，如图 3 中的 C 点。切点也是反射点中不可或缺的组成部分。因此，特征点包括中心点、交点和切点 3 种点。

每个反射点的信号基本上可用功率、传播延迟和载波相位变化 3 个参量来完整描述[11]。也就是说，信号建模实际上是在时频域上描述信号。而海面反射点的选取是在空间域进行的，因此有必要分析信号在空间域和时频域的关系，如图 3 所示，f_d 为多普勒频率，τ 为时延。

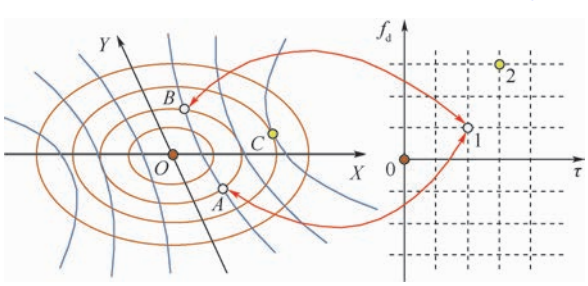


图 3　反射信号空间域和时频域的关系

Fig. 3　Relationship between spatial domain and time frequency domain of reflection signal

为简明起见，设空间域中的点 $\{O, A, B, C\}$ 组成集合 S，时频域中的点 $\{0, 1, 2\}$ 组合集合 T，那么集合 S 到集合 T 的映射如图 4 所示。

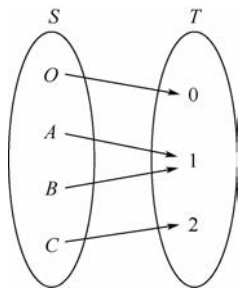


图 4　反射点的映射

Fig. 4　Reflection point mapping

该映射的对应法则是“电磁波的叠加原理”。设 A 点的信号功率为 P_A，B 点的信号功率为 P_B，点 1 的信号功率为 P_1，则

$$P_1 = P_A + P_B \tag{10}$$

通过这种映射就可建立起反射信号在空间域与时频域中的对应关系，为反射信号的选点和建模奠定了理论基础。

GNSS-R 海洋遥感是利用反射信号的时延多普勒相关功率来反推海洋表面信息的。作为基本的观测量，时延多普勒相关功率的波形是对称的。为保证接收到的海面反射信号相关功率的波形，特征点的选取应满足 2 个要求：①特征点的分布是较为对称且相对均匀的；②特征点的数量是足够的，且基本无冗余点。

基于以上考虑，本文根据 3 种点的特征分别来选取。中心点是唯一的，也是很明确的。对于交点，根据延迟和多普勒频率在海面的分布，设置延迟和多普勒频率的间隔，并选取合适的点。延迟间隔设定为 1/4 chip，共选取 14 个等延迟环，从内到外依次设为 $c_1, c_2, \cdots, c_{14}$。根据等多普勒线在海面分布的特点，以过镜面反射点的等多普勒线 l_0 为界，在其上方选取间隔为 15 Hz 的 14 条等多普勒线，从下到上依次设为 $l_1, l_3, \cdots, l_{27}$。在 l_0 下方选取间隔为 25 Hz 的 14 条等多普勒线，从上到下依次设为 $l_2, l_4, \cdots, l_{28}$。然后从 l_0 上方选取等延迟环和等多普勒线的交叉点，即依次是 c_1 和 l_1 的交点，c_2 和 l_3 的交点，$\cdots$，c_{14}和 l_{27}的交点。从 l_0 下方选取等延迟环和等多普勒线的交叉点，即依次是 c_1 和 l_2 的交点，c_2 和 l_4 的交点，$\cdots$，c_{14}和 l_{28}的交点。对于切点，在空间域中等多普勒线的对称轴上，在不同的等延迟环上获得。最终在海面上选取的反射点如图 5 所示。

北航学报 赠阅

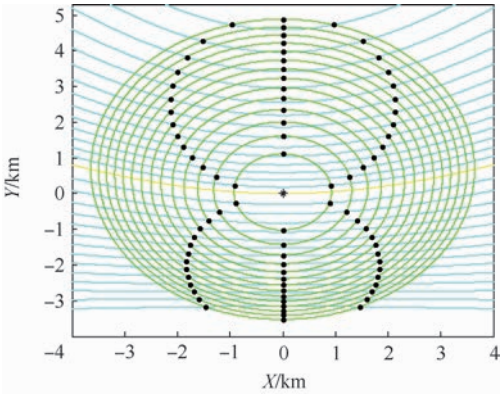


图 5 海面反射点选取

Fig. 5 Selection of sea surface reflection points

2.3 反射点对应单元的面积计算

反射单元的面积是影响反射信号功率的另一个重要因素。考虑到信号辐射的特点，每个反射点所对应的单元区域设为圆形。

设椭圆 c_1 的长轴为 a，短轴为 b，那么镜点反射单元的圆形半径为

$$r_0 = \alpha_0(a - \sqrt{a^2 - b^2}) \tag{11}$$

式中：α_0 为选择系数。

设椭圆 c_1 上的某个反射点 P 坐标为 (x_1, y_1)，点 O 到点 P 的射线 OP 斜率为

$$k = y_1/x_1 \tag{12}$$

可得射线 OP 与椭圆 c_2 的交点 Q，坐标设为 (x_2, y_2)。那么点 P 反射单元的圆形半径为

$$r = \alpha\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{13}$$

式中：α 为选择系数。为了每个反射单元的面积不重叠，选择系数应小于 0.5。设镜点反射单元的选择系数 $\alpha_0 = 0.4$，弧线 l_0 上方反射单元的选择系数 $\alpha = 0.25$，弧线 l_0 下方反射单元的选择系数 $\alpha = 0.18$。各个反射单元的散射面积如图 6 所示。

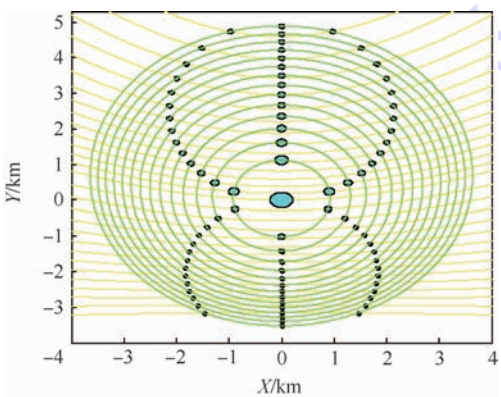


图 6 反射单元的面积

Fig. 6 Area of reflection unit

3 散射系数计算

散射系数是指单位面积上的雷达散射截面，是入射电磁波与海洋表面相互作用结果的度量。设海面的菲涅耳反射系数为 $\Re$，散射向量为 $\boldsymbol{q}$（q_x、q_y、q_z 为 $\boldsymbol{q}$ 的分量），则标准化的散射系数[12]可表示为

$$\sigma_0 = \pi|\Re|^2\left(\frac{\boldsymbol{q}}{q_z}\right)^4 P\left(-\frac{q_x}{q_z}, -\frac{q_y}{q_z}\right) \tag{14}$$

GNSS 信号经过海面反射后，信号极化由右旋圆极化变为左旋圆极化。其菲涅耳反射系数[13]为

$$\Re = \Re_{RL} = \frac{1}{2}(\Re_V - \Re_H) \tag{15}$$

其中：下标“R”、“L”、“V”、“H”分别代表右旋圆极化、左旋圆极化、垂直线极化和水平极化。运用菲涅耳等式可得到垂直极化和水平极化的菲涅耳反射系数为

$$\Re_V = \frac{\varepsilon\sin\theta - \sqrt{\varepsilon - \cos^2\theta}}{\varepsilon\sin\theta + \sqrt{\varepsilon - \cos^2\theta}} \tag{16}$$

$$\Re_H = \frac{\sin\theta - \sqrt{\varepsilon - \cos^2\theta}}{\sin\theta + \sqrt{\varepsilon - \cos^2\theta}} \tag{17}$$

式中：ε 为海面的复介电常数；θ 为卫星的高度角。

式(18)中的 P_s 为海面倾斜联合概率密度函数[14]。若海面服从高斯分布，在二维情况下，其表达式为

$$P_s(s_x, s_y) = \frac{1}{2\pi\sigma_{sx}\sigma_{sy}\sqrt{1 - b_{x,y}^2}}\exp\left(-\frac{1}{2(1 - b_{x,y}^2)}\right)\cdot\left(\frac{s_x^2}{\sigma_{sx}^2} - 2b_{x,y}\frac{s_x s_y}{\sigma_{sx}\sigma_{sy}} + \frac{s_y^2}{\sigma_{sy}^2}\right) \tag{18}$$

式中：$s_x = -q_x/|\boldsymbol{q}|$，$s_y = -q_y/|\boldsymbol{q}|$ 分别为沿 X 和 Y 方向的海面倾斜度；σ_{sx}^2、σ_{sy}^2 分别为海面沿 X 和 Y 方向的均方倾斜度；$b_{x,y}$ 为 s_x、s_y 的相关系数。海面倾斜方差和相关系数与海面风场之间可以通过波浪谱建立如下关系：

$$\sigma_{sx,sy}^2 = \langle s_{x,y}^2\rangle = \int_{-\infty}^{k_0}\int_{-\infty}^{k_0} k_{x,y}^2\psi(k_x, k_y)\,\mathrm{d}k_x\mathrm{d}k_y \tag{19}$$

$$b_{x,y} = \langle s_x, s_y\rangle/\sigma_{sx}\sigma_{sy} \tag{20}$$

$$\langle s_x, s_y\rangle = \int_{-\infty}^{k_0}\int_{-\infty}^{k_0} k_x k_y\psi(k_x, k_y)\,\mathrm{d}k_x\mathrm{d}k_y \tag{21}$$

其中：$\sigma_{sx,sy}^2$ 为海面倾斜方差；$s_{x,y}^2$ 为 s_x 和 s_y 的二维联合密度函数；$k_{x,y}$ 为 k_x 和 k_y 的合成波数，k_x 和 k_y 为海浪的波数；k_0 为截止波数；$\psi(k_x, k_y)$ 为海浪谱函数。若风向沿 X 方向或者 Y 方向，则 $b_{x,y} = 0$，否则 $b_{x,y} \neq 0$。设海浪谱主要是由海风产生的，

$\psi(k_x,k_y)$可通过 Elfouhaily 模型[15]来得到。

设海温为 25 ℃,盐度为 2%,海面风速为 5 m/s,风向为 0°,波龄为 0.84,则可求得闪烁区的散射系数,如图 7 所示。

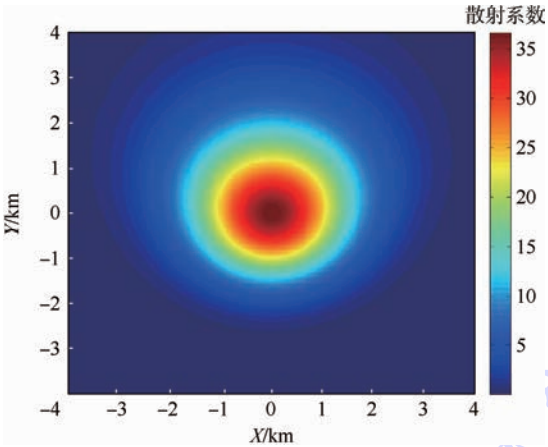


图 7　闪烁区的散射系数

Fig. 7　Scattering coefficient of glistening zone

4　海面反射信号的相关处理

本节对由这些反射信号组成的海反信号进行相关处理,以此来验证所取的海面反射点是否恰当以及计算的反射信号功率是否正确。按照文献[16]中的处理结构图,对模拟的海反信号进行相关处理仿真。在仿真中,设本地伪码间隔 0.25 chip,共 64 个相关器,本地载波间隔 20 Hz,范围是 −400 ~ 400 Hz。这样,得到的海反信号相关功率如图 8 所示。

在利用 GNSS 海面散射信号进行海面风场遥感的研究中,Zavorotny 和 Voronovich[17]建立了较为成熟的 GNSS 海面散射信号相关功率模型,即 ZV 模型。其时延一维相关功率是指在某个特定的多普勒频率下的相关功率,表达式如下:

$$\langle |Y(\tau,f_0)|^2\rangle = T_i^2\int \frac{D^2(\boldsymbol{\rho})\Lambda^2[\tau-(R_t+R_r)/c]}{4\pi R_t^2R_r^2}\cdot |S(f_0(\boldsymbol{\rho})-f_c)|^2\sigma_0'(\boldsymbol{\rho})\,\mathrm{d}^2\boldsymbol{\rho} \tag{22}$$

式中:T_i 为相干积分时间;D 为接收天线的增益;Λ 为 GNSS 伪码的自相关函数;S 为多普勒滤波函数;f_c 为载波中心频率;σ_0' 为海面的标准化散射截面;$\boldsymbol{\rho}$ 为镜面反射点到平均海面上某一点的向量;R_t 和 R_r 分别为卫星和接收机到 $\boldsymbol{\rho}$ 端点的距离。

利用 ZV 模型在相同的条件下即可得到理论的时延多普勒二维相关功率波形,如图 8(b)和图 8(d)所示。由图 8 可以看出两者的近似度很高,其相关系数为 0.925 5。

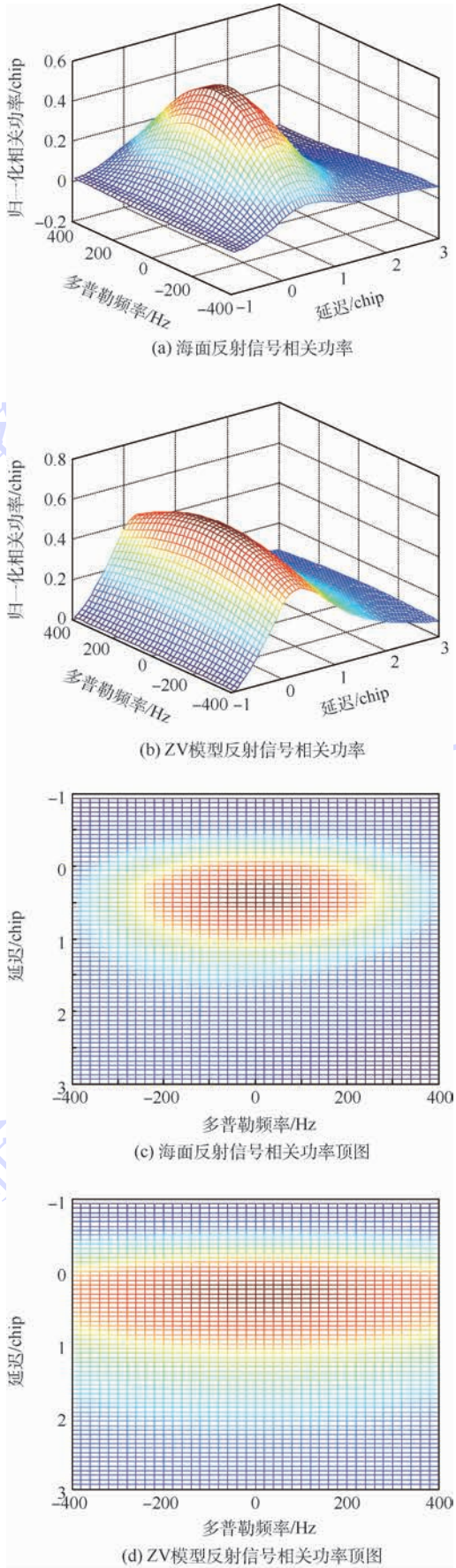


图 8　海面及 ZV 模型的反射信号相关功率及其顶图

Fig. 8　Reflection signal correlation power and its top view from sea surface and ZV model

5 结 论

本文建模分析和仿真可得到：

1）反射点的选取不仅符合时延多普勒海面分布的特点，又考虑到相关功率的对称性，且基本无冗余点。

2）反射单元面积的计算方法是科学合理的，使得能够应用双基雷达方程来计算GNSS海面反射信号的功率。

3）通过ZV模型和反射信号的相关处理的对比，验证了该建模方法的可行性和有效性。

参考文献（References）

[1] ZUFFADA C, LI Z J, NGHIEM S V, et al. The rise of GNSS reflectometry for earth remote sensing[C]//2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Piscataway, NJ: IEEE Press, 2015: 5111-5114.

[2] SHAH R, GARRISON J L, GRANT M S. Demonstration of bistatic radar for ocean remote sensing using communication satellite signals[J]. IEEE Geoscience and Remote Sensing Letters, 2012, 9(4): 619-623.

[3] MASHBURN J, AXELRAD P, LOWE S T, et al. An assessment of the precision and accuracy of altimetry retrievals for a Monterey Bay GNSS-R experiment[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016, 9(10): 4660-4668.

[4] SOISUVARN S, JELENAK Z, SAID F, et al. The GNSS reflectometry response to the ocean surface winds and waves[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016, 9(10): 4678-4699.

[5] VALENCIA E, ZAVOROTNY V U, AKOS D M, et al. Using DDM asymmetry metrics for wind direction retrieval from GPS ocean-scattered signals in airborne experiments [J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(7): 3924-3936.

[6] ZHANG Y, TIAN L M, MENG W T, et al. Feasibility of code-level altimetry using coastal BeiDou reflection (BeiDou-R) setups[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(8): 4130-4140.

[7] CARRENO-LUENGO H, PARK H, CAMPS A, et al. Submeter ocean altimetry with GPS L1 C/A signal[C]//2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Piscataway, NJ: IEEE Press, 2012: 7071-7074.

[8] GLEASON S, GEBRE-EGZIABHER D. GNSS applications and methods[M]. Norwood: Artech House, 2009: 399-433.

[9] 杨东凯，张其善. GNSS反射信号处理基础与实践[M]. 北京：电子工业出版社，2012：165-174.
YANG D K, ZHANG Q S. GNSS reflected signal processing: Fundamentals and applications[M]. Beijing: Publishing House of Electronics Industry, 2012: 165-174 (in Chinese).

[10] 白永星. 利用GNSS-R信号反演土壤湿度的关键技术研究[D]. 北京：北京航空航天大学，2013.
BAI Y X. Study on soil moisture remote sensing using GNSS-R signals[D]. Beijing: Beihang University, 2013 (in Chinese).

[11] 谢钢. 全球导航卫星系统原理——GPS、格洛纳斯和伽利略系统[M]. 北京：电子工业出版社，2013：288-289.
XIE G. Principles of GNSS: GPS, GLONASS, and Galileo[M]. Beijing: Publishing House of Electronics Industry, 2013: 288-289 (in Chinese).

[12] 祁永强，张波，杨东凯，等. 基于双基雷达原理的GNSS海面反射信号建模方法[J]. 北京航空航天大学学报，2017，43(8)：1610-1615.
QI Y Q, ZHANG B, YANG D K, et al. GNSS sea surface reflection signal modeling method based on the principle of bistatic radar[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(8): 1610-1615 (in Chinese).

[13] MAURICE W L. Radar reflectivity of land and sea[M]. Boston: Artech House, 2001: 65-94.

[14] ELFOUHAILY T, THOMPSON D R, LINDSTROM L. Delay-Doppler analysis of bistatical reflected signals from the ocean surface: Theory and application[J]. IEEE Transactions on Geoscience and Remote Sensing, 2002, 40(3): 560-573.

[15] ELFOUHAILY T, CHAPRON B, KATSAROS K, et al. A unified directional spectrum for and short wind- driven waves[J]. Journal of Geophysical Research: Oceans, 1997, 104(C7): 15781-15796.

[16] 杨东凯，丁文锐，张其善. 软件定义的GNSS反射信号接收机设计[J]. 北京航空航天大学学报，2009，35(9)：1048-1051.
YANG D K, DING W R, ZHANG Q S. Software defined GNSS reflections receiver design[J]. Journal of Beijing University of Aeronautics and Astronautics, 2009, 35(9): 1048-1051 (in Chinese).

[17] ZAVOROTNY V, VORONOVICH A. Scattering of GPS signals from the ocean with wind remote sensing application[J]. IEEE Transactions on Geoscience and Remote Sensing, 2000, 38(2): 951-964.

作者简介：

祁永强 男，博士研究生。主要研究方向：GNSS-R理论与应用。

张波 男，博士，讲师，硕士生导师。主要研究方向：扩频通信、卫星导航及其应用等。

杨东凯 男，博士，教授，博士生导师。主要研究方向：卫星导航信号处理及应用技术、无线数据传输等。

张彦仲 男，院士，博士生导师。主要研究方向：近代航空电子系统。

张健敏 女，硕士研究生。主要研究方向：室内定位方法。

北航学报 赠阅

Three-dimensional modeling method of GNSS sea surface reflection signal

QI Yongqiang, ZHANG Bo, YANG Dongkai*, ZHANG Yanzhong, ZHANG Jianmin

(School of Electronic and Information Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: In the application of global navigation satellite system-reflection (GNSS-R) technology, the GNSS-R signal simulator is needed to test the reflection signal receiver in order to reduce costs. A modeling method of global navigation satellite system (GNSS) sea surface reflection signal based on the principle of bistatic radar is presented. First, the remote sensing principle of GNSS-R bistatic radar was analyzed. Then, according to the distribution characteristics of the delay and the Doppler frequency on the sea surface, the reflection points of the sea surface were selected, and the area of corresponding reflection units was calculated. Subsequently, the calculation of the scattering coefficient was carried out. Finally, the simulation verification of the multiple combined signals was conducted. The simulation results indicate that the correlation coefficient of the simulated ocean reflection signal's correlation power curve and the theoretical curve of the ZV model is better than 0.92, which can be used to generate the GNSS ocean reflection signal effectively.

Keywords: global navigation satellite system (GNSS); ocean remote sensing; reflection signal; bistatic radar; ZV model

北京航空航天大学学报

Received: 2017-01-16; **Accepted**: 2017-02-15; **Published online**: 2017-04-28 15:12

URL: kns.cnki.net/kcms/detail/11.2625.V.20170428.1512.001.html

Foundation item: National Natural Science Foundation of China (61171070)

* **Corresponding author**. E-mail: yangdongkai@sina.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0031

纠错输出编码的留一误差界估计

薛爱军，王晓丹*
（空军工程大学 防空反导学院，西安 710051）

摘 要：纠错输出编码(ECOC)作为分解框架，将多类分类问题转化为二类分类问题，是解决多类分类问题的有效手段。为了提高ECOC的泛化性能，对ECOC基分类器的设计问题进行了研究。解决这一问题的关键是对ECOC的泛化性能进行估计。留一(LOO)误差作为泛化性能的无偏估计，研究了ECOC留一误差界的估计问题。先给出了ECOC留一误差的定义，基于此定义，再给出了基分类器为支持向量机(SVM)，解码方法为线性损失函数解码时，ECOC留一误差的上界和下界。在人工数据集和UCI数据集上的实验表明，ECOC留一误差的上界可以指导基分类器的参数选择，通过基分类器设计可以提高ECOC的泛化性能。此外，ECOC的训练误差可以作为ECOC留一误差的下界，对ECOC留一误差下界的研究可以作为未来的研究方向。

关 键 词：模式识别；多类分类；纠错输出编码(ECOC)；泛化性能(LOO)；留一误差
中图分类号：TP391
文献标识码：A **文章编号**：1001-5965(2018)01-0132-10

在现实世界中，很多应用都面临着多目标识别问题，比如，人脸识别[1]、手写体数字识别[2]、交通指示牌识别[3]等。在模式识别领域中，对多个目标进行识别属于多类分类问题，需要设计多类分类器。现有的分类方法中，有一些可以直接应用于解决多类分类问题，比如：决策树[4]和神经网络[2]等。但其他一些成熟的、性能较好的分类方法却难以直接应用到多类分类问题中，比如：支持向量机(SVM)[5]和Adaboost[6]等，它们属于二类分类器，可以很好地解决二类分类问题。将二类分类方法扩展到多类分类领域的一种解决思路，就是采取分而治之的策略，首先将多类分类问题分解为若干个二类分类问题，然后利用二类分类方法解决二类分类问题，最后将二类分类的结果融合得到多类分类的结果。

纠错输出编码(ECOC)提供了一种通用的分解框架，可以将复杂的多类分类问题分解为一系列的二类分类问题。自1995年由Dietterich和Bakiri[7]首次提出以来，得到了众多学者的关注。近年来，有关学者还将其应用于疾病诊断领域[8-9]。在实际应用中，科研人员更关心的是ECOC的泛化能力，即ECOC对未知类别样本的预测能力。影响ECOC泛化能力的因素主要有3个：编码方法、基分类器设计和解码方法。这里的基分类器就是前面提到的二类分类器。为了提高ECOC的泛化性能，很多学者对编码方法和解码方法进行了研究。Bautista等[10]研究了基于进化算法生成编码矩阵的方法；雷蕾等[11-12]提出了基于特征空间变换的编码方法和基于支持向量数据描述的层次纠错输出编码构造方法；周进登等[13]提出利用单层感知器作为学习框架，通过学习得到编码矩阵的方法。此外，Ismailoglu等[14]研究了通过

收稿日期：2017-01-17；**录用日期**：2017-05-12；**网络出版时间**：2017-06-19 15:53
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170619.1553.004.html
基金项目：国家自然科学基金(61273275，61703426)
* **通信作者**. E-mail：wang_afeu@126.com

引用格式：薛爱军，王晓丹. 纠错输出编码的留一误差界估计[J]. 北京航空航天大学学报，2018，44(1)：132-141.
XUE A J，WANG X D. Leave-one-out error bounds estimation for error correcting output codes[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(1)：132-141(in Chinese).

权值对分类器性能进行惩罚的加权解码方法;Passerini 等[15]提出了基于后验概率的解码方法,在此基础上,周进登等[16]又将其扩展到三符号编码的解码过程中。关于编码方法和解码方法更加详细的综述可以参考文献[17]。上述研究工作从不同角度为提高 ECOC 的泛化性能提供了解决思路。

但科研人员对于 ECOC 基分类器的设计关注较少,也鲜有文献对此展开研究。2002 年,Crammer 和 Singer[18]对现有的纠错输出编码方法进行了分析和总结,提出了 3 个学习问题,其中之一就是给定编码矩阵,如何找到使分类错误率最小的一组二类分类器。基分类器的设计问题就是在编码矩阵给定的前提下,如何找到一组最优的基分类器,使得 ECOC 的泛化误差最小。2004 年,Passerini 等[15]对 ECOC 基分类器的设计问题进行了研究,探讨了基分类器为 SVM 时的参数选择问题。

本文的目的就是在 Passerini 等[15]工作的基础上,对 ECOC 基分类器的设计问题开展进一步的研究,为提高 ECOC 的泛化性能做出一定的贡献。基分类器设计的目标是减小 ECOC 的泛化误差,因此,解决这一问题的关键就是如何计算 ECOC 的泛化误差。但由于样本的分布是未知的,ECOC 真实的泛化误差往往难以计算,因此,实际中科研人员经常使用的是泛化误差的估计值。留一(LOO)误差可以看做泛化误差的一种无偏估计。但当样本的数量较大时,计算留一误差需要耗费大量的时间。一种可行的解决思路就是估计留一误差的界,利用留一误差界对 ECOC 的泛化性能进行估计。本文的主要工作就是给出 ECOC 留一误差界的估计。

同时,为了简化问题,本文也只针对基分类器为 SVM 时的情形进行研究。当基分类器为 SVM 时,其设计问题就是如何选择最优的参数问题,这些参数包括核参数和正则化参数。不同的参数下,基分类器的泛化能力不同,进而影响 ECOC 整体的泛化性能。另外,对于 ECOC 来讲,其基分类器的参数选择有 2 种情形:一是每个基分类器的参数不相同,需要分别为每个基分类器选择最优的参数;二是所有的基分类器的参数相同,可以同时为所有的基分类器选择最优的参数。显然,前者的难度要远大于后者。因此,本文假设所有的基分类器具有相同的参数。

本文首先介绍了 ECOC 框架;然后给出了 ECOC 留一误差的定义,从此定义出发,分别给出了 ECOC 留一误差的上界和下界估计;最后给出了实验结果与分析。

1　纠错输出编码

ECOC 是一种利用二元或三元的编码矩阵将多类分类问题分解为若干个二类分类问题的通用集成框架。其中,二元编码矩阵可以表示为 $\boldsymbol{M}\in\{-1,+1\}^{k\times S}$,$k$ 为类别数,S 为编码长度,编码矩阵的行代表某一类,编码矩阵的列代表一个二类划分,“-1”表示其所在行对应的类在二类划分中被划分为负类,“+1”表示其所在行对应的类在二类划分中被划分为正类;三元编码矩阵可以表示为 $\boldsymbol{M}\in\{-1,0,+1\}^{k\times S}$,其中“0”表示其所在行对应的类在二类划分中被忽略。图 1 给出了 4 种常见的 ECOC 编码方法,分别是:“一对多”编码、“一对一”编码、密集随机编码、稀疏随机编码。图 1 中码元“+1”、“-1”和“0”分别用白色、黑色和灰色表示。

ECOC 解决多类分类问题时,通常可以划分为三个阶段:编码阶段、训练阶段和解码阶段。在编码阶段,形成多个二类划分,每个二类划分对应一个二类分类任务。在训练阶段,根据二类划分生成基分类器的训练数据集,完成对基分类器 f_s ($s=1,2,\cdots,S$)的训练。例如,在图 1(d)中对基分类器 f_3 进行训练时,选择白色对应的 C_2 为正类,黑色对应的 C_4 为负类,构成训练数据集,而忽略灰色对应的类别 C_1 和 C_3。

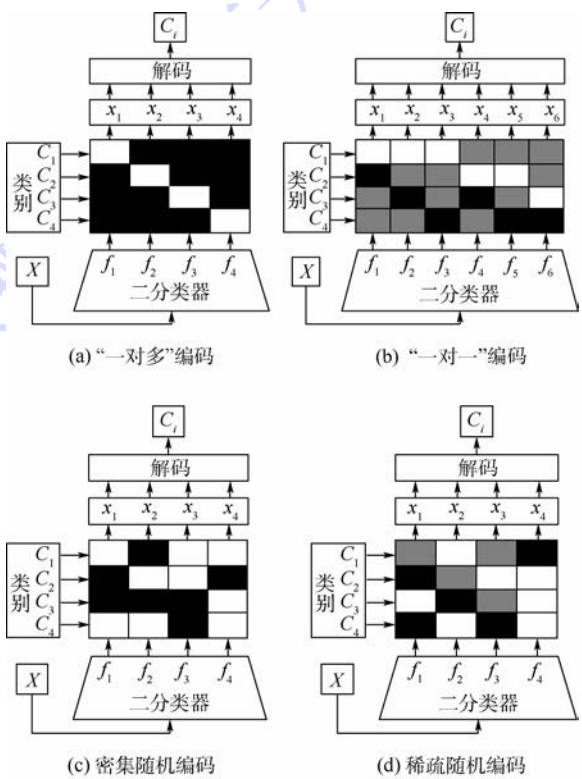


图 1　4 种常见的 ECOC

Fig. 1　Four commonly-used ECOCs

北航学报 赠阅

在解码阶段,根据多个基分类器的分类结果,利用某种解码规则(融合策略)得到最终的分类结果。例如,在图 1(d)给定一个测试样本 X,利用训练得到的基分类器对其进行分类,结果为一码字向量(x_1,x_2,x_3,x_4)(其中 $x_i\in\{-1,+1\}$),根据码字向量与编码矩阵行之间的最小汉明距离得到分类结果。

2 ECOC 留一误差及上下界的估计

2.1 ECOC 的留一误差

首先定义样本 x 的多分类间隔[15]:

$$g(x,y)=d_{\mathrm{L}}(\boldsymbol{m}_p,\boldsymbol{f}(x))-d_{\mathrm{L}}(\boldsymbol{m}_y,\boldsymbol{f}(x)) \tag{1}$$

式中:

$$p=\arg\min_{q\neq y} d_{\mathrm{L}}(\boldsymbol{m}_q,\boldsymbol{f}(x)) \tag{2}$$

$\boldsymbol{m}_y$ 为编码矩阵行向量;$\boldsymbol{f}(x)$ 为基分类器输出向量;$d_{\mathrm{L}}(\cdot)$为线性损失函数解码。

根据 ECOC 的解码规则,当 $g(x,y)<0$ 时,样本 x 被错分;而当 $g(x,y)\geqslant 0$ 时,样本 x 被正确分类。由此可见,根据多分类间隔 $g(x,y)$ 的取值,可以对样本 x 是否被 ECOC 错误分类进行判断。进而对于训练集中的所有样本,ECOC 的经验误差可以写为

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\theta(-g(x_i,y_i)) \tag{3}$$

式中:ℓ 为训练集中样本的个数;y_i 为样本 x_i 的类别;$\theta(\cdot)$为一函数,其定义为

$$\theta(\cdot)=\begin{cases}1 & x>0\\ 0 & \text{其他}\end{cases}$$

参考 ECOC 经验误差的定义,可以给出 ECOC 留一误差的定义。假设在留一测试过程中,样本 x_i 是被剔除的样本。此时,剔除样本后得到的新的训练集的样本个数为 $\ell-1$。在新的训练集上训练 ECOC,训练后得到一组基分类器,记为 $\boldsymbol{f}^i(\cdot)$。接下来利用多分类间隔对样本 x_i 是否被 ECOC 错误分类进行判断。被剔除样本 x_i 的多分类间隔记为 $g^i(x_i,y_i)$,其定义为

$$g^i(x_i,y_i)=d_{\mathrm{L}}(\boldsymbol{m}_{p^i},\boldsymbol{f}^i(x_i))-d_{\mathrm{L}}(\boldsymbol{m}_{y_i},\boldsymbol{f}^i(x_i)) \tag{4}$$

式中:

$$p^i=\arg\min_{q\neq y_i} d_{\mathrm{L}}(\boldsymbol{m}_q,\boldsymbol{f}^i(x_i)) \tag{5}$$

同样,根据 ECOC 的决策规则,当 $g^i(x_i,y_i)<0$时,样本 x_i 将被错误分类;否则,样本 x_i 将被正确分类。即根据 $g^i(x_i,y_i)$ 的取值,可以对样本 x_i 在留一测试过程中是否会被错误分类进行判断。因此,对于训练集中的所有样本,ECOC 的留一误差可以定义为

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\theta(-g^i(x_i,y_i)) \tag{6}$$

通常,科研人员认为留一误差可以看作一个学习算法泛化误差的无偏估计,可以有效地指导学习算法中的参数选择。但在实际应用中,计算某一算法的留一误差的复杂度很高,时间开销很大。假设某一训练集的样本个数为 ℓ,则计算某一算法的留一误差,需要对该学习算法进行 ℓ 次训练。特别的,当训练集中的样本数很大时,计算学习算法的留一误差成为了不可能完成的任务。因此,在解决实际问题时,科研人员经常致力于寻找留一误差的界,利用这一界值对学习算法的留一误差进行估计,进而估计该学习算法的泛化误差。

接下来,从 ECOC 的留一误差的定义出发,分别给出 ECOC 的留一误差的上界和下界。

2.2 ECOC 留一误差的上界和下界估计

估计 ECOC 留一误差的界的关键在于,如何判断被剔除样本 x_i 在留一测试过程中是否被错误分类。这一点可以通过对多分类间隔 $g^i(x_i,y_i)$取值的判断来实现。当 $g^i(x_i,y_i)<0$ 时,就可以认为样本 x_i 在留一测试过程中被错误分类。因此,估计 ECOC 的留一误差界的问题就转化为估计多分类间隔 $g^i(x_i,y_i)$的取值问题。

为了简化问题,本文假设解码方法为线性损失函数解码,其定义为

$$d_{\mathrm{L}}(\boldsymbol{m}_q,\boldsymbol{f})=-\boldsymbol{m}_q\cdot\boldsymbol{f}=-\sum_{s=1}^{S}m_{qs}f_s \tag{7}$$

此时,多分类间隔 $g^i(x_i,y_i)$可以写为

$$g^i(x_i,y_i)=\boldsymbol{m}_{y_i}\cdot\boldsymbol{f}^i(x_i)-\boldsymbol{m}_{p^i}\cdot\boldsymbol{f}^i(x_i) \tag{8}$$

在留一测试过程中,每剔除掉一个样本都要重新训练一次 $\boldsymbol{f}^i(\cdot)$,这是导致计算留一误差时间开销比较大的主要原因。因此,为了减少计算开销,需要消除 $\boldsymbol{f}^i(\cdot)$对多分类间隔 $g^i(x_i,y_i)$取值的影响。幸运的是,文献[15]中的引理 1 建立了 $\boldsymbol{f}^i(\cdot)$与$\boldsymbol{f}(\cdot)$之间的联系。由于$\boldsymbol{f}(\cdot)$仅需要在整个训练集上训练一次,因此,如果利用 $\boldsymbol{f}(\cdot)$代替 $\boldsymbol{f}^i(\cdot)$,理论上讲,只需要在整个训练集上训练一次 ECOC,就可以判断所有样本的多分类间隔 $g^i(x_i,y_i)$的取值。从而,大大减小计算开销。

引理 1[15] 假设 f 为一个在整个训练集上训练得到的 SVM,f^i为剔除掉样本 x_i 后在新的训练集上训练得到的 SVM。则有式(9)成立:

$$y_i f(x_i)-\alpha_i G_{ii}\leqslant y_i f^i(x_i)\leqslant y_i f(x_i) \tag{9}$$

式中:α_i 为求解优化问题得到的拉格朗日系数;$G_{ii}=K(x_i,x_i)$,$K(\cdot,\cdot)$为核函数。

引理 1 的证明过程见文献[15]。将引理 1 应用于 ECOC 的每一个基分类器,不等式(9)可以重写为

$$f_s^i(x_i)=f_s(x_i)-\lambda_s m_{y_i^s} \qquad s\in\{1,2,\cdots,S\} \tag{10}$$

式中:$\lambda_s\in[0,\alpha_i^s G_{ii}^s]$。

将式(10)代入多分类间隔 $g^i(x_i,y_i)$ 的定义中,可以得到

$$g^i(x_i,y_i)=\sum_{s=1}^{S}(m_{y_i^s}-m_{p^{is}})f_s^i(x_i)=\sum_{s=1}^{S}[(m_{y_i^s}-m_{p^{is}})f_s(x_i)-m_{y_i^s}(m_{y_i^s}-m_{p^{is}})\lambda_s] \tag{11}$$

这样,在多分类间隔 $g^i(x_i,y_i)$ 的定义中,消除了 $\boldsymbol{f}^i(\cdot)$ 的影响。但仍然需要指出的是,在式(11)中,有 2 个参数是无法确定的,分别是 p^i 和 λ_s。这 2 个参数的未知性,导致无法准确的得到多分类间隔 $g^i(x_i,y_i)$ 的取值。因此,本文通过对多分类间隔 $g^i(x_i,y_i)$ 的缩放,对多分类间隔 $g^i(x_i,y_i)$ 的取值进行估计,分别得到了 ECOC 的留一误差上界和下界。

首先,需要说明多分类间隔 $g^i(x_i,y_i)$ 的取值的缩放与 ECOC 留一误差上界和下界之间的联系。假设多分类间隔 $g^i(x_i,y_i)$ 的取值满足下面的不等式,或者说多分类间隔 $g^i(x_i,y_i)$ 的取值在区间$[A,B]$上。

$$A\leqslant g^i(x_i,y_i)\leqslant B \tag{12}$$

式中:A 和 B 的取值是可以确定的。这样,通过 A 和 B 就可以估计多分类间隔 $g^i(x_i,y_i)$ 的取值。

当 $g^i(x_i,y_i)<0$ 时,一定有 $A<0$ 成立;但是反之则不然,当 $A<0$ 时,不一定有 $g^i(x_i,y_i)<0$ 成立。因此,可以得出结论,$A<0$ 是 $g^i(x_i,y_i)<0$ 成立的必要条件而非充分条件。进一步,如果根据 A 的取值对 ECOC 的留一误差进行估计,得到的一定是 ECOC 留一误差的上界,即有

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\theta(-A_i)\geqslant\frac{1}{\ell}\sum_{i=1}^{\ell}\theta(-g^i(x_i,y_i)) \tag{13}$$

式中:A_i 为基于样本 x_i 计算得到的 A。

另一方面,当 $B<0$ 时,一定有 $g^i(x_i,y_i)<0$ 成立;但反之,当 $g^i(x_i,y_i)<0$ 时,不一定有 $B<0$ 成立。因此,可以得出结论,$B<0$ 是 $g^i(x_i,y_i)<0$ 成立的充分条件而非必要条件。进一步,如果根据 B 的取值对 ECOC 的留一误差进行估计,得到的一定是 ECOC 留一误差的下界,即有

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\theta(-B_i)\leqslant\frac{1}{\ell}\sum_{i=1}^{\ell}\theta(-g^i(x_i,y_i)) \tag{14}$$

式中:B_i 为基于样本 x_i 计算得到的 B。

由此可以看出,如何找到合适的 A 和 B 是得到 ECOC 留一误差上界和下界的关键,也是本文致力于解决的问题。下面通过定理 1 和定理 2 分别给出 ECOC 的留一误差上界和下界。

定理 1　当基分类器为 SVM,解码方法为线性损失函数解码时,ECOC 的留一误差上界为

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\theta\left(\min_{q\neq y_i}\sum_{s=1}^{S}[(m_{qs}-m_{y_i^s})f_s(x_i)+m_{y_i^s}(m_{y_i^s}-m_{qs})\alpha_i^s G_{ii}^s]\right)$$

证明　首先考虑 $m_{y_i^s}(m_{y_i^s}-m_{p^{is}})$ 的取值。$m_{y_i^s}$ 和 $m_{p^{is}}$ 都有 3 种可能的取值,分别是 +1,0 和 −1。

当 $m_{y_i^s}=0$ 时,$m_{y_i^s}(m_{y_i^s}-m_{p^{is}})=0$。

当 $m_{y_i^s}=+1$ 时,

$$m_{y_i^s}(m_{y_i^s}-m_{p^{is}})=\begin{cases}0 & m_{p^{is}}=+1\\ 1 & m_{p^{is}}=0\\ 2 & m_{p^{is}}=-1\end{cases}$$

当 $m_{y_i^s}=-1$ 时,

$$m_{y_i^s}(m_{y_i^s}-m_{p^{is}})=\begin{cases}2 & m_{p^{is}}=+1\\ 1 & m_{p^{is}}=0\\ 0 & m_{p^{is}}=-1\end{cases}$$

所以,$m_{y_i^s}(m_{y_i^s}-m_{p^{is}})$ 的取值是非负的。同时,考虑到 $\lambda_s\in[0,\alpha_i^s G_{ii}^s]$,则有

$$g^i(x_i,y_i)=\sum_{s=1}^{S}(m_{y_i^s}-m_{p^{is}})f_s^i(x_i)=\sum_{s=1}^{S}[(m_{y_i^s}-m_{p^{is}})f_s(x_i)-m_{y_i^s}(m_{y_i^s}-m_{p^{is}})\lambda_s]\geqslant\sum_{s=1}^{S}[(m_{y_i^s}-m_{p^{is}})f_s(x_i)-m_{y_i^s}(m_{y_i^s}-m_{p^{is}})\alpha_i^s G_{ii}^s]$$

式中:p^i 与 $\boldsymbol{f}^i(\cdot)$ 相关,其值仍然难以确定。事实上,p^i 代表了编码矩阵 $\boldsymbol{M}$ 中的某一行。因此可得

$$g^i(x_i,y_i)\geqslant\sum_{s=1}^{S}[(m_{y_i^s}-m_{p^{is}})f_s(x_i)-m_{y_i^s}(m_{y_i^s}-m_{p^{is}})\alpha_i^s G_{ii}^s]\geqslant\min_{q\neq y_i}\sum_{s=1}^{S}[(m_{y_i^s}-m_{qs})f_s(x_i)-m_{y_i^s}(m_{y_i^s}-m_{qs})\alpha_i^s G_{ii}^s]$$

当 $g^i(x_i,y_i)<0$ 时,则一定有

$$\min_{q\neq y_i}\sum_{s=1}^{S}[(m_{y_i^s}-m_{qs})f_s(x_i)-m_{y_i^s}(m_{y_i^s}-m_{qs})\alpha_i^s G_{ii}^s]<0$$

因此,可以得到

$$A_i=\min_{q\neq y_i}\sum_{s=1}^{S}[(m_{y_i^s}-m_{qs})f_s(x_i)-$$

北航学报 赠阅

$m_{y_i s}(m_{y_i s}-m_{qs})\alpha_i^s G_{ii}^s]$ **证毕**

定理 2 当基分类器为 SVM,解码方法为线性损失函数解码时,ECOC 的留一误差下界为

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\theta\left(\min_{q\neq y_i}\sum_{s=1}^{S}[(m_{qs}-m_{y_i s})f_s(x_i)]\right)$$

证明 考虑到 $m_{y_i s}(m_{y_i s}-m_{p^i s})$ 和 λ_s 的非负性,则有

$$g^i(x_i,y_i)=\sum_{s=1}^{S}(m_{y_i s}-m_{p^i s})f_s^i(x_i)=\sum_{s=1}^{S}[(m_{y_i s}-m_{p^i s})f_s(x_i)-m_{y_i s}(m_{y_i s}-m_{p^i s})\lambda_s]\leqslant\sum_{s=1}^{S}[(m_{y_i s}-m_{p^i s})f_s(x_i)]$$

同样的,p^i 代表了编码矩阵 $\boldsymbol{M}$ 中的某一行。因此,则有

$$g^i(x_i,y_i)\leqslant\min_{q\neq y_i}\sum_{s=1}^{S}[(m_{y_i s}-m_{qs})f_s(x_i)]\leqslant\sum_{s=1}^{S}[(m_{y_i s}-m_{p^i s})f_s(x_i)]$$

则当 $\min_{q\neq y_i}\sum_{s=1}^{S}[(m_{y_i s}-m_{qs})f_s(x_i)]<0$ 成立时,一定有 $g^i(x_i,y_i)<0$ 成立。

所以,可以得到

$$B_i=\min_{q\neq y_i}\sum_{s=1}^{S}[(m_{y_i s}-m_{qs})f_s(x_i)]$$ **证毕**

还需要注意的一点是,定理 2 给出的 ECOC 留一误差的下界可以进一步写为

$$\frac{1}{\ell}\sum_{i=1}^{\ell}\theta\left(\min_{q\neq y_i}\sum_{s=1}^{S}[(m_{qs}-m_{y_i s})f_s(x_i)]\right)=\frac{1}{\ell}\sum_{i=1}^{\ell}\theta\left(\min_{q\neq y_i}\sum_{s=1}^{S}m_{qs}f_s(x_i)-\sum_{s=1}^{S}m_{y_i s}f_s(x_i)\right)=\frac{1}{\ell}\sum_{i=1}^{\ell}\theta(d_{\mathrm{L}}(\boldsymbol{m}_{y_i},\boldsymbol{f}(x_i))-d_{\mathrm{L}}(\boldsymbol{m}_p,\boldsymbol{f}(x_i)))=\frac{1}{\ell}\sum_{i=1}^{\ell}\theta(-g(x_i,y_i)) \tag{15}$$

由此可以看出,ECOC 留一误差的下界实际上就是 ECOC 的经验误差,也称为训练误差。

至此,本文分别给出了 ECOC 留一误差的上界和下界。下面将通过实验对这 2 个界的有效性进行验证。

3 实 验

下面从实验数据、实验设计以及实验结果与分析 3 个方面对实验进行介绍。

3.1 实验数据

实验数据为一个人工数据集和多个 UCI 数据集[19]。人工数据集共包含 5 类样本,每一类的样本个数为 50,各类的数据分布如图 2 所示。每一类特征向量的维数为 2,图中分别用 F_1 和 F_2 表示,并且服从正态分布,其概率密度函数为

$$p(\boldsymbol{x}\mid C_i)=\frac{1}{2\pi|\boldsymbol{\Sigma}_i|^{1/2}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right)\quad i=1,2,\cdots,5 \tag{16}$$

式中:$\boldsymbol{x}$ 为二维特征向量;$\boldsymbol{\mu}_i$ 为平均值向量;$\boldsymbol{\Sigma}_i$ 为协方差矩阵;各参数的设置见表 1。

实验中,从 UCI 数据集中共选取了 6 个多类的数据集。表 2 给出了这些数据集的样本的数目、特征量的维数以及包含类别的个数。

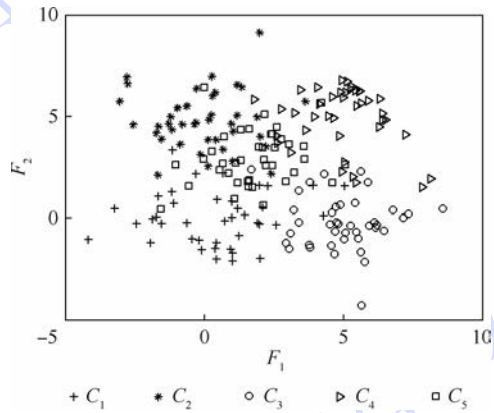


图 2 人工数据集的数据分布

Fig. 2 Data distribution of synthetic dataset

表 1 人工数据集的参数设置

Table 1 Parameter setting for synthetic dataset

类 别	先验概率	平均值向量	协方差矩阵
C_1	$P(C_1)=\frac{1}{5}$	$\boldsymbol{\mu}_1=(0,0)^{\mathrm{T}}$	$\boldsymbol{\Sigma}_1=\begin{pmatrix}2&0\\0&2\end{pmatrix}$
C_2	$P(C_2)=\frac{1}{5}$	$\boldsymbol{\mu}_2=(0,5)^{\mathrm{T}}$	$\boldsymbol{\Sigma}_2=\begin{pmatrix}2&0\\0&2\end{pmatrix}$
C_3	$P(C_3)=\frac{1}{5}$	$\boldsymbol{\mu}_3=(5,0)^{\mathrm{T}}$	$\boldsymbol{\Sigma}_3=\begin{pmatrix}2&0\\0&2\end{pmatrix}$
C_4	$P(C_4)=\frac{1}{5}$	$\boldsymbol{\mu}_4=(5,5)^{\mathrm{T}}$	$\boldsymbol{\Sigma}_4=\begin{pmatrix}2&0\\0&2\end{pmatrix}$
C_5	$P(C_5)=\frac{1}{5}$	$\boldsymbol{\mu}_5=(2,3)^{\mathrm{T}}$	$\boldsymbol{\Sigma}_5=\begin{pmatrix}2&0\\0&2\end{pmatrix}$

表 2 实验中用到的 UCI 数据集

Table 2 UCI datasets used in experiment

数 据 集	样本个数	特征维数	类 别 数
vowel	990	13	11
balance	625	4	3
glass	214	10	6
vehicle	846	18	4
letter	1 214	16	26
segmentation	2 310	19	7

3.2 实验设计

本实验设计的目的是为了检验定理 1 和定

理 2 给出的 ECOC 留一误差上界和下界的有效性。留一误差界的有效性是指留一误差的上界或者下界具有与留一误差相同或者相近的变化趋势。这样,当留一误差的上界或下界减小时,能够推断出留一误差也应该相应的减小。

由于人工数据集的样本个数较少,共有 200 个样本,因此,在人工数据集上进行留一误差测试,统计实际的留一误差。考虑到 UCI 数据集的样本个数较多,进行留一误差测试耗时较长,所以,在 UCI 数据集上进行 20 重交叉验证,以交叉验证的结果来近似代替留一误差。

实验中基分类器为 SVM,核函数为高斯径向基核函数。编码方法包括“一对多”编码和“一对一”编码。解码方法为线性损失函数解码。正则化参数 C 和核参数 σ 的变化范围均设置为$\{2^{-5},2^{-4},\cdots,2^{8}\}$。为了简化实验,所有基分类器的正则化参数和核参数的设置相同。

3.3　实验结果与分析

3.3.1　人工数据集

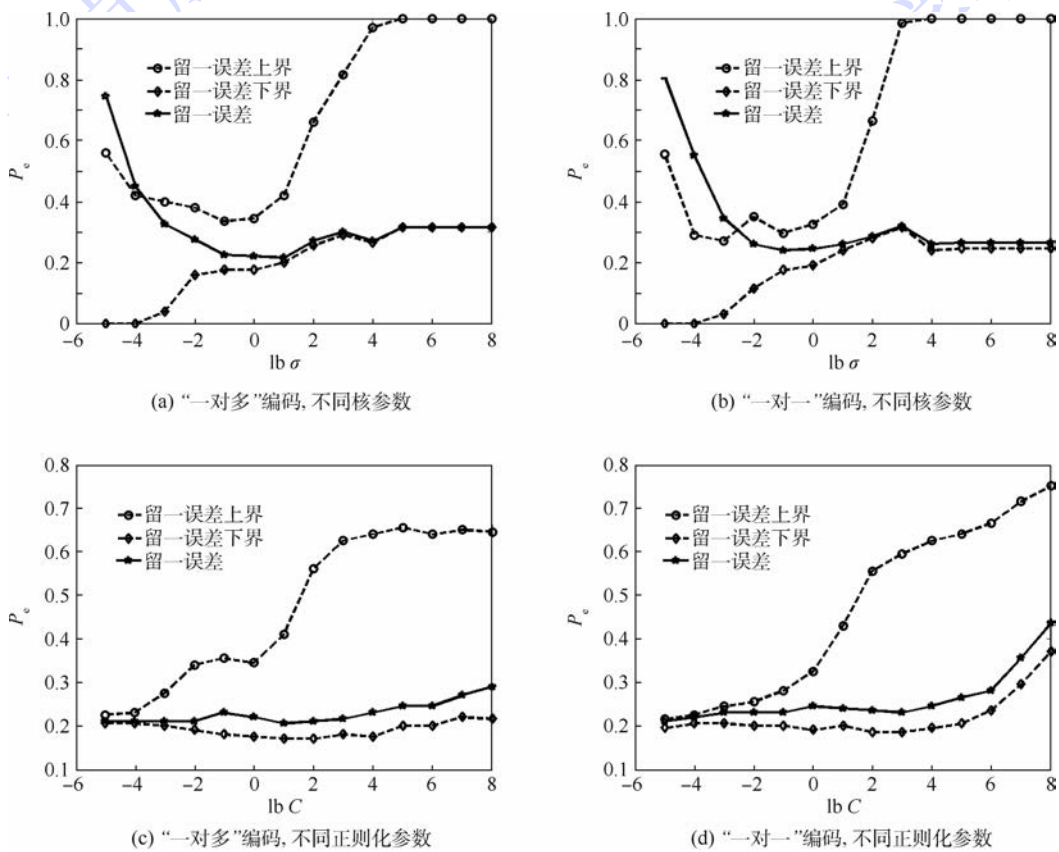
图 3 分别给出了不同核参数和正则化参数下的留一误差和留一误差上下界的变化情况,分类错误率为P_e。由图3可以看出,留一误差随着这 2 个参数的变化起伏较大,这证明了通过对正则化参数和核参数进行优化,可以有效地提高 ECOC 的泛化性能。对比图 3(a)、(b)和图 3(c)、(d)中留一误差变化曲线的起伏程度可知,核参数对 ECOC 的留一误差的影响要大于正则化参数的影响。这说明,对核参数进行优化对于提高 ECOC 的泛化性能更加重要。

由图 3 还可以看出,留一误差上界最小值出现的位置与留一误差最小值出现的位置非常接近。因此,利用留一误差上界可以指导本文选择最优的核参数和正则化参数。可以说,定理 1 给出的留一误差的上界是有效的。同时,还需要说明的是,定理 1 中留一误差上界的计算仅需要将分类器在训练集上训练一次,不需要增加额外的计算代价。

但是,图 3(a)、(b)中留一误差下界最小值出现的位置却与留一误差最小值出现的位置偏差很大。这说明,留一误差的下界是不能有效地指导选择最优参数。应该注意到,定理 2 给出的留一误差下界实际上就是 ECOC 的训练误差。这也就从另一个角度验证了最小化训练误差并不一定能够最小化泛化误差。

3.3.2　UCI 数据集

图 4 和图 5 分别给出了 UCI 数据集上不同核



(a) “一对多”编码, 不同核参数　(b) “一对一”编码, 不同核参数

(c) “一对多”编码, 不同正则化参数　(d) “一对一”编码, 不同正则化参数

图 3　人工数据集上不同核参数和正则化参数对应的留一误差和留一误差上下界

Fig. 3　LOO error and LOO error's upper and lower bounds with different kernel parameters and regularization parameters on synthetic dataset

北航学报 赠阅

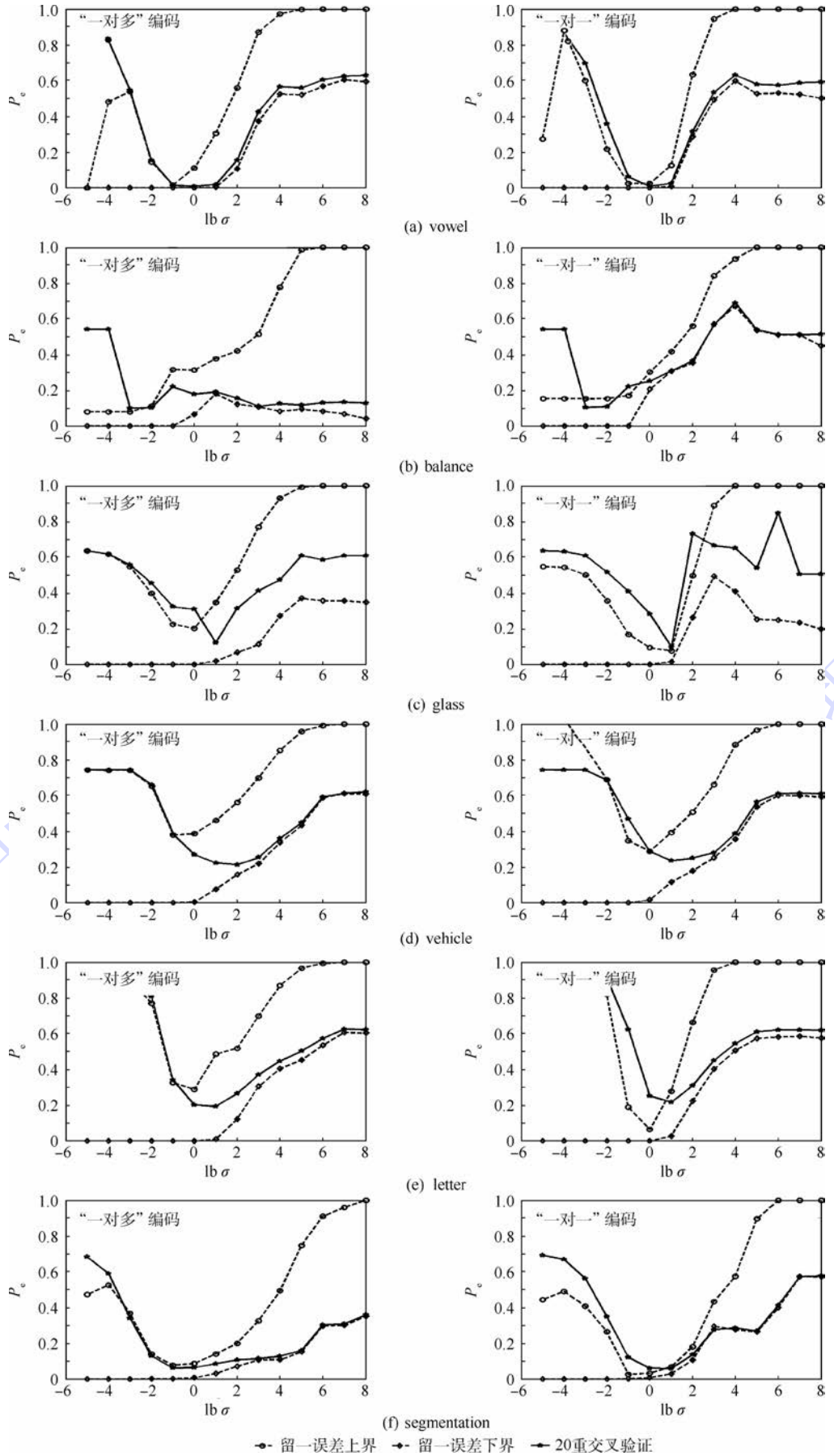


图 4 UCI 数据集上不同核参数下 20 重交叉验证的结果及留一误差上下界

Fig. 4 20-fold cross validation results and LOO error's upper and lower bound with different kernel parameters on UCI datasets

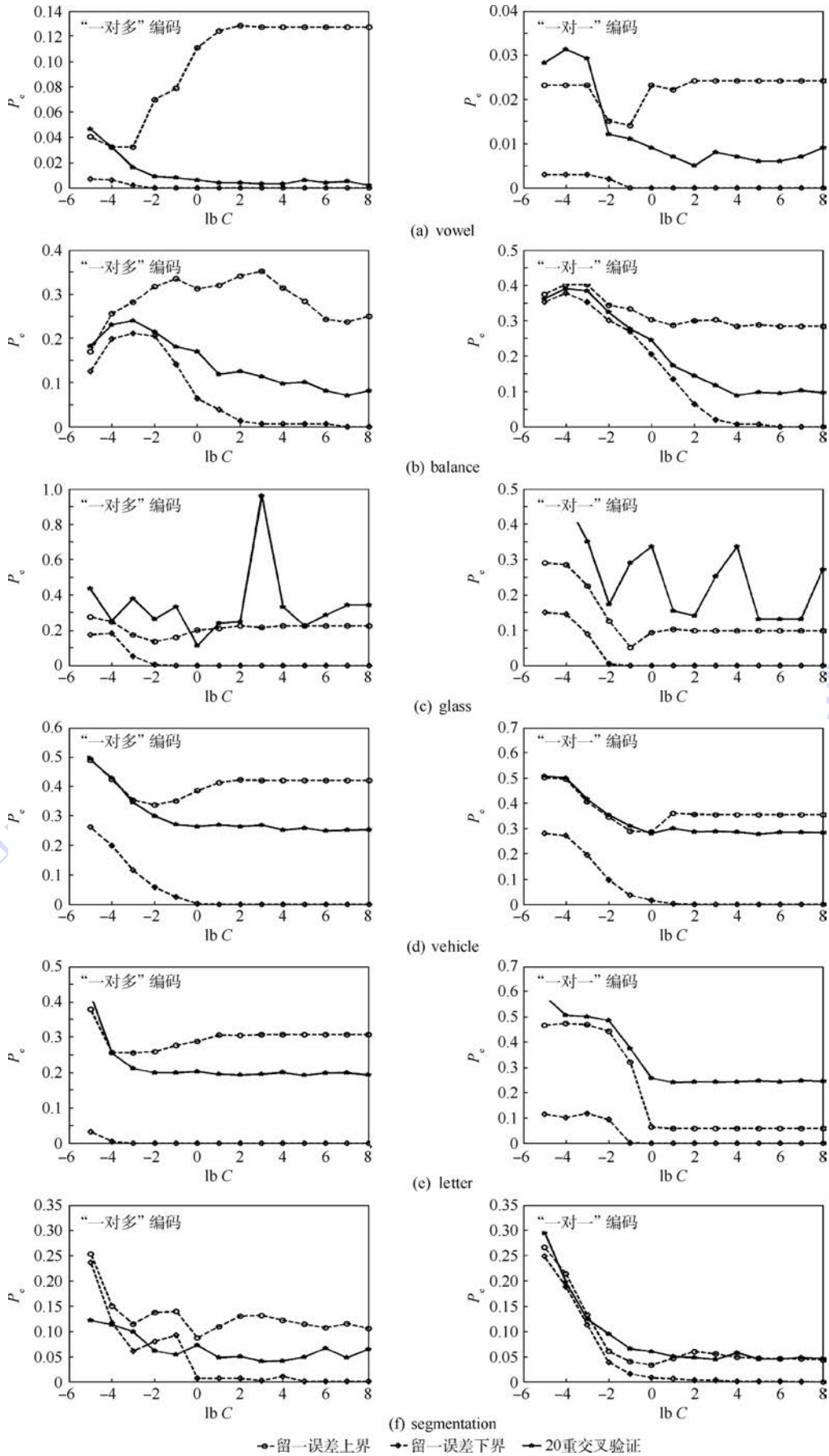


图 5　UCI 数据集上不同正则化参数下 20 重交叉验证的结果及留一误差上下界

Fig. 5　20-fold cross validation results and LOO error's upper and lower bound with different regularization parameters on UCI datasets

参数和正则化参数下 20 重交叉验证的结果及留一误差上下界的变化情况。由图中可以看出,根据留一误差上界的最小值可以找到一个相对较小的 20 重交叉验证结果。图中尽管留一误差上界的最小值与 20 重交叉验证结果最小值之间存在一定的偏差,但根据留一误差上界最小值对参数进行择优仍然可以很大程度上提高分类器的泛化能力。因此,可以说,在 UCI 数据集上定理 1 给出的留一误差上界仍然可以有效地指导核参数和正则化参数的选择。

另一方面,从图 4 中可以看到,20 重交叉验证结果的最小值和留一误差上界的最小值几乎都出现在了留一误差下界等于 0 的位置上,也就是训练误差为 0 的位置。这似乎可以说明最小化训练误差是最小化泛化误差的必要条件,但非充分条件。更有趣的现象是,图 3(c)、(d)和图 5 中在不同的正则化参数下,与留一误差上界相比,留一误差下界具有与留一误差或者 20 重交叉验证结果更加相似的变化趋势。并且,在图 3(a)、(b)和图 4 中,当核参数取值较大时,留一误差上界都出现了完全收敛于留一误差或者 20 重交叉验证结果的现象。由于知识积累的不足,现阶段本文无法对上述现象给出合理的解释,也无法对定理 2给出的留一误差下界在 ECOC 模型选择中的应用给出准确的说明。但文献[20]已经指出,泛化误差界的下界估计可以提供经验风险收敛于期望风险的本质属性,从而可以更本质的判别学习算法的性能。因此,未来对 ECOC 的留一误差下界进行深入的研究将是一项富有意义的工作。

4 结 论

1) 本文给出的 ECOC 留一误差的上界可以有效地指导基分类器参数的选择,其中,对核参数的优化重要性更大。

2) 通过为基分类器选择最优的参数,可以有效地提高 ECOC 的泛化性能。

3) ECOC 的经验误差或者说训练误差可以作为 ECOC 留一误差的下界。

4) 本文给出的 ECOC 留一误差的下界在某些情况下可以完全收敛于真实的留一误差。

未来为了提高对 ECOC 泛化能力的估计精度,需要对 ECOC 留一误差的下界进行更加深入的研究。此外,当每个基分类器的参数不相同时,研究 ECOC 的基分类器的设计问题也是一项富有挑战性的工作。探讨其它解码方法下的 ECOC 留一误差界的估计也可以作为未来的一个研究方向。

参考文献(References)

[1] NI J, XU X Z, DING S F, et al. An adaptive extreme learning machine algorithm and its application on face recognition[J]. International Journal of Computing Science and Mathmatics, 2015, 6(6): 611-619.

[2] QURESHI M S, QURESHI M B, NABI M G, et al. Handwritten digit recognition system using neural network[J]. Energy Procedia, 2011, 13: 4326-4336.

[3] BERKAYA S K, GUNDUZ H, OZSEN O, et al. On circular traffic sign detection and recognition[J]. Expert System with Applications, 2016, 48: 67-75.

[4] NITHYA R, SANTHI B. Decision tree classifiers for mass classification[J]. International Journal of Signal and Imaging System Engineering, 2015, 8(1/2): 39-45.

[5] 边肇祺,张学工. 模式识别[M]. 2 版. 北京:清华大学出版社,2000:296-303.
BIAN Z Q, ZHANG X G. Pattern recognition[M]. 2nd ed. Beijing: Tsinghua University Press, 2000: 296-303 (in Chinese).

[6] FREUND Y, SHAPIRE R E. A decision-theoretic generalization of online learning and an application to boosting[J]. Journal of Computer and System Sciences, 1997, 55(1): 119-139.

[7] DIETTERICH T G, BAKIRI G. Solving multiclass learning problems via error-correcting output codes[J]. Journal of Artificial Intelligence Research, 1995, 2(1): 263-286.

[8] BAI X L, NIWAS S I, LIN W S, et al. Learning ECOC code matrix for multiclass classification with application to glaucoma diagnosis[J]. Journal of Medical Systems, 2016, 40(4): 78.

[9] LIU K H, ZENG Z H, NG V T Y. A hierarchical ensemble of ECOC for cancer classification based on multi-class microarray data[J]. Information Sciences, 2016, 349-350: 102-118.

[10] BAUTISTA M A, ESCALERA S, BARO X, et al. On the design of an ECOC-compliant genetic algorithm[J]. Pattern Recognition, 2014, 47(2): 865-884.

[11] 雷蕾,王晓丹,罗玺,等. 基于特征空间变换的纠错输出编码[J]. 控制与决策,2015,30(9):1597-1602.
LEI L, WANG X D, LUO X, et al. Error-correcting output codes based on feature space transformation[J]. Control and Decision, 2015, 30(9): 1597-1602 (in Chinese).

[12] 雷蕾,王晓丹,罗玺,等. 基于 SVDD 的层次纠错输出编码研究[J]. 系统工程与电子技术,2015,37(8):1916-1921.
LEI L, WANG X D, LUO X, et al. Hierarchical error-correcting output codes based on SVDD[J]. Systems Engineering and Electronics, 2015, 37(8): 1916-1921 (in Chinese).

[13] 周进登,周红建,杨云,等. 基于神经网络的纠错输出编码方法研究[J]. 电子学报,2013,41(6):1114-1121.
ZHOU J D, ZHOU H J, YANG Y, et al. Coding design for error correcting output codes based on neural network[J]. Acta Electronica Sinica, 2013, 41(6): 1114-1121 (in Chinese).

[14] ISMAILOGLU F, SPRINGHUIZEN I G, SMIRNOV E, et al. Fractional programming weighted decoding for error-correcting output codes[J]. Lecture Note in Computer Science, 2015, 9132: 38-50.

[15] PASSERINI A, PONTIL M, FRASCONI P. New results on error

correcting output codes of kernel machines[J]. IEEE Transactions on Neural Networks,2004,15(1):45-54.

[16] ZHOU J D,WANG X D,ZHOU H J,et al. Decoding design based on posterior probabilities in ternary error-correcting output codes[J]. Pattern Recognition,2012,45(4):1802-1818.

[17] 雷蕾,王晓丹,罗玺,等. ECOC 多类分类研究综述[J]. 电子学报,2014,42(9):1794-1800.
LEI L,WANG X D,LUO X,et al. An overview of multi-classification based on error-correcting output codes[J]. Acta Electronica Sinica,2014,42(9):1794-1800(in Chinese).

[18] CRAMMER K,SINGER Y. On the learnability and design of output codes for multiclass problems[J]. Machine Learning,2002,47(2-3):201-233.

[19] ASUNCION A,NEWMAN D. UCI machine learning repository[D]. Irvine: University of California,2007.

[20] 张海,徐宗本. 学习理论综述(Ⅰ):稳定性与泛化性[J]. 工程数学学报,2008,25(1):1-9.
ZHANG H,XU Z B. A survey on learning theory(Ⅰ):Stability and generalization[J]. Chinese Journal of Engineering Mathematics,2008,25(1):1-9(in Chinese).

作者简介:

薛爱军　男,博士研究生。主要研究方向:模式识别。

王晓丹　女,教授,博士生导师。主要研究方向:机器学习。

Leave-one-out error bounds estimation for error correcting output codes

XUE Aijun, WANG Xiaodan*

(Air and Missile Defense College, Air Force Engineering University, Xi'an 710051, China)

Abstract: Error correcting output codes (ECOC) is a decomposition framework, which can transform a complex multiclass classification problem into a series of two-class classification problems. It can complete one multiclass classification task efficiently. To improve its generalization performance, we studied the design of its base classifier, which is also known as model selection in ECOC. The key point is how to estimate the generalization error of ECOC. Leave-one-out (LOO) error is an almost unbiased estimator of generalization error, so we studied how to estimate the LOO error bounds for ECOC. First, we provided the definition of LOO error for ECOC. And then, based on this definition, upper bound and lower bound of LOO error for ECOC was given under the condition that base classifiers were support vector machines (SVM) and decoding method was linear loss function. The experiments on synthetic dataset and UCI dataset show that the upper bound of LOO error for ECOC leads to good estimates of parameters in base classifiers, and designing base classifiers can improve the generalization performance of ECOC. Furthermore, we also report that training error is one lower bound of LOO error for ECOC, and the application of this lower bound should be studied in the future.

Keywords: pattern recognition; multiclass classification; error correcting output codes (ECOC); generalization performance; leave-one-out (LOO) error

Received: 2017-01-17; **Accepted**: 2017-05-12; **Published online**: 2017-06-19 15:53

URL: kns.cnki.net/kcms/detail/11.2625.V.20170619.1553.004.html

Foundation items: National Natural Science Foundation of China (61273275, 61703426)

* **Corresponding author**. E-mail: wang_afeu@126.com

北京航空航天大学学报

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0035

纤维金属层板金属层应变测量及应力预测方法

孟维迎[1,2], 谢里阳[1,2,*], 胡杰鑫[1,2], 吕骁[3], 秦波[1,2], 王博文[1,2]

(1. 东北大学 机械工程与自动化学院, 沈阳 110819; 2. 东北大学 航空动力学装备振动及控制教育部重点实验室, 沈阳 110819;
3. 上海微创医疗器械(集团)有限公司, 上海 201203)

摘 要: 纤维金属层板作为一种新型复合材料,已开始应用于航空航天领域。脱离传统应变测量方法,应用一种新测量方法——数字化光学应变法,实现了层板中金属层应变的测量;同时以子层刚度理论获得层板的等效刚度矩阵,修正经典层板理论中整体刚度矩阵的求解方法,实现了金属层应力的更准确预测。以纤维增强铝锂合金 2/1 及 3/2 层板为例,使用光学应变法测量其金属层应变进而计算金属层应力,利用有限元仿真分析、经典层板理论及修正方法分别对其进行金属层应力预测。通过对比光学应变测量结果和有限元仿真结果,2/1 及 3/2 层板光学应变测量结果与仿真结果最大误差分别为 2.12% 和 3.68%,验证了新测量方法的准确性及实用性;通过对比光学应变测量结果和层板理论预测结果,2/1 及 3/2 层板模型修正后结果比修正前准确率分别提升了 2.91% 和 5.83%,验证了修正模型的有效性及先进性。

关 键 词: 纤维金属层板; 应变测量; 应力预测; 等效刚度; 经典层板理论
中图分类号: V257
文献标识码: A **文章编号**: 1001-5965(2018)01-0142-09

纤维金属层板是一种由合金层与嵌入式固化纤维层以胶结形式交替组成的材料。纤维金属层板由交替的单向、双向或多向纤维增强合金片组成,这些纤维可铺放成不同的方向[1]。相比于常规板(只包括金属或纤维层),纤维金属层板利用金属与纤维层的结合来提高机械性能。由于质量轻、损伤容限高等优点,纤维金属层板被应用于航空结构,代替高强度铝合金[2]。例如,在空中客车 A380 中,为减重和提高损伤容限,目前纤维金属层板已应用在其机身、前缘等部位[3-4]。

纤维金属层板的金属层伴有抵抗裂纹增长的纤维[5]。当疲劳裂纹在金属层中增长时,这些纤维不破坏,限制疲劳裂纹的张开行为。在疲劳裂纹萌生阶段和裂纹扩展初始阶段,纤维桥接机制不存在,即纤维没有限制金属层裂纹张开的作用。Alderliesten[6]研究指出,对于不同层板,纤维开始起作用的临界裂纹长度不同,一般在几毫米到十几毫米之间,取决于层板的结构及层间黏结情况。对于疲劳寿命的研究表明,在桥接机制起作用之前,金属层应力起主要作用;桥接机制起作用之后,金属层应力和桥接应力共同起作用[7-8]。为了预测疲劳寿命,需要考虑金属层板的实际应力水平。而对于多层结构层板(3/2 及以上层板),层板内部金属层受到两侧界面黏结力作用,相应的金属层应力小于外部金属层应力[1],这也是外部金属层首先破坏的原因。故层板金属层应力的研究往往针对的是外部金属层应力。

层板中金属层应力的测定能为纤维金属层板的研制及应用提供重要参考指标,为理论应力求解模型的选取提供参照标准,对于研究纤维金属

收稿日期: 2017-01-18; **录用日期**: 2017-03-17; **网络出版时间**: 2017-04-13 15:08
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170413.1508.003.html
基金项目: 国家自然科学基金 (51335003); 辽宁重大装备制造协同创新中心资助
* **通信作者**. E-mail: lyxie@me.neu.edu.cn

引用格式: 孟维迎, 谢里阳, 胡杰鑫, 等. 纤维金属层板金属层应变测量及应力预测方法[J]. 北京航空航天大学学报, 2018, 44(1): 142-150. MENG W Y, XIE L Y, HU J X, et al. Strain measurement and stress prediction methods of metal layer in fiber metal laminates[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(1): 142-150 (in Chinese).

层板的疲劳寿命问题具有重要意义。然而，应力的测定不能直接通过试验方法实现，通常只能由试验测得材料表面应变情况，进而由理论公式计算得出。目前，国内外针对复合材料层板应变的测量方法较少。Murphy 和 Duke[9] 成功地将信号臂埋置于环氧树脂内，使用光导纤维干涉仪实现了复合材料层板内应力的测量。马子广等[10] 通过在复合材料层板试验件上粘贴应变片的方法，记录了应力-应变曲线。屈蓓[11] 和耿红霞[12] 等分别应用光学应变技术测定了铝板冲击振动情况及带孔铝板的拉伸弹性模量。传统的静态应力测量通过电阻应变片、光弹、钢弦振动、机械传动、液压传递等得出与应力相关的效应，计算应力值。其中应用最广泛的是电阻应变片法。以上测量方法，或使用条件受到限制、或操作繁琐、精度低，往往只能测量局部区域的平均应变情况。

为了估算纤维金属层板的疲劳裂纹萌生寿命及预测纤维金属层板疲劳寿命，实现纤维金属层板金属层应力的预测至关重要。目前国内外对于复合材料层板，均利用经典层板理论及派生方法来实现应力预测。闻柏承和顾震隆[13] 用经典层板理论分析了复合材料层板中的面内应力；甄文强等[14] 在经典层板理论的基础上，考虑到大载荷产生的拉剪损伤耦合效应的影响，实现了平纹编织复合材料偏轴下应力的预测；对于纤维金属层板材料，虽然组分材料不同导致了不同形式的应力分布及变形，但仍可考虑应用经典层板理论进行分析。Spronk 等[15] 和 Homan[1] 引入热膨胀系数的影响，修正经典层板理论，实现纤维金属层板金属层应力预测。经典层板理论及其派生方法中层板整体刚度矩阵的计算是各层刚度根据其组分的体积分数简单相加，均未考虑实际铺层情况及层间相互作用，所求层板整体刚度矩阵与实际情况有一定差异[16]，使得金属层应力的预测存在较大误差。

本文采用光学应变法来实现纤维金属层板中外部金属层应变的测量；在不考虑层板固化产生的残余应力的情况下，利用子层刚度理论来修正经典层板理论，实现层板金属层应力更准确地预测。具体方法是，在经典层板理论的基础上，通过子层刚度理论获得层板整体等效刚度矩阵，对层板理论中的整体刚度矩阵进行修正。

1　光学应变法测量层板金属层应变

1.1　测试方法及原理

光学应变法是借助数字化光学应变仪，利用 2 个高清相机，基于双目立体视觉原理，采用三维数字图像相关方法，可对被测物体表面的三维形貌和载荷作用下的三维变形场进行测量。主要原理如下：首先，利用经过校验的标定板对双目相机系统进行标定，获得 2 个相机的内外参数；其次，利用相关匹配算法，得到左右相机采集图像中对应点的视差，从图像中各点的视差数据和预先获得的标定参数重建物体表面的三维形貌；再次，通过比较载荷作用下测量区域内各点的三维形貌的变化，得到被测物体全场三维位移分布及应变分布；最后，利用层板组分材料的弹性模量转换成试样表面的应力，即纤维金属层板金属层应力。数字化光学应变仪测试系统如图 1 所示。



图 1　数字化光学应变仪测试系统

Fig. 1　Testing system of digital optical strain gauge

1.2　测试方法的优点

光学应变法能实时测量试样关键部位的全局应变，测量精度为 0.01%，可以确定最大应变的位置及其周围应变情况，为试样的应变、应力分析提供可靠的保证。光学应变法相比于常用的传统的应变片测量法，有以下优势：①准备试验时，不损伤试样原表面，并且避免了胶的厚度、固化温度及粘贴压力的影响，保证了试验结果真实性。②应变片最大量程为 4% ~5%，往往因为量程的限制不能测量整个过程的应变，而光学应变法量程为 0.05% ~100%，则避免了此限制。③测量范围大：用同样的传感器测量不同的物体，尺寸范围为 1 ~1 000 mm。④全场的图形化结果，结果包含大量的数据点，高密度的数据点和测量结果的图形化显示，可以帮助更好地了解零件性能。实现最大应变位置准确定位，在一定程度上可以代替并优于软件仿真，因为该方法是实际情况的真实测试。⑤具有先进的后处理模块，能显示各点的主应变方向和等值线以及各种过程参量，并分析计算各点的 X 向应变分量、Y 向应变分量、Mises 应变、厚度减薄率、Tresca 应变、剪切应变、剪切角及各方向位移等。

1.3　测试操作步骤

具体测试过程如下：①用酒精擦拭清洁试样

北航学报 赠阅

表面。②在试样表面喷涂一层亚光白色油漆,然后再用亚光黑色漆,以散点法落于试件表面。③根据所需观察试样的应变范围,选取标定板,标定应变仪。④开始试验,首先使用应变仪拍取基准状态(无载荷状态)试样表面照片,再拍取各加载状态下的试样表面照片,如图 2 所示。⑤结束试验,对应变仪所拍摄的照片进行处理,选取所要观察应变的区域,去除不必要区域。⑥满足合理参考误差前提下设置参考点。⑦分析计算以获得试样表面应变情况,如图 3、图 4 所示。⑧寻找最大应变部位并读取应变值,根据应力应变关系,获得金属层应力。

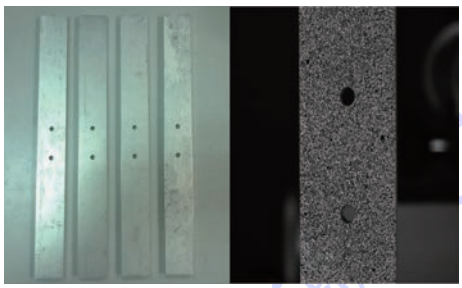


图 2 原始试样及喷涂试样照片

Fig. 2 Photos of original and sprayed specimen

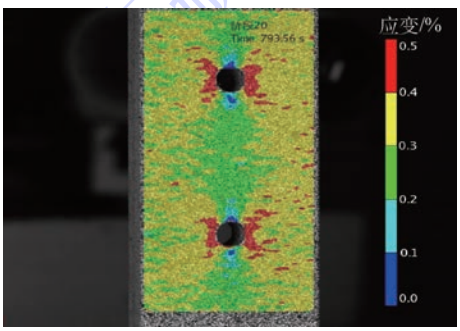


图 3 试样孔心截面 Y 向应变

Fig. 3 Y-direction strain of specimen section through pinhole center

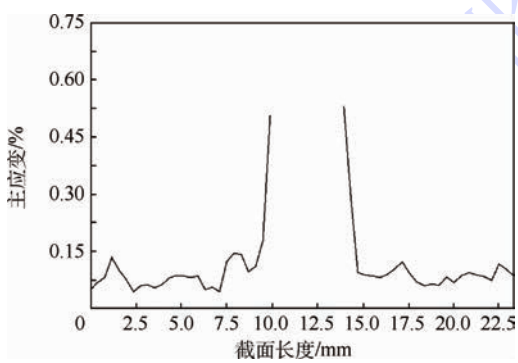


图 4 试样孔心水平截面应变

Fig. 4 Strain of specimen horizontal section through pinhole center

2 层板金属层应力预测模型

2.1 经典层板理论

许多文献对经典层板理论做了概括。通过总结归纳,经典层板理论模型实现主要有 3 个步骤[17]:①各组分材料的本构关系;②层板的整体刚度矩阵;③计算各层材料应力。

2.1.1 各组分材料的本构关系

纤维金属层板结构由相互平行的金属及非金属组分材料交替叠加而成,层板中各组分材料本构关系的通用形式为[18]

$$\begin{pmatrix}\sigma_x\\ \sigma_y\\ \sigma_z\\ \tau_{yz}\\ \tau_{zx}\\ \tau_{xy}\end{pmatrix}=\begin{bmatrix}c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16}\\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26}\\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} & c_{36}\\ c_{41} & c_{42} & c_{43} & c_{44} & c_{45} & c_{46}\\ c_{51} & c_{52} & c_{53} & c_{54} & c_{55} & c_{56}\\ c_{61} & c_{62} & c_{63} & c_{64} & c_{65} & c_{66}\end{bmatrix}\begin{pmatrix}\varepsilon_x\\ \varepsilon_y\\ \varepsilon_z\\ \gamma_{yz}\\ \gamma_{zx}\\ \gamma_{xy}\end{pmatrix} \tag{1}$$

式中:$(\sigma_x \quad \sigma_y \quad \sigma_z \quad \tau_{yz} \quad \tau_{zx} \quad \tau_{xy})^{\mathrm{T}}$ 为应力向量;$(\varepsilon_x \quad \varepsilon_y \quad \varepsilon_z \quad \gamma_{yz} \quad \gamma_{zx} \quad \gamma_{xy})^{\mathrm{T}}$ 为应变向量;刚度矩阵为 $\begin{bmatrix}c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16}\\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26}\\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} & c_{36}\\ c_{41} & c_{42} & c_{43} & c_{44} & c_{45} & c_{46}\\ c_{51} & c_{52} & c_{53} & c_{54} & c_{55} & c_{56}\\ c_{61} & c_{62} & c_{63} & c_{64} & c_{65} & c_{66}\end{bmatrix}$。

在求解实际问题或进行复合材料设计时,经常遇到整体坐标系与局部坐标系不重合的现象。这时,需要利用刚度矩阵坐标转换公式将局部坐标系下的刚度矩阵 $\boldsymbol{c}$ 转换为整体坐标系下的刚度矩阵 $\bar{\boldsymbol{c}}$[19]:

$$\bar{\boldsymbol{c}} = \boldsymbol{T}\boldsymbol{c}\boldsymbol{T}^{\mathrm{T}} \tag{2}$$

式中:

$$\boldsymbol{T}=\begin{bmatrix} l_1^2 & m_1^2 & n_1^2 & 2m_1n_1 & 2n_1l_1 & 2l_1m_1\\ l_2^2 & m_2^2 & n_2^2 & 2m_2n_2 & 2n_2l_2 & 2l_2m_2\\ l_3^2 & m_3^2 & n_3^2 & 2m_3n_3 & 2n_3l_3 & 2l_3m_3\\ l_2l_3 & m_2m_3 & n_2n_3 & m_2n_3+m_3n_2 & n_2l_3+n_3l_2 & l_2m_3+l_3m_2\\ l_3l_1 & m_3m_1 & n_3n_1 & m_3n_1+m_1n_3 & n_3l_1+n_1l_3 & l_3m_1+l_1m_3\\ l_1l_2 & m_1m_2 & n_1n_2 & m_1n_2+m_2n_1 & n_1l_2+n_2l_1 & l_1m_2+l_2m_1\end{bmatrix}$$

其中:l_i, m_i, n_i $(i=1,2,3)$ 为材料的局部坐标系($O123$)和整体坐标系($Oxyz$)中各个坐标轴的方向余弦,见表 1。

北航学报 赠阅

表 1　不同坐标系下坐标轴间的余弦值

Table 1　Cosine value between coordinate axes in different coordinate systems

坐标轴	x	y	z
1	l_1	m_1	n_1
2	l_2	m_2	n_2
3	l_3	m_3	n_3

利用转换公式将层板中各层局部坐标系下的刚度矩阵转换为整体坐标系下的刚度矩阵，其通用形式为

$$[\bar{\boldsymbol{c}}]_k=\begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} \\ d_{41} & d_{42} & d_{43} & d_{44} & d_{45} & d_{46} \\ d_{51} & d_{52} & d_{53} & d_{54} & d_{55} & d_{56} \\ d_{61} & d_{62} & d_{63} & d_{64} & d_{65} & d_{66} \end{bmatrix} \tag{3}$$

式中：k 表示第 k 层材料。

考虑到纤维金属层板各层属性，其刚度矩阵可表示为

$$[\bar{\boldsymbol{c}}]_k=\begin{bmatrix} d_{11} & d_{12} & d_{13} & 0 & 0 & d_{16} \\ d_{21} & d_{22} & d_{23} & 0 & 0 & d_{26} \\ d_{31} & d_{32} & d_{33} & 0 & 0 & d_{36} \\ 0 & 0 & 0 & d_{44} & d_{45} & 0 \\ 0 & 0 & 0 & d_{54} & d_{55} & 0 \\ d_{61} & d_{62} & d_{63} & 0 & 0 & d_{66} \end{bmatrix} \tag{4}$$

2.1.2　整体刚度矩阵

层板的整体刚度矩阵 $\boldsymbol{Q}$ 及柔度矩阵 $\boldsymbol{S}$ 为

$$\boldsymbol{Q}=\sum_{k=1}^{n}\left([\bar{\boldsymbol{c}}]_k\frac{t_k}{t_{\mathrm{lam}}}\right) \tag{5}$$

$$\boldsymbol{S}=\boldsymbol{Q}^{-1} \tag{6}$$

式中：t_k 为第 k 层材料厚度；t_{lam} 为层板总厚度。

2.1.3　计算各层材料压力

外载荷下层合板中面应变为

$$\begin{pmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \\ \gamma_{yz} \\ \gamma_{zx} \\ \gamma_{xy} \end{pmatrix}=\begin{bmatrix} S_{11} & S_{12} & S_{13} & 0 & 0 & S_{16} \\ S_{21} & S_{22} & S_{23} & 0 & 0 & S_{26} \\ S_{31} & S_{32} & S_{33} & 0 & 0 & S_{36} \\ 0 & 0 & 0 & S_{44} & S_{45} & 0 \\ 0 & 0 & 0 & S_{54} & S_{55} & 0 \\ S_{61} & S_{62} & S_{63} & 0 & 0 & S_{66} \end{bmatrix}\begin{pmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{yz} \\ \tau_{zx} \\ \tau_{xy} \end{pmatrix} \tag{7}$$

整体坐标系下张量应变转换为局部坐标系下张量应变，各层材料主方向应变通用形式为

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \gamma_{23} \\ \gamma_{31} \\ \gamma_{12} \end{pmatrix}=\boldsymbol{T}^{\mathrm{T}}\begin{pmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \\ \gamma_{yz} \\ \gamma_{zx} \\ \gamma_{xy} \end{pmatrix} \tag{8}$$

各层材料主方向应力为

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \tau_{23} \\ \tau_{31} \\ \tau_{12} \end{pmatrix}=[\bar{\boldsymbol{c}}]_k\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \gamma_{23} \\ \gamma_{31} \\ \gamma_{12} \end{pmatrix} \tag{9}$$

2.2　修正层板刚度矩阵

在 2.1 节经典层板理论实现过程中，层板整体刚度矩阵的计算是各层刚度根据其组分的体积分数简单相加，未能考虑到实际铺层情况及层间相互作用，造成所求层板整体刚度矩阵与实际情况有一定差异，从而使得金属层应力的预测存在较大误差。为了解决此问题，本文引入等效刚度矩阵的概念来对经典层板理论中整体刚度矩阵的求解方法进行修正。

2.2.1　层板材料性能等效及子层刚度法等效刚度理论

由于复合材料内部结构的多样性和多尺度性（即复合材料层合板是由多层单向板铺放而成，每一层单向板的铺放角度、铺放厚度不同），导致了各层材料在整体坐标系下不同的材料性能，从而使得复合材料整体层合板的材料性能复杂多样[20]。虽然试验方法可以对复合材料的有效性能进行测试，但进行材料力学性能的试验研究给出的一般都是经验公式，受加工条件、制备方法等因素的影响，给出的测试结果往往具有一定的局限性。因此，通过建立力学模型，用理论方法分析和预测复合材料的力学性能，最终把复合材料层合板的性能等效成各向异性材料的性能，对于复合材料结构性能分析及结构优化设计有重要的意义。

Sun 和 Li[16] 基于界面连续性假设，提出的子层刚度法直接从组分的本构关系出发，以整体角度把弹性张量进行统一化处理，进一步应用统计平均思想推导出刚度公式，求解出材料的整体等效刚度矩阵[21]。文献[20]针对正交各向异性复合材料有 3 个正交的弹性性能对称面的特殊结构，以子层刚度法为理论基础，通过应变能密度守恒定律，推导出适用于正交各向异性复合材料的

北航学报 赠阅

等效刚度矩阵：

$$\boldsymbol{C}=\begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{21} & C_{22} & C_{23} & 0 & 0 & 0 \\ C_{31} & C_{32} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{bmatrix} \tag{10}$$

式中：相关刚度系数可根据子层刚度公式求出[16]

$$C_{11}=\sum_{k=1}^{n}V_k d_{11}^{(k)}+\sum_{k=2}^{n}\frac{(d_{13}^{(k)}-\lambda_{13})V_k(d_{13}^{(1)}-d_{13}^{(k)})}{d_{33}^{(k)}}$$

$$C_{12}=\sum_{k=1}^{n}V_k d_{12}^{(k)}+\sum_{k=2}^{n}\frac{(d_{13}^{(k)}-\lambda_{13})V_k(d_{23}^{(1)}-d_{23}^{(k)})}{d_{33}^{(k)}}$$

$$C_{13}=\sum_{k=1}^{n}V_k d_{13}^{(k)}+\sum_{k=2}^{n}\frac{(d_{33}^{(k)}-\lambda_{33})V_k(d_{13}^{(1)}-d_{13}^{(k)})}{d_{33}^{(k)}}$$

$$C_{22}=\sum_{k=1}^{n}V_k d_{22}^{(k)}+\sum_{k=2}^{n}\frac{(d_{23}^{(k)}-\lambda_{23})V_k(d_{23}^{(1)}-d_{23}^{(k)})}{d_{33}^{(k)}}$$

$$C_{23}=\sum_{k=1}^{n}V_k d_{23}^{(k)}+\sum_{k=2}^{n}\frac{(d_{33}^{(k)}-\lambda_{33})V_k(d_{23}^{(1)}-d_{23}^{(k)})}{d_{33}^{(k)}}$$

$$C_{33}=1\Big/\sum_{k=1}^{n}\frac{V_k}{d_{33}^{(k)}}$$

$$C_{44}=\frac{\sum_{k=1}^{n}\frac{V_k d_{44}^{(k)}}{M_k}}{M}$$

$$C_{55}=\frac{\sum_{k=1}^{n}\frac{V_k d_{55}^{(k)}}{M_k}}{M}$$

$$C_{66}=\sum_{k=1}^{n}V_k d_{66}^{(k)}+\sum_{k=2}^{n}\frac{(d_{36}^{(k)}-\lambda_{36})V_k(d_{36}^{(1)}-d_{36}^{(k)})}{d_{33}^{(k)}}$$

$$\lambda_{13}=C_{13}$$

$$\lambda_{23}=C_{23}$$

$$\lambda_{33}=C_{33}$$

$$\lambda_{36}=C_{36}$$

$$M=\left(\sum_{k=1}^{n}\frac{V_k d_{44}^{(k)}}{M_k}\right)\left(\sum_{k-1}^{n}\frac{V_k d_{55}^{(k)}}{M_k}\right)-\left(\sum_{k-1}^{n}\frac{V_k d_{45}^{(k)}}{M_k}\right)^2$$

$$M_k=d_{44}^{(k)}d_{55}^{(k)}-(d_{45}^{(k)})^2$$

其中：V_k 为第 k 层材料的体积分数。

2.2.2 修正经典层板理论

通过子层刚度法理论求得层板的等效刚度矩阵后，代入到经典层板理论中。修正后的经典层板理论如下：①各组分材料在整体坐标系下的刚度矩阵，通过式(1)、式(2)、式(4)求解；②采用改进的子层刚度理论计算层板的等效刚度矩阵，见式(10)；③等效刚度矩阵的逆矩阵式(6)代入到式(7)，得到层板中面应变

$$\begin{pmatrix}\varepsilon_x\\ \varepsilon_y\\ \varepsilon_z\\ \gamma_{yz}\\ \gamma_{zx}\\ \gamma_{xy}\end{pmatrix}=\begin{bmatrix} A_{11} & A_{12} & A_{13} & 0 & 0 & 0 \\ A_{21} & A_{22} & A_{23} & 0 & 0 & 0 \\ A_{31} & A_{32} & A_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & A_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & A_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & A_{66} \end{bmatrix}\begin{pmatrix}\sigma_x\\ \sigma_y\\ \sigma_z\\ \tau_{yz}\\ \tau_{zx}\\ \tau_{xy}\end{pmatrix} \tag{11}$$

将式(11)代入到式(8)、式(9)中，即可获得所求金属层应力。

3 实例验证

为了验证该光学应变法测量的准确性及预测模型的有效性，本文选取 2060 铝锂合金 2/1 结构及 3/2 结构的纤维金属层板进行加载试验。采用平板开孔试件，试件沿材料纵向取材，形式如图 5 所示，层板结构如图 6 所示，其组分材料属性见表 2。其中 2/1 及 3/2 层板应力集中系数分别为 2.57 和 2.56。本文以层板受单向拉伸载荷情况为例。考虑到研究目标为材料的线弹性性能，边界条件为分别对层板施加 50、60、70、80、90、100 MPa 远程应力。分别利用光学应变法对其

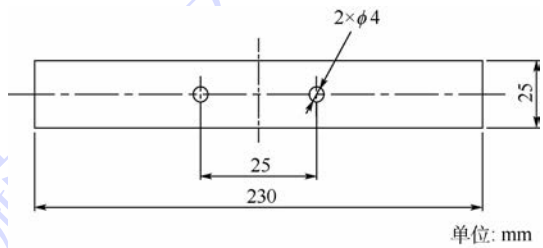


图 5 铝锂合金层板试件

Fig. 5 Test specimen of Al-Li alloy laminates

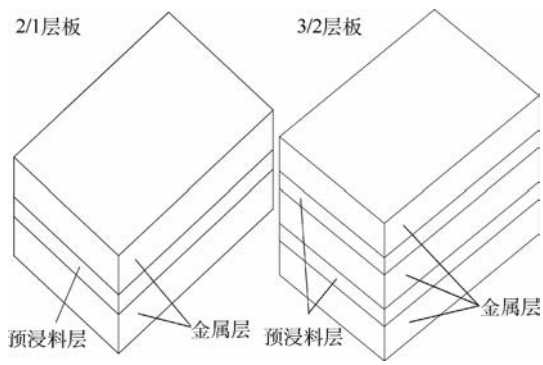


图 6 2/1 及 3/2 层板结构

Fig. 6 Structure of 2/1 and 3/2 laminates

北航学报 赠阅

表 2　层板组分材料属性及尺寸

Table 2　Material properties and sizes of laminates' components

材　料	弹性模量/GPa	泊 松 比	厚度/mm
金属层	72.4	0.3	1.9
预浸料层	54.6	0.252	0.9

进行金属层应变测量，使用 ANSYS 软件对其进行有限元仿真分析，采用经典层板理论及修正方法对其进行金属层应力预测。由于其试件所受载荷为试件长度方向，该方向应力、应变作用最大，则后文所指的金属层应力应变均为试件长度方向应力应变。

3.1　有限元仿真分析

以试件为研究对象，本文通过商用有限元软件 ANSYS 来实现有限元仿真分析。对于纤维金属层板试样各组分厚度影响其所受应力情况，采用三维实体建模方式，单元类型为 SOLID185。由于主要研究对象为孔周部位，故对孔周边进行网格细化处理。层板厚度方向网格划分情况为：每层金属层划分成 4 个网格，每层纤维层划分成 2 个网格。其每层材料单元的材料属性赋予情况如表 2 所示。为了模拟实际作用效果，在试件长度方向一端的底面施加全约束，另一端的 2 个侧平面施加相应载荷。其有限元模型如图 7 ~ 图 9 所示。通过有限元仿真分析，获得不同应力下层板金属层的应变情况，见图 10。

3.2　光学应变法测量层板金属层应力

使用数字化光学应变测量仪，按照上述介绍的光学应变法试验步骤，分别测量 2/1 层板及3/2 层板试样不同远程应力下的应变，取最大应变部位（孔周部位）为研究对象，其试验过程见图 2 ~ 图 4。由光学应变法获得孔周的应变情况，提取各应力下的最大应变。测试获得的应变数据根

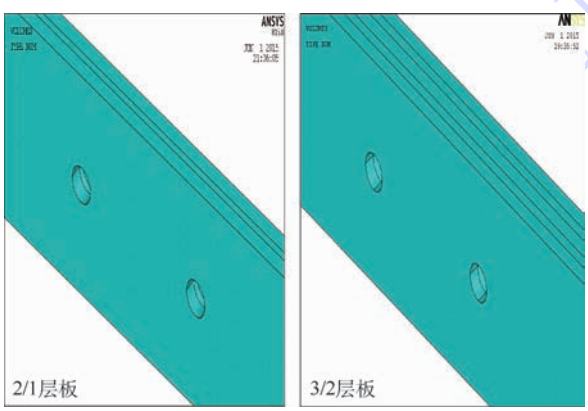


图 7　2/1 及 3/2 层板几何模型

Fig. 7　Geometric models of 2/1 and 3/2 laminates

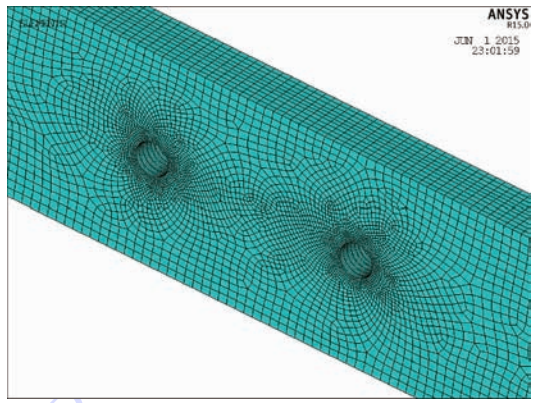


图 8　有限元模型网格划分情况

Fig. 8　Meshing case of finite element model

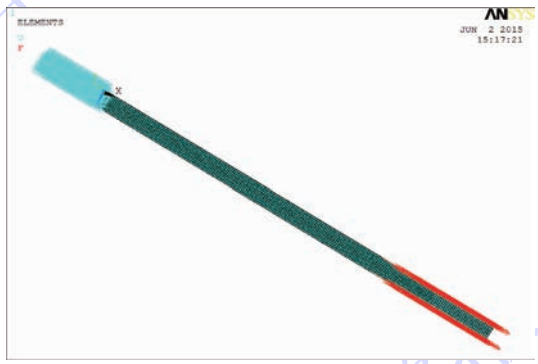


图 9　有限元模型边界条件

Fig. 9　Boundary conditions of finite element model

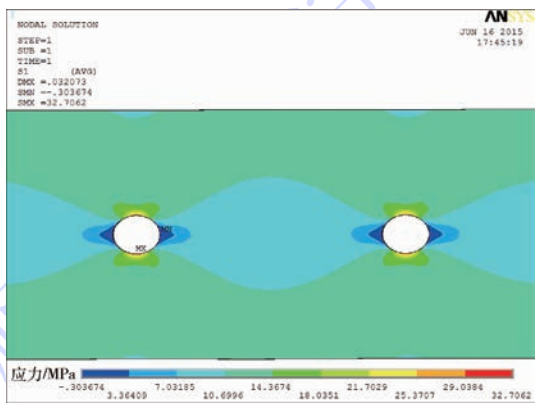


图 10　有限元模型应力分析结果

Fig. 10　Stress analysis results of finite element model

据胡克定律转换成应力张量，并对该试验数据进行线性拟合。2/1 层板及 3/2 层板各应力下最大应力的有限元仿真结果及应变测试结果如图 11（a）、图 12（a）所示。

3.3　数学模型预测层板金属层应力

分别采用经典层板理论及修正后金属层应力预测模型，对 2/1 层板及 3/2 层板试件在不同远程应力下的金属层应力进行预测，考虑应力集中影响，所得结果如图 11（b）、图 12（b）所示。

北航学报 赠阅

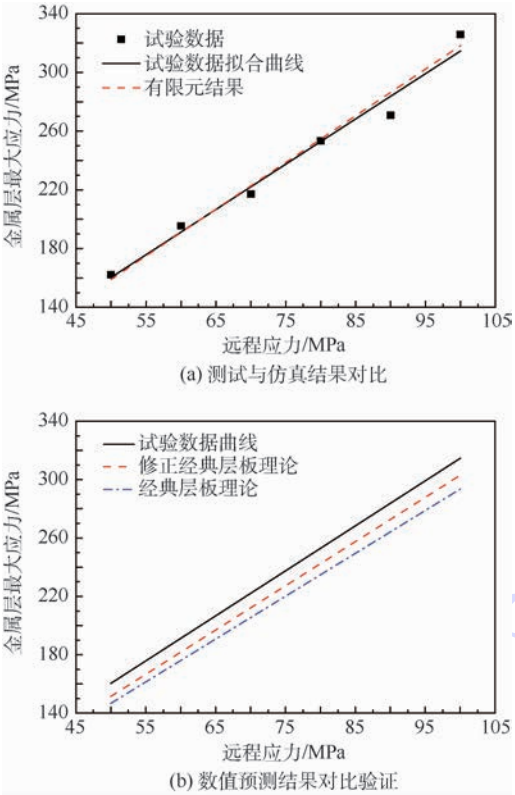


图 11 2/1 层板金属层最大应力测量及预测结果

Fig. 11 Measurement and prediction results of maximum stress in metal layer of 2/1 laminates

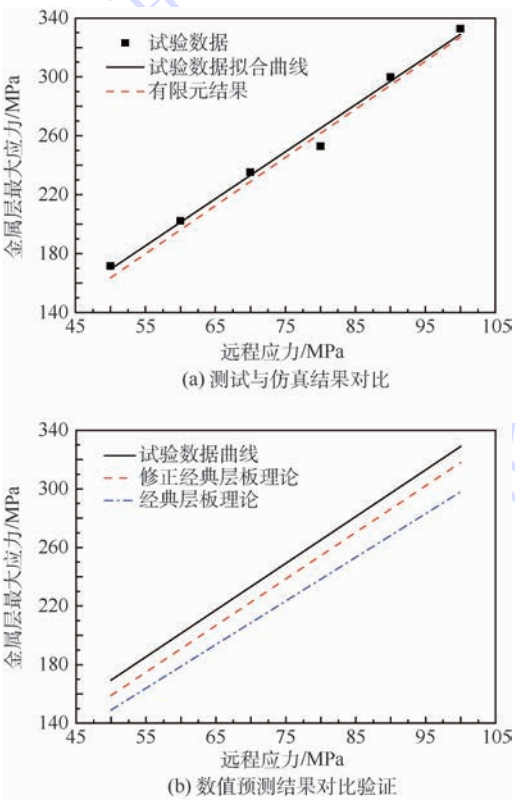


图 12 3/2 层板金属层最大应力测量及预测结果

Fig. 12 Measurement and prediction results of maximum stress in metal layer of 3/2 laminates

3.4 结果对比分析

由图 11(a)、图 12(a)可知,数字化光学应变法结果与有限元仿真结果较吻合,2/1 层板及 3/2 层板最大误差分别为 2.12% 和 3.68%,从而证明了该测试方法的准确性及实用性。由图 11(b)、图 12(b)可知通过经典层板理论预测结果与光学应变结果比较,2/1 层板及 3/2 层板最大误差分别为 9.91% 和 13.15%;修正后的层板理论预测结果较经典层板理论与光学应变结果更接近,2/1 层板及 3/2 层板最大误差分别为 7% 和 7.32%,准确率分别提升了 2.91% 和 5.83%,证明了该修正模型的有效性和先进性。

4 结 论

1) 数字化光学应变法实现了纤维金属层板金属层应力的测量。通过实例应用及对比有限元仿真结果,分析表明:2/1 层板及 3/2 层板最大误差分别为 2.12% 和 3.68%,验证了该方法的准确性。

2) 利用子层刚度理论修正经典层板理论,实现了纤维金属层板金属层应力更准确的预测。通过实例分析并对比光学应变测量结果及经典层板理论预测结果,研究表明:2/1 层板及 3/2 层板预测准确率分别提升了 2.91% 和 5.83%,验证了该模型的有效性、先进性。

参考文献(References)

[1] HOMAN J J. Fatigue initiation in fiber metal laminates[J]. International Journal of Fatigue, 2006, 28(4): 366-374.

[2] 郭亚军,吴学仁. 纤维金属层板疲劳裂纹扩展速率与寿命预测的唯象模型[J]. 航空学报, 1998, 19(3): 275-282.
GUO Y J, WU X R. Phenomenological model for predicting fatigue crack growth in fiber reinforced metal laminates[J]. Acta Aeronautica et Astronautica Sinica, 1998, 19(3): 275-282(in Chinese).

[3] CHANG P Y, YEH P C, YANG J M. Fatigue crack initiation in hybrid boron/glass/aluminum fiber metal laminates[J]. Materials Science and Engineering A, 2008, 496: 273-280.

[4] FRIZZELL R M, MCCRATHY C T, MCCARTHY M A. An experimental investigation into the progression of damage in pin-loaded fiber metal laminates[J]. Composites Part B: Engineering, 2008, 39(6): 907-925.

[5] MARISSEN R. Fatigue crack growth in ARALL: A hybrid aluminum-aramid composite material crack growth mechanisms and quantitative predictions of the crack growth rates[D]. Delft: Delft University of Technology, 1988.

[6] ALDERLIESTEN R C. Development of an empirical fatigue crack growth prediction model for the fibre metal laminate glare[D]. Delft: Delft University of Technology, 1999.

北航学报 赠阅

[7] 郭亚军. 纤维金属层板的疲劳损伤与寿命预测[D]. 北京：北京航空材料研究院，1997.
GUO Y J. Fatigue damage and life prediction of fiber reinforced metal laminates[D]. Beijing: Beijing Institute of Aeronautical Materials, 1997(in Chinese).

[8] KIEBOOM O. Fatigue crack initiation and early crack growth in glare at different temperatures[D]. Delft: Delft University of Technology, 2000.

[9] MURPHY K, DUKE J. A rugged optical fiber interferometer for strain measurements inside a composite material laminate[J]. Journal of Composites, Technology and Research, 1988, 10(1): 11-15.

[10] 马子广，陈庆童，王卫卫. 复合材料层压板开孔拉伸力学性能探究[J]. 直升机技术，2015(1)：64-69.
MA Z G, CHEN Q T, WANG W W. The exploring of composite materials' mechanical properties under tensile loading[J]. Helicopter Technique, 2015(1): 64-69(in Chinese).

[11] 屈蓓，付小龙，何俊武，等. 非接触式光学应变测量技术研究进展[J]. 计测技术，2013，33(5)：10-15.
QU B, FU X L, HE J W, et al. Research on optical non-contact strain measurement[J]. Metrology & Measurement Technology, 2013, 33(5): 10-15(in Chinese).

[12] 耿红霞，付朝华，蒋小林，等. 光学应变测量系统在研究生实验教学中的应用[J]. 实验室研究与探索，2015，34(3)：220-224.
GENG H X, FU C H, JIANG X L, et al. Application of optical strain measurement system in experiment teaching for postgraduates[J]. Re-search and Exploration in Laboratory, 2015, 34(3): 220-224(in Chinese).

[13] 闻柏承，顾震隆. 带或不带脱层的复合材料层板强度的预报[J]. 复合材料学报，1989，2(1)：48-54.
WEN B C, GU Z L. Probabilistic prediction of laminates with or without delamination[J]. Acta Materiae Compositae Sinica, 1989, 2(1): 48-54(in Chinese).

[14] 甄文强，王波，李潘，等. 平面编织 C/SiC 复合材料层合板偏轴拉伸性能研究[J]. 机械强度，2014，36(1)：856-861.
ZHEN W Q, WANG B, LI P, et al. Study of off-axis tensile properties of plain-woven C/SiC composites[J]. Journal of Mechanical Strength, 2014, 36(1): 856-861(in Chinese).

[15] SPRONK S W F, SEN I, ALEDRLIESTEN R C. Predicting fatigue crack initiation in fibre metal laminates based on metal fatigue test data[J]. International Journal of Fatigue, 2015, 70: 428-439.

[16] SUN C T, LI S J. Three dimensional effective elastic constants for thick laminates[J]. Composite Material, 1988, 22(7): 629-639.

[17] BRUNBAUER J, PINTER G. Fatigue life prediction of carbon fibre reinforced laminates by using cycle-dependent classical laminate theory[J]. Composites Part B: Engineering, 2015, 70: 167-174.

[18] 张振瀛. 复合材料力学基础[M]. 北京：航空工业出版社，1989：10.
ZHANG Z Y. Mechanical basis of composite[M]. Beijing: Aviation Industry Press, 1989: 10(in Chinese).

[19] 赵美英，陶梅贞. 复合材料结构力与结构设计[M]. 西安：西北工业大学出版社，2007：12.
ZHAO M Y, TAO M Z. Structural mechanics and structural design of composite[M]. Xi'an: Northwestern Polytechnical University Press, 2007: 12(in Chinese).

[20] 孟维迎. 大型风机叶片复合纤维性能等效算法研究及应用[D]. 呼和浩特：内蒙古工业大学，2013.
MENG W Y. Research and application of performance equivalence agorithm of composite for large scale wind turbine blade[D]. Huhhot: Inner Mongolia University of Technology, 2013(in Chinese).

[21] 孙鹏文，岳彩宾，张兰挺，等. 基于子层刚度法性能等效的风机叶片分析研究[J]. 太阳能学报，2013，34(11)：1974-1977.
SUN P W, YUE C B, ZHANG L T, et al. Research on wind turbine blade analysis of composite equivalent elastic constants based on sub-laminate stiffness method[J]. Acta Energiae Solaris Sinica, 2013, 34(11): 1974-1977(in Chinese).

作者简介：

孟维迎　男，博士研究生。主要研究方向：纤维金属层板疲劳寿命性能。

谢里阳　男，博士，教授，博士生导师。主要研究方向：机械疲劳强度与可靠性。

北航学报 赠阅

Strain measurement and stress prediction methods of metal layer in fiber metal laminates

MENG Weiying[1,2], XIE Liyang[1,2,*], HU Jiexin[1,2], LYU Xiao[3], QIN Bo[1,2], WANG Bowen[1,2]

(1. School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China;
2. Key Laboratory of Vibration and Control of Aero-propulsion Systems Ministry of Education, Northeastern University, Shenyang 110819, China;
3. Shanghai MicroPort Medical (Group) Co., Ltd., Shanghai 201203, China)

Abstract: Fiber metal laminates, as a new-type composite material, have been applied in aerospace field. Digital optical strain method is used to realize strain measurement of metal layer instead of traditional method of strain measurement. Meanwhile, in order to predict the stress distribution in metal layer accurately, the global stiffness matrix obtained from classic laminate theory is modified by the equivalent stiffness matrix from sub-laminate stiffness theory. Taking 2/1 and 3/2 laminates of glass fiber reinforced Al-Li alloy as an example, the stress distribution in metal layer of the laminates is determined based on the measured strain, finite element analysis, classical laminate theory and modified method. The comparison of stress distributions obtained from the measured strain and finite element analysis shows that the maximum errors are only 2.12% and 3.68% for 2/1 and 3/2 laminates, respectively, which verifies the accuracy and practicability of the optical strain method. By comparing the stress distributions from the optical strain method and laminate theory, the prediction accuracy of the modified model increases by 2.91% and 5.83% compared with that of original model for 2/1 and 3/2 laminates, respectively, which proves the effectiveness and advancement of the modified model.

Keywords: fiber metal laminates; strain measurement; stress prediction; equivalent stiffness; classical laminate theory

Received: 2017-01-18; **Accepted**: 2017-03-17; **Published online**: 2017-04-13 15:08

URL: kns.cnki.net/kcms/detail/11.2625.V.20170413.1508.003.html

Foundation items: National Natural Science Foundation of China (51335003); Subsidized by the Collaborative Innovation Center of Major Machine Manufacturing in Liaoning

* **Corresponding author**. E-mail: lyxie@me.neu.edu.cn

北京航空航天大学学报

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0038

基于傅里叶域卷积表示的目标跟踪算法

朱日东，杨小远*，王敬凯

（北京航空航天大学 数学与系统科学学院，北京 100083）

摘 要：针对目标跟踪问题，提出基于傅里叶域卷积表示的目标跟踪算法，将目标跟踪问题转化为卷积表示模型，通过求解最优滤波器，得到对目标函数的最佳表示，可以实现快速鲁棒的跟踪。多通道卷积表示模型在傅里叶域等价于求解线性方程的最佳近似解。首先，通过广义逆理论求得该方程的最优通解，给出一般滤波器的表示形式；然后，利用前一时刻的滤波器和当前特征模板生成当前滤波器，利用满秩算法快速求解广义逆；最后，在位移和尺度上更新、应用该滤波器。在目标跟踪基准（OTB）数据库中的大量实验表明，本文算法比当前部分较为先进的跟踪算法具有更好的表现，并提供了更加灵活多样的滤波器设计。

关 键 词：目标跟踪；卷积表示；Moore-Penrose 广义逆；傅里叶变换；最佳逼近
中图分类号：O29；TP391.4
文献标识码：A **文章编号**：1001-5965(2018)01-0151-09

目标跟踪是计算机视觉领域中的一个重要问题，在过去十多年中一直是计算机视觉领域中非常活跃的研究内容。目标跟踪的过程是：在初始图像帧中给出特定目标的初始状态，然后估计该目标在后续各图像帧中的状态。目标跟踪在实际生活中有广泛的应用，如人机互动、智能监控、视频检索和自动驾驶等[1-2]。目标跟踪技术尽管已经取得了很大进步，但是设计出对各种场景都能做到实时稳定跟踪的算法仍然是一项很大的挑战。这是因为实际中的场景十分复杂，跟踪过程中会遇到各种问题，例如光照变化、形态变化、尺度变化、遮挡问题以及背景相似等，这些因素使得跟踪过程不稳定，增加了跟踪难度。

最近卷积表示开始应用到目标跟踪并实现了非常好的跟踪效果，在跟踪速度和精度上都得到了很大的提升，吸引了国内外学者的关注。Bolme 等利用最小输出误差平方和求解滤波器，提出 MOSSE 跟踪法[3]，该算法是基于单通道特征多样本训练的跟踪算法，通过少量样本训练实现了高速跟踪。Henriques 等提出的 KCF 跟踪器[4]，是利用核函数矩阵的循环结构，通过核化的岭回归函数求解卷积滤波器，实现了快速的目标检测和跟踪。Danelljan 等通过多通道特征单样本卷积表示求解滤波器，并将此结构应用到尺度分析上，提出了 DSST 跟踪算法[5]，将位移和尺度上的跟踪同时进行，提高了跟踪的稳定性。

注意到大多数基于卷积的目标跟踪算法都是通过正则化约束求解卷积滤波器的近似解[5-7]，对于卷积表示无正则化模型很少涉及。卷积表示无正则化模型是求解卷积表示问题的精确解，该模型经过离散傅里叶变换可以得到若干个相互独立的子问题。子问题可以通过 Moore-Penrose 广义逆进行求解，进而得到模型的通解，给出滤波器的一般设计算法，最后利用跟踪过程的前后连续性迭代生成滤波器，从而得到适合不同场景的滤波器。其中，广义逆可以通过满秩算法快速求解。

收稿日期：2017-01-18；**录用日期**：2017-04-07；**网络出版时间**：2017-05-16 14:30
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170516.1430.001.html
基金项目：国家自然科学基金（61671002）；北京市自然科学基金（4152029）
* **通信作者**. E-mail：xiaoyuanyang@vip.163.com

引用格式：朱日东，杨小远，王敬凯. 基于傅里叶域卷积表示的目标跟踪算法[J]. 北京航空航天大学学报，2018，44(1)：151-159.
ZHU R D，YANG X Y，WANG J K. Convolution representation-based object tracking algorithm in Fourier domain[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(1)：151-159 (in Chinese).

北航学报 赠阅

在此基础上,提出一种在傅里叶域通过广义逆理论求解卷积表示模型的目标跟踪算法,实验结果表明该算法比部分较为先进的跟踪算法具有更好的跟踪效果。

1 Moore-Penrose 广义逆矩阵

定义 1[8] 设 $\boldsymbol{A}\in\mathbf{C}^{n\times m}$,如果存在 $\boldsymbol{X}\in\mathbf{C}^{m\times n}$,满足:$\boldsymbol{AXA}=\boldsymbol{A}$,$\boldsymbol{XAX}=\boldsymbol{X}$,$(\boldsymbol{AX})^{*}=\boldsymbol{AX}$,$(\boldsymbol{XA})^{*}=\boldsymbol{XA}$,则称 $\boldsymbol{X}$ 是矩阵 $\boldsymbol{A}$ 的 Moore-Penrose 广义逆矩阵,记为 $\boldsymbol{A}^{\dagger}$。

其中:“*”表示矩阵的共轭转置。

引理 1[9] 设 $\boldsymbol{A}\in\mathbf{C}^{n\times m}$ 的秩为 k,则存在列满秩矩阵 $\boldsymbol{L}$ 和行满秩矩阵 $\boldsymbol{R}$,使得 $\boldsymbol{A}=\boldsymbol{LR}$。

定理 1[8] 设 $\boldsymbol{A}\in\mathbf{C}^{n\times m}$,$\boldsymbol{A}$ 的秩为 k,则

$$\boldsymbol{A}^{\dagger}=\boldsymbol{R}^{\dagger}\boldsymbol{L}^{\dagger}=\boldsymbol{R}^{*}(\boldsymbol{RR}^{*})^{-1}(\boldsymbol{L}^{*}\boldsymbol{L})^{-1}\boldsymbol{L}^{*}$$

推论 1 设 $\boldsymbol{A}\in\mathbf{C}^{1\times m}$,$\boldsymbol{A}\neq\boldsymbol{0}$,则

$$\boldsymbol{A}^{\dagger}=\frac{\boldsymbol{A}^{*}}{\|\boldsymbol{A}\|^{2}}$$

定理 2 线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最佳近似解为 $\bar{\boldsymbol{x}}=\boldsymbol{y}-\boldsymbol{A}^{\dagger}(\boldsymbol{Ay}-\boldsymbol{b})$

其中:$\boldsymbol{A}\in\mathbf{C}^{1\times m}$,$\boldsymbol{A}\neq\boldsymbol{0}$;$\boldsymbol{x}\in\mathbf{C}^{m\times 1}$;$\boldsymbol{b}\in\mathbf{C}$;$\boldsymbol{y}\in\mathbf{C}^{m\times 1}$。

证明 因为 $\boldsymbol{Ax}-\boldsymbol{b}=\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{A}^{\dagger}\boldsymbol{b})+(\boldsymbol{AA}^{\dagger}-\boldsymbol{I})\boldsymbol{b}$,其中:$\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{A}^{\dagger}\boldsymbol{b})$ 和 $(\boldsymbol{AA}^{\dagger}-\boldsymbol{I})\boldsymbol{b}$ 垂直。

因为

$$((\boldsymbol{AA}^{\dagger}-\boldsymbol{I})\boldsymbol{b})^{*}\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{A}^{\dagger}\boldsymbol{b})=\boldsymbol{b}^{*}(\boldsymbol{AA}^{\dagger}-\boldsymbol{I})^{*}\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{A}^{\dagger}\boldsymbol{b})=\boldsymbol{b}^{*}((\boldsymbol{AA}^{\dagger})^{*}-\boldsymbol{I})\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{A}^{\dagger}\boldsymbol{b})$$

由广义逆的定义知

$$(\boldsymbol{AA}^{\dagger})^{*}=\boldsymbol{AA}^{\dagger}$$

$$\boldsymbol{AA}^{\dagger}\boldsymbol{A}=\boldsymbol{A}$$

所以

$$\boldsymbol{b}^{*}((\boldsymbol{AA}^{\dagger})^{*}-\boldsymbol{I})\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{A}^{\dagger}\boldsymbol{b})=\boldsymbol{b}^{*}(\boldsymbol{AA}^{\dagger}-\boldsymbol{I})\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{A}^{\dagger}\boldsymbol{b})=\boldsymbol{b}^{*}(\boldsymbol{AA}^{\dagger}\boldsymbol{A}-\boldsymbol{A})(\boldsymbol{x}-\boldsymbol{A}^{\dagger}\boldsymbol{b})=0$$

所以 $\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{A}^{\dagger}\boldsymbol{b})$ 和 $(\boldsymbol{AA}^{\dagger}-\boldsymbol{I})\boldsymbol{b}$ 垂直。

而 $(\boldsymbol{AA}^{\dagger}-\boldsymbol{I})\boldsymbol{b}=(\boldsymbol{AA}^{\dagger}-\boldsymbol{I})(\boldsymbol{b}-\boldsymbol{Ay})$,由勾股定理得

$$\|\boldsymbol{Ax}-\boldsymbol{b}\|^{2}=\|\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{A}^{\dagger}\boldsymbol{b})\|^{2}+\|(\boldsymbol{AA}^{\dagger}-\boldsymbol{I})(\boldsymbol{b}-\boldsymbol{Ay})\|^{2}\geqslant\|(\boldsymbol{AA}^{\dagger}-\boldsymbol{I})(\boldsymbol{b}-\boldsymbol{Ay})\|^{2}$$

而

$$\|(\boldsymbol{AA}^{\dagger}-\boldsymbol{I})(\boldsymbol{b}-\boldsymbol{Ay})\|^{2}=\|\boldsymbol{AA}^{\dagger}\boldsymbol{b}-\boldsymbol{AA}^{\dagger}\boldsymbol{Ay}+\boldsymbol{Ay}-\boldsymbol{b}\|^{2}=\|\boldsymbol{A}(\boldsymbol{y}-\boldsymbol{A}^{\dagger}(\boldsymbol{Ay}-\boldsymbol{b}))-\boldsymbol{b}\|^{2}$$

因此

$$\|\boldsymbol{Ax}-\boldsymbol{b}\|^{2}\geqslant\|\boldsymbol{A}(\boldsymbol{y}-\boldsymbol{A}^{\dagger}(\boldsymbol{Ay}-\boldsymbol{b}))-\boldsymbol{b}\|^{2}$$

所以 $\bar{\boldsymbol{x}}=\boldsymbol{y}-\boldsymbol{A}^{\dagger}(\boldsymbol{Ay}-\boldsymbol{b})$ 是最佳近似解。 **证毕**

定理 2 的应用举例:求解方程 $3x_1+5x_2+2x_3+7x_4+x_5=1$。

令 $\boldsymbol{a}=[3,5,2,7,1]$,$\boldsymbol{x}=\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\end{bmatrix}$,方程可表示为 $\boldsymbol{ax}=1$。方程的通解为 $\boldsymbol{x}=\boldsymbol{y}-\frac{\boldsymbol{a}^{*}}{\boldsymbol{aa}^{*}}(\boldsymbol{ay}-1)$,$\boldsymbol{y}$ 为任意 $\mathbf{C}^{5\times 1}$ 中的向量。

如取 $\boldsymbol{y}=\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}$,则对应的解为

$$\boldsymbol{x}_0=\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}-\frac{[3,5,2,7,1]^{*}}{[3,5,2,7,1][3,5,2,7,1]^{*}}\cdot\left([3,5,2,7,1]\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}-1\right)=\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}-\frac{17}{88}\begin{bmatrix}3\\5\\2\\7\\1\end{bmatrix}=\begin{bmatrix}37/88\\3/88\\54/88\\-31/88\\71/88\end{bmatrix}$$

将此解代入方程,可以验证它为方程的解。

2 卷积表示模型

卷积表示模型是用一组模板 $\boldsymbol{D}=\{\boldsymbol{D}_1,\boldsymbol{D}_2,\cdots,\boldsymbol{D}_K\}$ 和滤波器 $\bar{\boldsymbol{X}}=\{\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_K\}$,通过卷积运算表示目标函数 $\boldsymbol{G}$,即

$$\boldsymbol{G}=\sum_{k=1}^{K}\boldsymbol{D}_k*\boldsymbol{X}_k \tag{1}$$

式中:$*$ 为卷积符号;$\boldsymbol{D}_k\in\mathbf{R}^{N\times M}$;$\boldsymbol{X}_k\in\mathbf{R}^{N\times M}$;$\boldsymbol{G}\in\mathbf{R}^{N\times M}$。

由循环卷积的性质,当模板发生平移时,目标函数也会发生同样的平移,此时只需根据目标函数的最大值,即可确定目标的位置,如图 1 所示。因此该模型非常适合目标运动的检测。为此,需要找到一组滤波器,使其满足式(1)。因此建立

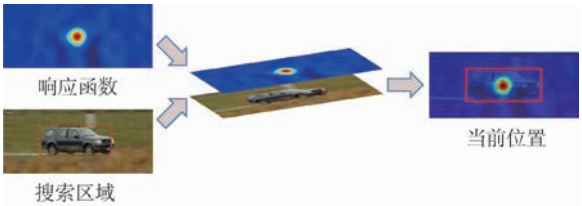


图 1　响应函数预测位置

Fig. 1　Location estimation by response function

以下模型:

$$\arg\min_{\{\boldsymbol{X}_k\}}\left\|\sum_{k=1}^{K}\boldsymbol{D}_k * \boldsymbol{X}_k - \boldsymbol{G}\right\|_{\mathrm{F}}^2 \tag{2}$$

为了便于分析和求解,该问题通常改为带正则化项的模型[5-7],即

$$\arg\min_{\{\boldsymbol{X}_k\}}\left\|\sum_{k=1}^{K}\boldsymbol{D}_k * \boldsymbol{X}_k - \boldsymbol{G}\right\|_{\mathrm{F}}^2 + \lambda\sum_{k=1}^{K}\|\boldsymbol{X}_k\|_{\mathrm{F}}^2 \tag{3}$$

式中:λ 为正则化权重参数。

类似模型已经被用于目标跟踪[5]、目标检测[6]和目标配准[7]等。

正则化项是为了求得在预定条件下的确定解,这种方法便于问题的分析与求解,但是可以看出满足模型式(3)的解是原问题式(2)的近似解。为了更好地分析原问题,本文将研究无正则化项的模型式(2)。

因为卷积运算计算量特别大,而在傅里叶变换域可以转化为计算量少的点积运算,所以将该问题转到傅里叶变换域。由 Parseval's 等式可知,式(2)等价于

$$\arg\min_{\{\hat{\boldsymbol{X}}_k\}}\frac{1}{NM}\left\|\sum_{k=1}^{K}\hat{\boldsymbol{D}}_k \odot \hat{\boldsymbol{X}}_k - \hat{\boldsymbol{G}}\right\|_{\mathrm{F}}^2 \tag{4}$$

式中:$\odot$ 为点积运算;$\hat{\boldsymbol{D}}_k$、$\hat{\boldsymbol{X}}_k$、$\hat{\boldsymbol{G}}$ 为 $\boldsymbol{D}_k$、$\boldsymbol{X}_k$、$\boldsymbol{G}$ 的傅里叶变换矩阵。

由弗罗贝尼乌斯范数的性质可知,问题式(4)等价于求解 $N\times M$ 个子问题

$$\arg\min_{\{\hat{x}_k\}}\frac{1}{NM}\left\|\sum_{k=1}^{K}\hat{d}_k\hat{x}_k - \hat{g}\right\|_{\mathrm{F}}^2 \tag{5}$$

式中:$\hat{d}_k=(\hat{\boldsymbol{D}}_k)_{ij}$;$\hat{x}_k=(\hat{\boldsymbol{X}}_k)_{ij}$;$\hat{g}=\hat{\boldsymbol{G}}_{ij}$。

最后模型可以表示为

$$\arg\min_{\{\hat{x}\}}(\hat{\boldsymbol{d}}\hat{\boldsymbol{x}} - \hat{g})^2 \tag{6}$$

式中:$\hat{\boldsymbol{d}}=[\hat{d}_1,\hat{d}_2,\cdots,\hat{d}_K]$;$\hat{\boldsymbol{x}}=[\hat{x}_1,\hat{x}_2,\cdots,\hat{x}_K]^{\mathrm{T}}$。

求解问题式(6)最常用的方法是通过凸优化的方法,即对优化函数求导使其等于零。但是这种方法会使得问题变得复杂,由定理 2 可知,该问题的最优解为

$$\boldsymbol{y} - \hat{\boldsymbol{d}}^{\dagger}(\hat{\boldsymbol{d}}\boldsymbol{y} - \hat{g}) = \arg\min_{\{\hat{x}\}}(\hat{\boldsymbol{d}}\hat{\boldsymbol{x}} - \hat{g})^2 \tag{7}$$

在本文问题中,$\hat{\boldsymbol{d}}$ 奇异,该问题有无穷解,因此要在这些解中寻找最适合目标跟踪的解。从解的形式看,该解性质比较好,可以通过给定不同的 $\boldsymbol{y}$ 得到满足要求的解,或通过迭代得到满足给定条件的解。例如,最易想到的是 $\boldsymbol{y}=\boldsymbol{0}$ 时所得的解,即 $\hat{\boldsymbol{x}}=\hat{\boldsymbol{d}}^{\dagger}\hat{g}$,由广义逆理论知,此解为该问题的最小范数解。

下面根据目标跟踪过程中目标特征在前后帧中相似的特点,结合滤波器的更新,给出以下设计滤波器的方法:

$$\hat{\boldsymbol{x}}^1 = (\hat{\boldsymbol{d}}^1)^{\dagger}\hat{g} \tag{8}$$

$$\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}^{n-1} - (\hat{\boldsymbol{d}}^n)^{\dagger}(\hat{\boldsymbol{d}}^n\hat{\boldsymbol{x}}^{n-1} - \hat{g}) \tag{9}$$

$$\hat{\boldsymbol{x}}^n = (1-\alpha)\hat{\boldsymbol{x}}^{n-1} + \alpha\hat{\boldsymbol{x}} \tag{10}$$

式中:α 为学习率。

由于

$$\begin{aligned}\hat{\boldsymbol{x}}^n &= (1-\alpha)\hat{\boldsymbol{x}}^{n-1} + \alpha[\hat{\boldsymbol{x}}^{n-1} - (\hat{\boldsymbol{d}}^n)^{\dagger}(\hat{\boldsymbol{d}}^n\hat{\boldsymbol{x}}^{n-1} - \hat{g})] = \\ &(1-\alpha)\hat{\boldsymbol{x}}^{n-1} + \alpha\hat{\boldsymbol{x}}^{n-1} - \alpha(\hat{\boldsymbol{d}}^n)^{\dagger}(\hat{\boldsymbol{d}}^n\hat{\boldsymbol{x}}^{n-1} - \hat{g}) = \\ &\hat{\boldsymbol{x}}^{n-1} - \alpha(\hat{\boldsymbol{d}}^n)^{\dagger}\hat{\boldsymbol{d}}^n\hat{\boldsymbol{x}}^{n-1} + \alpha(\hat{\boldsymbol{d}}^n)^{\dagger}\hat{g} = \\ &[\boldsymbol{I} - \alpha(\hat{\boldsymbol{d}}^n)^{\dagger}\hat{\boldsymbol{d}}^n]\hat{\boldsymbol{x}}^{n-1} + \alpha(\hat{\boldsymbol{d}}^n)^{\dagger}\hat{g}\end{aligned}$$

令 $\boldsymbol{H}^n=(\hat{\boldsymbol{d}}^n)^{\dagger}\hat{\boldsymbol{d}}^n$,$\boldsymbol{T}^n=(\hat{\boldsymbol{d}}^n)^{\dagger}\hat{g}$,则滤波器可表示为

$$\hat{\boldsymbol{x}}^1 = \boldsymbol{T}^1 \tag{11}$$

$$\hat{\boldsymbol{x}}^n = (\boldsymbol{I} - \alpha\boldsymbol{H}^n)\hat{\boldsymbol{x}}^{n-1} + \alpha\boldsymbol{T}^n \tag{12}$$

根据滤波器设计的规则可知,该算法需要求解 $N\times M$ 个广义逆。再对无正则化项模型进行分析,在目标跟踪过程中,问题中的 $\boldsymbol{G}$ 为响应函数,它是根据目标尺度确定的尖峰函数,对其做完傅里叶变换后的函数的模是一个 4 个顶点突起的函数,如图 2 所示。分析该函数,可以发现大部分函数值的模接近零,是一个十分稀疏的矩阵。由式(8)~式(10)可知,函数值的模接近零处的滤波器系数也接近零,因此可以省去这部分滤波器的计算。离散傅里叶变换的系数具有共轭对称性[10],因此滤波器系数具有共轭对称性,因此只需要求解约一半的剩余滤波器系数。实验分析,在保证精度的条件下,需要求解的滤波器系数可以降到 5% 以下,大大降低了计算量。

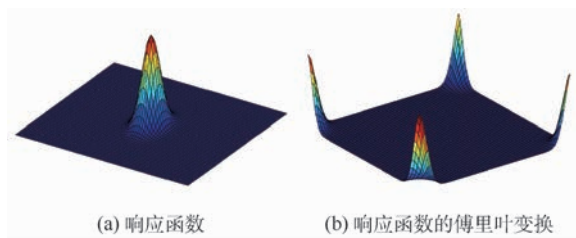


图 2　响应函数及其傅里叶变换

Fig. 2　Response function and its Fourier transform

北航学报 赠阅

本文跟踪算法如下：

1 初始化，输入初始参数

2 生成平移和尺度目标函数，并做傅里叶变换

3 输入初始帧图像 $\boldsymbol{F}_1$、初始目标位置 $\boldsymbol{P}_1$，提取目标特征

4 通过推论 1 快速求解广义逆，然后利用式(8)求解初始平移滤波器 $\boldsymbol{X}_1^{\text{tran}}$ 与初始尺度滤波器 $\boldsymbol{X}_1^{\text{scal}}$

5 $k=2$

6 重复

7 输入当前帧图像 $\boldsymbol{F}_k$，提取候选目标特征

8 由式(1)求解当前平移响应函数

9 根据平移响应函数峰值确定当前位置 $\boldsymbol{P}_k$

10 由式(1)求解当前尺度响应函数

11 根据尺度响应函数峰值确定当前尺度 S_k

12 在当前位置和尺度下，提取目标特征，利用推论 1 快速求解广义逆，通过式(9)求解当前平移滤波器 $\boldsymbol{X}^{\text{tran}}$ 和尺度滤波器 $\boldsymbol{X}^{\text{scal}}$

13 由式(10)更新尺度滤波器得到 $\boldsymbol{X}_k^{\text{scal}}$，更新平移滤波器得到 $\boldsymbol{X}_k^{\text{tran}}$

14 $k=k+1$

15 直到最后一帧

3 实验结果与分析

本文采用 DSST[5] 所提出的平移-尺度跟踪框架，采用多通道特征和多尺度分析，目标特征为灰度图像和 HoG 特征[11]，如图 3 所示。利用本文提出的滤波器设计方法构建滤波器。初始化参数主要有：搜索目标范围为目标大小的 4 倍，尺度样本数量为 33 个，相邻尺度样本之间的倍数为 1.015，平移滤波器学习率为 0.15，尺度滤波器学习率为 0.12 等。本文在 CPU：Inter(R) core(TM) i3-2310M 2.10 GHz RAM：2G 机器上进行实验。实验选取了目标跟踪基准(OTB)图像库中的 32 个图像序列对本文算法进行评估。这些序列包含了跟踪过程中会遇到的遮挡、尺度变化、快速运动、形变、面内旋转、面外旋转等情况，所选图像序列如图 4 所示。实验结果与 9 个当前比较先进的跟踪算法进行对比，这些跟踪算法分别是 KCF[4]、DSST[5]、SRDCF[12]、MEEM[13]、Struck[14]、ASLA[15]、SCM[16]、L1-APG[17]、CT[18]。这些算法是近年来目标跟踪领域常用来进行效果对比的算法。这些算法的跟踪结果来自 OTB[19] 数据库网站以及各自作者的网站。

目标跟踪效果评估标准是位置误差和重叠

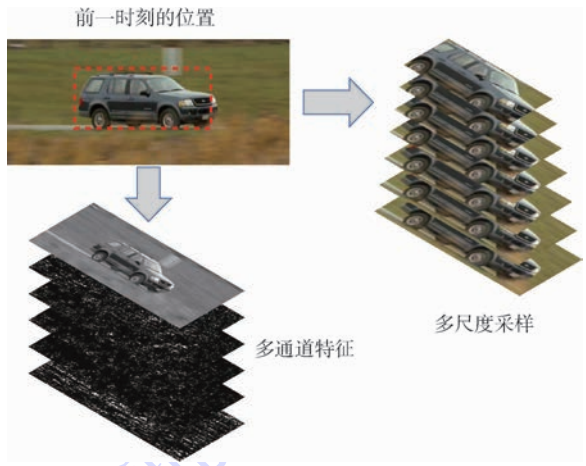


图 3 多通道特征和多尺度采样

Fig. 3 Multi-channel features and multi-scale sampling

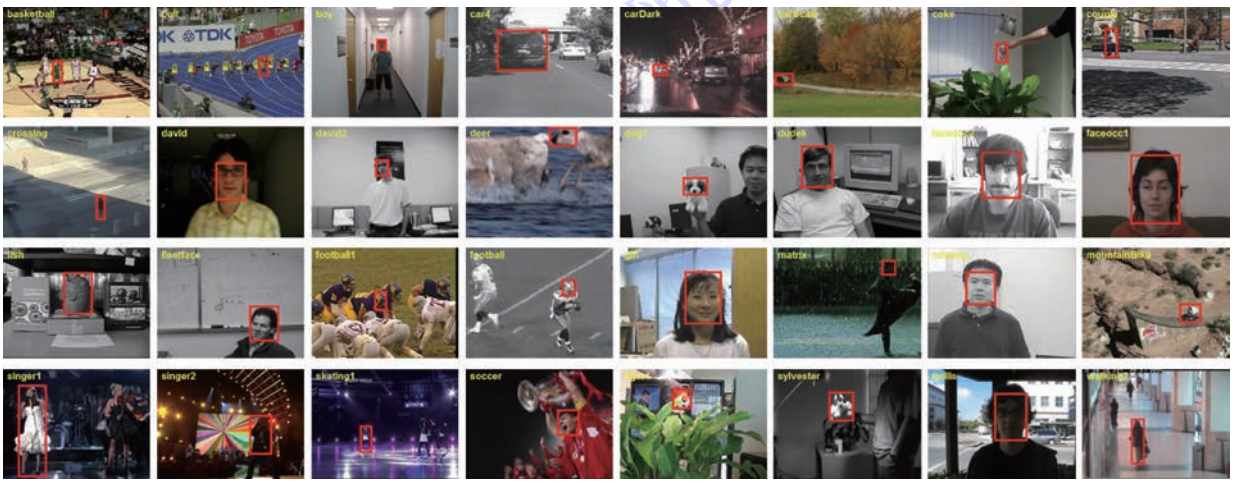


图 4 OTB[19] 数据库中的 32 个图像序列

Fig. 4 32 image sequences of OTB[19] database

北航学报 赠阅

率[19]。位置误差,即跟踪到的目标中心位置与目标实际的中心位置之间的距离。如图 5 所示,$\boldsymbol{O}_1$ 为目标真实位置的中心点,$\boldsymbol{O}_2$ 为跟踪到的目标的中心点,L 为真实位置的中心点与跟踪到的中心点的欧氏距离,L 就是位置误差,L 越小,则说明跟踪效果越好,反之越差。这种评估标准没有考虑到被跟踪目标的大小变化,当被跟踪目标大小发生变化时,这种评估标准不能很好地说明跟踪方法在估计目标大小上的性能。重叠率,即目标实际所在区域与跟踪确定目标所在区域相交的区域占相并的区域的比例,能更好地反映跟踪算法对目标的位置与大小估计。如图 5 所示,$\boldsymbol{Q}_1$ 表示实线边框围成的区域,$\boldsymbol{Q}_2$ 表示虚线边框围成的区域,$\boldsymbol{Q}_3$ 表示阴影区域,是 $\boldsymbol{Q}_1$ 与 $\boldsymbol{Q}_2$ 相交的区域,重叠率为

$$\rho = \frac{|\boldsymbol{Q}_1 \cap \boldsymbol{Q}_2|}{|\boldsymbol{Q}_1 \cup \boldsymbol{Q}_2|}$$

其中:$|\cdot|$表示区域的大小,即区域包含的像素数目。ρ 越大,说明跟踪效果越好。

根据重叠率大于阈值的帧数占总帧数的比例可以绘制出成功率曲线图,如图 6(a)所示。根据跟踪位置误差小于阈值的帧数占总帧数的比例可以绘制出精度曲线图,如图 6(b)所示。

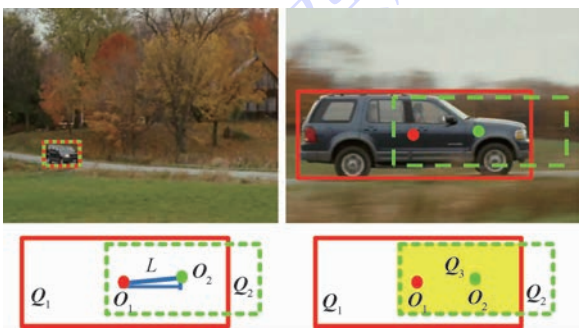


图 5　评估算法

Fig. 5　Evaluation algorithm

图 6(a)、(b)为本文算法与 KCF[4]、DSST[5]、SRDCF[12]、MEEM[13]、Struck[14]、ASLA[15]、SCM[16]、L1-APG[17]、CT[18] 10 个算法在这 32 个图像序列上跟踪效果的综合表现,曲线反映了在不同阈值下,各跟踪算法的跟踪效果。图 6(a)中,在重叠率阈值为 0.4 ~ 0.6 之间,本文算法的跟踪成功率较大程度上优于其他各算法。图 6(b)中,在位置误差阈值为 10 ~ 20 像素之间时,本文算法的精度较大程度上优于其他各算法。依据坐标与成功率曲线围成的下边区域的面积(AUC),可对各算法进行成功率排名,图 6(a)中方括号给出了这 10 个算法的 AUC 值。依据精度曲线位置误差阈值为 20 像素时的精度,可对各算法进行精度排名,图 6(b)中方括号给出了这10 个算法在阈值为 20 像素时的精度值。表 1 和表 2 给出了2 种评估算法的排名,在这 2 种评估算法下,本文算法排名都是最好的,表明本文算法总体跟踪效果优于其他算法。

图 7 给出了本文算法与 KCF[4]、DSST[5]、MEEM[13]、Struck[14] 5 个算法跟踪的跟踪效果图,其中本文算法用实线方框跟踪目标。Fleetface 图像序列主要是面内旋转场景,目标人物的脸部向面内旋转,从而使目标特征发生较大变化,在 690 帧当人物脸部转回时,本文算法仍然能够非常好的跟踪到他的位置和大小,其他 4 种算法跟踪到的目标在位置和大小估计上都比较差。Soccer图像序列主要是形变场景和遮挡场景,

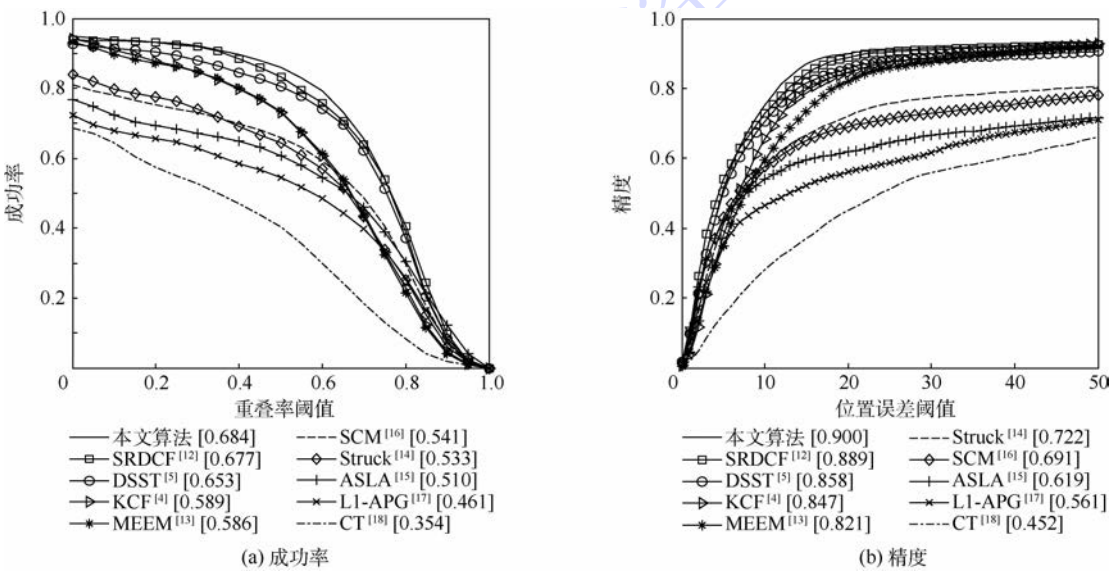


图 6　在 32 个图像序列上的成功率和精度曲线

Fig. 6　Success rate and precision curves over 32 image sequences

北航学报 赠阅

表 1 成功率排名

Table 1 Ranking in success rate

算法	本文算法	SRDCF[12]	DSST[5]	KCF[4]	MEEM[13]	SCM[16]	Struck[14]	ASLA[15]	L1-APG[17]	CT[18]
AUC	0.684	0.677	0.653	0.589	0.586	0.541	0.533	0.510	0.461	0.354
排名	1	2	3	4	5	6	7	8	9	10

表 2 精度排名

Table 2 Ranking in precision

算法	本文算法	SRDCF[12]	DSST[5]	KCF[4]	MEEM[13]	Struck[14]	SCM[16]	ASLA[15]	L1-APG[17]	CT[18]
精度	0.900	0.889	0.858	0.847	0.821	0.722	0.691	0.619	0.561	0.452
排名	1	2	3	4	5	6	7	8	9	10



图 7 本文算法与其他算法在 Fleetface、Soccer、Walking2 和 Skating1 序列上的跟踪结果

Fig. 7 Tracking results of proposed algorithm and other algorithms on Fleetface, Soccer, Walking2 and Skating1 sequences

目标球员的脸部发生较强的形变以及被前面的纸片遮挡,本文算法对目标实现持续跟踪,其他 4 种算法都出现丢失目标的情况。Walking2 图像序列主要是尺度变化和遮挡场景,行走中的女士慢慢走向镜头的远处,从而慢慢变小,中间过程中出现男士,在一段时间内挡住了她,本文算法和 DSST 跟踪器都对其进行了准确的跟踪。Skating1 主要是遮挡和光照变化,女生在滑行过程中被其他人部分遮挡,整个滑行场所光照发生很大变化,本文算法对其实现了稳定的跟踪。

图 8 为本文算法与 KCF[4]、DSST[5]、SRDCF[12]、MEEM[13]、Struck[14]、ASLA[15]、SCM[16]、L1-APG[17]、CT[18] 10 个算法在快速运动、形变、尺度变化、遮挡、面外旋转、面内旋转 6 种场景下的跟踪性能对比。图 8 在各个子图中的方括号中给出了这 10 种跟踪器在不同场景下的 AUC 值。在形变、遮挡、面外旋转、面内旋转这 4 种场景下,本文算法的 AUC 值分别为 0.656、0.657、0.651、0.674,均高于其他算法的 AUC 值,表明本文算法在这 4 种场景下的跟踪效果最好。在快速运动场景下,基于多样本学习的 SRDCF 算法的 AUC 值为0.646,本文算法为0.575,这是因为SRDCF

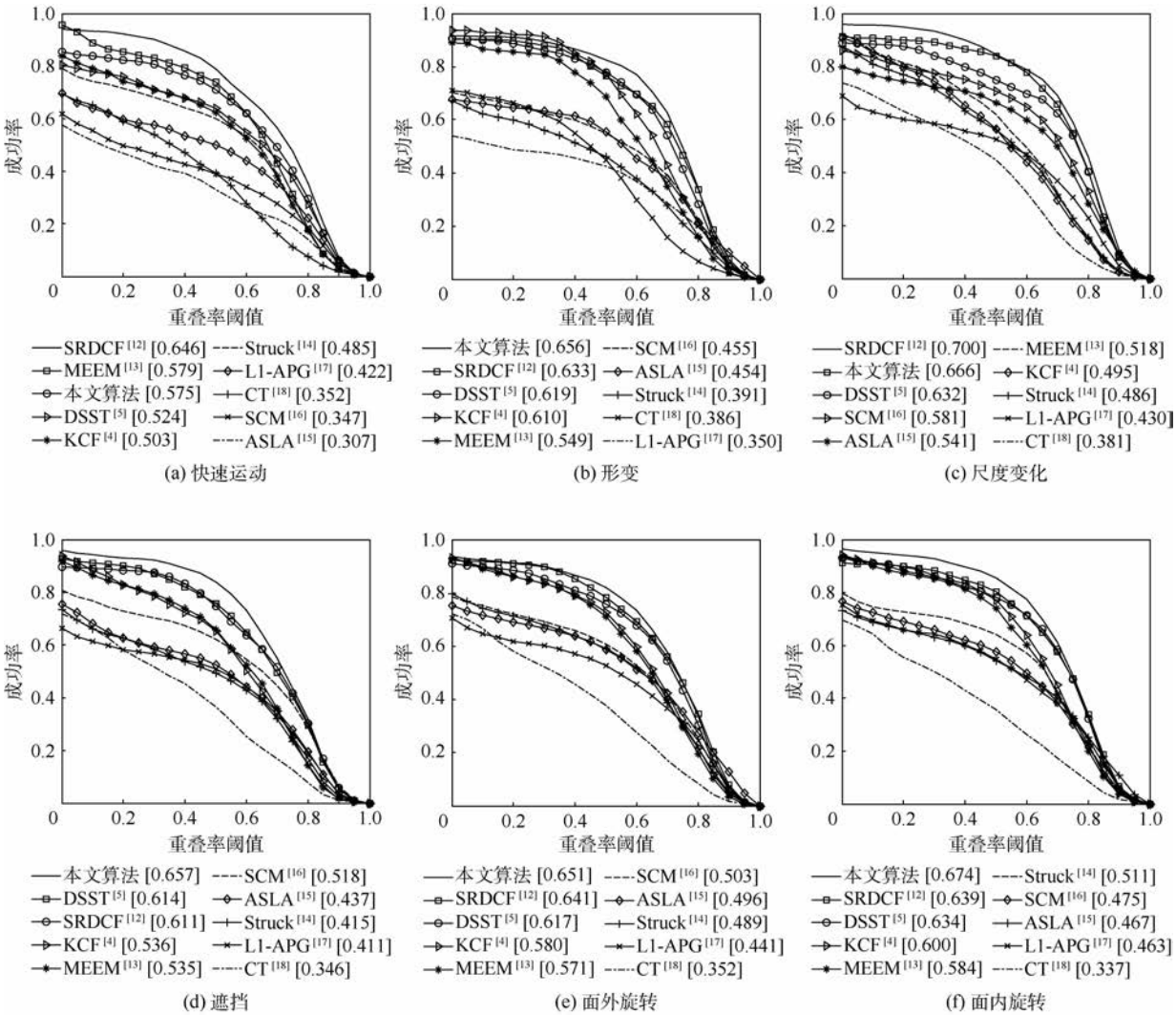


图 8　快速运动、形变、尺度变化、遮挡、面外旋转、面内旋转场景下的成功率曲线

Fig. 8　Success rate curves of fast motion, deformation, scale variation, occlusion, out-of-plane rotation and in-plane rotation

算法进行了更广区域的目标搜索,能够更好地捕捉到快速运动的目标。在尺度变化场景下,本文算法 AUC 值为 0.666,仅次于 SRDCF 算法的 0.700。在各种场景下,本文算法的 AUC 值都高于基于正则化约束的卷积表示 DSST 算法,表明本文所提算法提高了跟踪效果。

本文跟踪算法在实现更好跟踪效果的同时,跟踪速度也与其他跟踪算法相近,在 32 个图像序列上的平均跟踪速度达到 9.5 帧/s,对于 29 像素 ×23 像素大小的目标,可以实现 26 帧/s 的跟踪速度,达到实时跟踪效果。跟踪算法 DSST[5]、SRDCF[12]、MEEM[13]在各自文献中给出的速度分别为 24、5 和 10 帧/s。为了进一步对比速度之间的差异,将跟踪速度最快的 DSST[5]跟踪算法在本文算法的实验机器上运行,得到的跟踪速度为 12.3 帧/s,这与本文算法的跟踪速度相近,说明了本文算法也具有较快的跟踪速度。

4　结　论

1)本文提出一种在傅里叶域求解卷积表示的目标跟踪算法。将卷积表示问题变换到傅里叶域,可以通过广义逆的满秩算法快速求解,得到多样的滤波器设计方法。

2)信号在傅里叶域的共轭对称性和稀疏性可以辅助降低算法的计算复杂度。

3)在 OTB 图像库的 32 个序列上进行跟踪实验,与当前比较先进的 9 种目标跟踪算法进行对比。实验结果显示,本文跟踪算法在跟踪性能上优于其他 9 种跟踪算法。

后续工作会将本文提出的滤波器设计方法用于多样本训练模型,进一步提高目标跟踪的速度和精度。

参考文献(References)

[1] YILMAZ A, JAVED O, SHAH M. Object tracking: A survey

北航学报 赠阅

[J]. ACM Computing Surveys, 2006, 38(4): 1-45.

[2] SMEULDERS A W M, CHU D M, CUCCHIARA R, et al. Visual tracking: An experimental survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(7): 1442-1468.

[3] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]// 2010 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2010: 2544-2550.

[4] HENRIQUES J, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.

[5] DANELLJAN M, HAGER G, KHAN F S, et al. Accurate scale estimation for robust visual tracking[C]// Proceedings of the British Machine Vision Conference 2014. Durham: BMVA Press, 2014: 1-11.

[6] KIANI H, SIM T, LUCEY S. Multi-channel correlation filters[C]// 2013 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2013: 3072-3079.

[7] BODDETI N V, KANADE T, KUMAR B V K V. Correlation filters for object alignment[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2013: 2291-2298.

[8] 张跃辉. 矩阵理论与应用[M]. 北京: 科学出版社, 2011: 207-231.
ZHANG Y H. Matrix theory and application[M]. Beijing: Science Press, 2011: 207-231(in Chinese).

[9] PUNTANEN S, STYAN P H G, ISOTALO J. Matrix tricks for linear statistical models[M]. Berlin: Springer-Verlag, 2011: 349-350.

[10] OPPENHEIM V A, WILLSKY S A. Signals and systems[M]. Upper Saddle River: Prentice Hall, 1983: 322.

[11] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// 2005 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2005: 886-893.

[12] DANELLJAN M, HAGER G, KHAN S F, et al. Learning spatially regularized correlation filters for visual tracking[C]// 2015 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2015: 4310-4318.

[13] ZHANG J M, MA S G, SCLAROFF S. MEEM: Robust tracking via multiple experts using entropy minimization[C]// 2014 European Conference on Computer Vision. Berlin: Springer-Verlag, 2014: 188-203.

[14] HARE S, GOLODETZ S, SAFFARI A, et al. Struck: Structured output tracking with kernels[C]// 2011 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2011: 263-270.

[15] JIA X, LU H C, YANG M H. Visual tracking via adaptive structural local sparse appearance model[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2012: 1822-1829.

[16] ZHONG W, LU H C, YANG M H. Robust object tracking via sparsity-based collaborative model[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2012: 1838-1845.

[17] BAO C L, WU Y, LING H B, et al. Real time robust L1 tracker using accelerated proximal gradient approach[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2012: 1830-1837.

[18] ZHANG K H, ZHANG L, YANG M H. Real-time compressive tracking[C]// 2012 European Conference on Computer Vision. Berlin: Springer-Verlag, 2012: 866-879.

[19] WU Y, LIM J, YANG M H. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.

作者简介：

朱日东 男，博士研究生。主要研究方向：目标跟踪。

杨小远 女，博士，教授，博士生导师。主要研究方向：应用调和分析与图像处理。

王敬凯 男，博士研究生。主要研究方向：多源遥感图像融合。

北京航空航天大学学报

北航学报 赠阅

Convolution representation-based object tracking algorithm in Fourier domain

ZHU Ridong, YANG Xiaoyuan*, WANG Jingkai

(School of Mathematics and Systems Science, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: A novel object tracking algorithm based on convolution representation in Fourier domain is proposed for object tracking. Object tracking question can be treated as a convolution representation model. By finding the best filters, which reconstruct the target function with minimum loss, fast and robust object tracking can be realized. When the optimal multi-channel convolution representation model is mapped to the Fourier domain, it is equal to solving the least squares solution to linear equations. First, all solutions of the system of linear equations can be expressed through the theory of pseudo inverse, which provide a general format of convolution filters. Then, filters updated in the previous frame and feature templates extracted from current frame are used to generate current filters, and the pseudo inverse can be obtained fast through the full rank algorithm. Finally, tracking filters are updated and applied in both translation and scale. Experimental results on the object tracking benchmark (OTB) database show that our algorithm performs better than some state-of-the-art tracking methods in terms of accuracy and offers a general format to design filters.

Keywords: object tracking; convolution representation; Moore-Penrose pseudo inverse; Fourier transform; optimal approximation

北京航空航天大学学报

Received: 2017-01-18; **Accepted**: 2017-04-07; **Published online**: 2017-05-16 14:30

URL: kns.cnki.net/kcms/detail/11.2625.V.20170516.1430.001.html

Foundation items: National Natural Science Foundation of China (61671002); Beijing Natural Science Foundation (4152029)

* **Corresponding author**. E-mail: xiaoyuanyang@vip.163.com

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2016.0959

基于 CT 的泡沫铝三维细观模型重建及应用

李侯贞强[1,2], 张亚栋[1,*], 张锦华[1], 姜春琳[1]
(1. 中国人民解放军陆军工程大学 爆炸冲击防灾减灾国家重点实验室, 南京 210007; 2. 太原卫星发射中心, 太原 036304)

摘 要: 为了建立更加真实的闭孔泡沫铝三维细观分析模型,提出了一种基于计算机层析成像(CT)的有限元模型构建新方法。首先,对 CT 扫描得到的泡沫铝试件的扫描图像进行 Otsu 算法分析,确定了区分基体材料和空气的灰度最佳阈值。其次,基于映射网格思想直接从扫描图像生成了试件的有限元分析模型,实现了泡沫金属三维细观分析模型的重建。最后,以此为基础进行了泡沫铝试件准静态压缩和动态冲击试验的数值模拟,结果表明,准静态压缩下泡沫铝的内部变形随机分布于整个试件,且与其三维结构密切相关;而动态冲击下变形在冲击端附近首先发生,体现出显著的局部化效应。本文方法能真实地描述泡沫金属内部的细观结构,实现了对泡沫铝试件在准静态压缩和动态冲击作用下的受力、变形与破坏过程更加详细的模拟分析。

关 键 词: 闭孔泡沫铝; CT 图像; 细观模型; 重建; 数值模拟
中图分类号: TU311; V256
文献标识码: A **文章编号**: 1001-5965(2018)01-0160-09

全下文载

闭孔泡沫金属的胞孔大小和形态因制备工艺不同而变化,内部细观结构具有无序、非均匀、不等壁厚等特点,使其具备许多独特的性能而在工程中得到广泛应用[1-3]。研究表明,泡沫金属的物理力学特性与其细观结构息息相关,都是源于材料内部结构的随机性[4-5]。在细观尺度上开展泡沫金属材料的力学行为和损伤破坏机制的研究,建立其宏观力学性能与细观结构特征之间的联系,是此类材料研究的热点之一。

泡沫金属材料内部结构受力后的变形演化复杂,常规试验研究手段存在诸多的困难,因此基于精细化分析模型的数值模拟成为有效的研究手段。Gibson 和 Ashby 基于对泡沫金属实际结构的高度抽象,提出了用立方体单胞元模型来近似泡沫铝细观结构的思想[1]。以此为基础,研究人员对单元形状不断改进,先后发展了改进的立方体模型[6-7]、十四面体(Kelvin 模型)[8-9]及八面体模型[10-11]等。这些单胞元模型能为材料的力学性能分析提供一种简便的方法,但其简单的单元形状与规则排列的特点过于理想化,和实际结构相差甚远。考虑胞孔的位置和大小在一定范围内随机分布的特点,Voronoi 拓扑算法被很多学者用来构建泡沫金属的细观随机模型。Zhang[12]、Zheng[13]和 Li[14]等先后应用该方法建立细观模型,并开展了泡沫金属材料的动静态力学性能和损伤破坏模式的研究。但该模型基于等壁厚单元的假设与实际情况不符。为此,三维随机椭球体模型[15]和三维随机多面体模型[16-17]先后被提出来,并克服了单胞元模型规则排列的缺点和胞孔等壁厚的问题,但这些算法仍存在所建模型受随机因子影响大以及算法稳定性不高的问题,实际应用尚不十分令人满意。

收稿日期: 2016-12-21; **录用日期**: 2017-03-17; **网络出版时间**: 2017-04-13 17:48
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170413.1748.007.html
基金项目: 国家重点研发计划(2014YFC0305200); 国家自然科学基金(51478464, 51678566); 国家重大科学仪器设备开发专项(2014YQ24044509)
* **通信作者**. E-mail: zhydjs@139.com

引用格式: *李侯贞强, 张亚栋, 张锦华, 等. 基于 CT 的泡沫铝三维细观模型重建及应用[J]. 北京航空航天大学学报, 2018, 44(1): 160-168. LI H Z Q, ZHANG Y D, ZHANG J H, et al. Reconstruction and application of three-dimensional mesoscopic model of aluminum foam based on CT[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(1): 160-168 (in Chinese).*

北航学报 赠阅

近年来，一些学者尝试利用计算机层析成像（Computed Tomography，CT）扫描技术对泡沫金属材料进行研究[18-20]，能够将材料的细观结构形态逐层以灰度图像直观、无损地展示出来，并可进一步利用图像重建技术得到材料的三维细观结构模型[21]，最大程度地还原了泡沫金属材料的胞孔形态和拓扑结构特征，具有其他方法无法比拟的真实性。目前，基于 CT 的建模方法复杂，中间处理步骤繁多，往往需要借助于大型商用软件来实施，限制了其广泛的应用性。

本文基于 CT 扫描图像，利用 Fortran 语言编写了材料属性判别程序，基于映射网格思想建立了闭孔泡沫铝的细观有限元分析模型。建模过程无需生成点、线、面、体等几何元素，而是直接生成节点和单元，避免了图像重建过程中网格剖分质量控制的难题。所建立的闭孔泡沫铝三维细观分析模型，更加接近实际材料的细观结构特征，分析结果与试验结果吻合良好，为泡沫铝材料的细观结构分析提供一种新的途径。

1　CT 试验

CT 技术在不破坏被扫描物体结构的前提下，根据材料内部不同密度的成分对 X 射线衰减值不同的原理，由多角度收集 X 射线穿过物体衰减后的投影数据，再通过数学方法处理得到物体中不同点的 X 射线衰减值信息，并逐层转化为灰度图像以进行观察。该技术经医学领域拓展到工业、农业、工程、安全检测等方面，能够真实还原出物体内部复杂的细观结构形态和拓扑结构特征。

1.1　试件制作

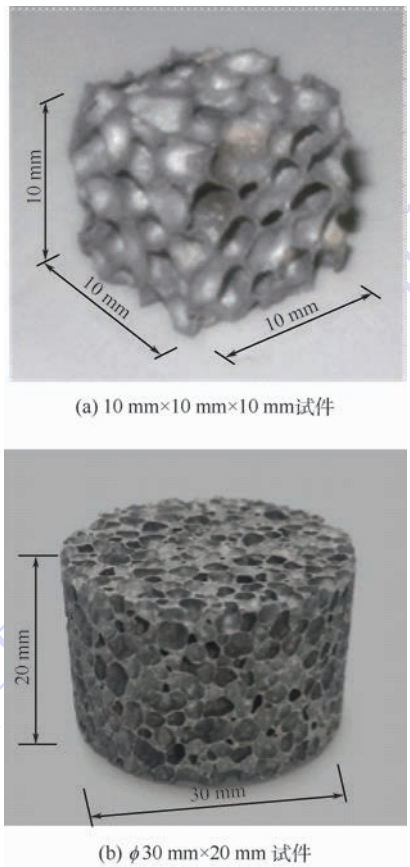
本文闭孔泡沫铝试件利用粉末冶金技术制备，基体材料为 Al-7% Si（7% 为质量分数）合金，孔径主要分布在 1 ~ 3 mm，通过称重法得到孔隙率为 85.52%。

依据 Jeon 和 Asahina[22] 实验研究，只要泡沫铝的胞孔不存在大的缺陷，如局部的胞孔相连、缺失、坍塌等，小尺寸泡沫铝试件的力学特性就能有效地表征整个材料的力学特性。同时对于尺寸大小固定的 CT 探测器，试件越小扫描的放大倍数越大，得到的扫描结果精度也更高。鉴于此，本文立方体试件和圆柱体试件分别按 10 mm × 10 mm × 10 mm 和 ϕ30 mm × 20 mm 的尺寸加工，如图 1 所示。

1.2　CT 扫描

CT 扫描采用丹东奥龙射线有限责任公司研制的 ICT-3400 型工业 CT 检测系统，该 CT 系统的空间分辨率为 25 LP/cm，密度分辨率为 0.25%，扫描模式为锥束连续扫描，为兼顾扫描效果和后期数据的处理，选用的扫描精度为 33 μm，通过 360°旋转扫描共得到 720 个不同角度的扫描数据。对这些数据进行滤波、校正等处理，10 mm × 10 mm × 10 mm 试件经由 FDK 算法[23] 计算得到 292 个截面的切片数据。本文基于这些数据进行三维细观分析模型的重建。

上述切片数据可以用灰度图像进行显示，如图 2 所示。由于铝和空气对 X 射线的衰减系数相差较大，在灰度图像中能明显的区分出铝基体材料和空气，清晰地展示了泡沫铝内部的细观结构。从图中可以看出，泡沫铝内部结构非常复杂，不同截面切片结构的形态大不相同，胞孔的大



(a) 10 mm×10 mm×10 mm试件

(b) ϕ30 mm×20 mm 试件

图 1　闭孔泡沫铝试件

Fig. 1　Closed-cell aluminum foam specimens

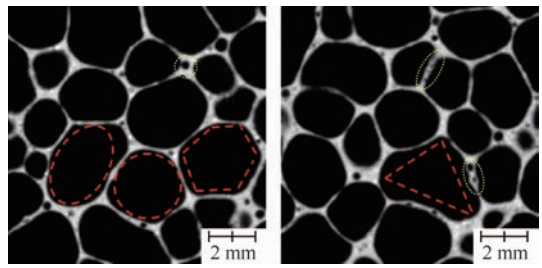


图 2　CT 扫描得到的泡沫铝灰度图像

Fig. 2　Gray images of aluminum foam by CT scanning

北航学报 赠阅

小分布毫无规律,胞孔形状复杂多变,孔壁厚度变化互不相同,许多胞壁内存在数量不等的细微孔洞,造成胞壁的局部肿大或缺陷。可见,高精度的CT扫描能够很好地还原材料内部的微缺陷等细观结构特征,可基于此建立复杂真实的细观结构模型来精细地研究泡沫铝的力学性能及破坏形态。

2 分析模型的重建

CT扫描得到的灰度图像能清晰地识别出材料的内部结构特征。本文以10 mm × 10 mm × 10 mm试件为例,基于扫描得到的图像,进行三维细观模型的重建,为有限元计算打下基础,其步骤见2.1节~2.4节。

2.1 映射网格的生成

试件扫描后的切片图像大小为292像素×292像素,共292层。将这些图像逐层叠放,得到292×292×292的点云矩阵,称为衰减值矩阵,矩阵中每一个数值代表该区域的X射线衰减值。对切片数据的处理,通常是先进行二值化转化,即针对衰减值矩阵取定一个阈值,将高于该阈值的区域重新赋予数值1并判定为胞壁,低于该阈值的将重新赋予数值0并判定为胞孔,从而将灰度图像转化为具有明显对比度的黑白图像。要对泡沫铝进行细观模型重建,通常生成包含点、线、面、体等大量的中间过渡数据,并对重建的几何模型划分网格以进行有限元分析[20-21]。然而,由于泡沫铝内部结构的复杂性,即使用二值化的图像进行几何模型重建也是很困难的,并且很难划出高质量的有限元分析网格。

为避免上述重建过程的困难,本文直接利用三维衰减值矩阵生成有限元模型。首先,对扫描图像进行二值化处理,判定图像中的各点属性为胞壁或胞孔,建立新的点云矩阵;然后,取矩阵中第1个点,以其为中心生成一个正方体单元和相应的8个节点,如果该点在二值化过程中判定为胞壁,则确定此单元的属性为胞壁,否则确定为胞孔(空气)。以此方式逐行、逐列、逐层推进,就在整个点云矩阵中生成了规则有序、均匀划分的有限元模型。显然,这是一种映射化的网格模型,其中包含了胞壁(基体铝)和胞孔(空气)2种成分。

图3给出了典型切面的二值化图像与有限元模型的对比情况,可以看到有限元模型很好地还原了CT扫描图像中的结构细节特征。

2.2 材料阈值的判定

泡沫铝是铝基体胞壁和空气胞孔组成的两相材料,尽管二者的X射线衰减值有明显差别,但

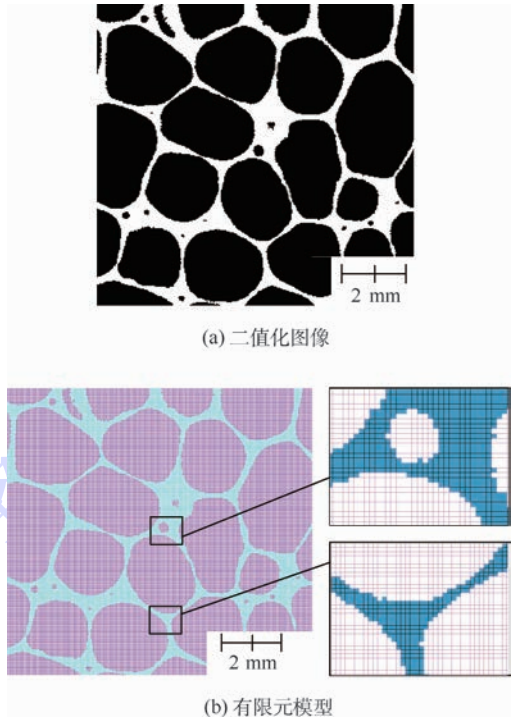


图3 二值化图像与有限元模型的对比

Fig. 3 Comparison between binary image and finite element model

要从原始灰度图像中区分出空气与胞孔的界面,就存在一个灰度阈值的选取问题(如图2所示)。如果阈值取得过高,则会有一部分铝基体被误判为空气;如果阈值取得过低,则一部分过多的空气误判为铝基体。为便于计算机处理,利用CT数据进行三维细观模型重建时,先将切片图像进行了二值化处理,为此必须选取合适的灰度阈值。

泡沫铝是一种组分鲜明的材料,扫描后的图像对比度好,对于这种前景与背景具有明显差别的,图像判别中的阈值可采用Otsu算法[24]进行估计。该算法假定某一个阈值将数据分成两组,当2组数据类间方差最大时则认为二值化效果最佳。也就是对于所选定的阈值,在判定图像上某点是属于目标还是背景时出错的概率最小。Otsu是数字图像处理中的常用算法,在MATLAB软件中通过Multithresh函数直接调用。

本文CT扫描得到的X射线衰减值最大为0.016,最小为0,基体材料衰减值主要处于10^{-3}量级,空气衰减值主要处于10^{-4}量级,通过MATLAB计算得到区分基体材料和空气的阈值为0.0022。此时,对应的孔隙率为80.16%,比称重法测得的孔隙率小5.36%。误差产生的主要原因是,泡沫铝在制备过程中除生成胞孔外,还在基体材料中产生大量微细孔洞[25]。尺寸小于扫描精度的微细孔洞在图像处理中均判定为基体,从而降低了计算的孔隙率。Mcdonald等[25]研究表

明，当泡沫铝压缩到达屈服应力及孔壁出现坍塌时，相对大的胞孔优先发生破坏，而微细孔洞对材料的力学性能影响非常小。这表明，上述误差难以消除但可接受。

2.3　单元尺寸选择

上述建模过程中，单元的边长即搜索步长需人为给定。若按照 CT 系统原始扫描精度构造空间立方体单元，单元尺寸为 0.033 mm，整个试件包含的单元数目将超过 2 000 万个，这对于目前的计算机求解能力来说是很困难的。对扫描图像的分析表明，本文泡沫铝试件的壁厚主要分布在 0.1 ~ 0.3 mm 之间。综合考虑计算效率和精度的要求，重建中对原始衰减值矩阵进行间隔取点，在行列层方向上每隔 3 个点取一个点，来表征闭孔征泡沫铝的主要结构形态，即使得单元的最小尺寸变为 0.132 mm。这样，在保留胞孔主要结构形态的基础上大大减少了单元的数量。

根据以上方法，编制了 Fortran 程序，生成的分析模型用 ANSYS 软件的 APDL 语言描述，其中包括节点数据、单元数据和材料属性数据，可以直接读入 ANSYS 进行分析。典型试件的三维细观有限元模型如图 4 所示，单元总数量为 389 017，其中基体材料单元为 77 185，空气单元为 311 832。

2.4　方法的验证

在所建立三维细观模型中等距离截取 9 个切片 S1 ~ S9，分别统计各切片上基体材料单元和空气单元个数，计算每个切片的孔隙率 ρ，并与相应灰度图像计算的孔隙率进行对比，同时计算得到两者之间的误差，结果如图5所示。从图中

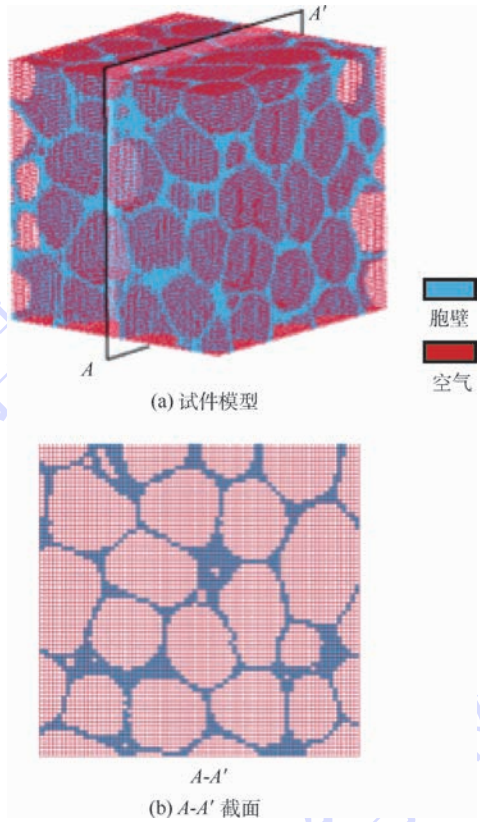


图 4　闭孔泡沫铝三维细观有限元模型

Fig. 4　Three-dimensional mesoscopic finite element model of closed-cell aluminum foam

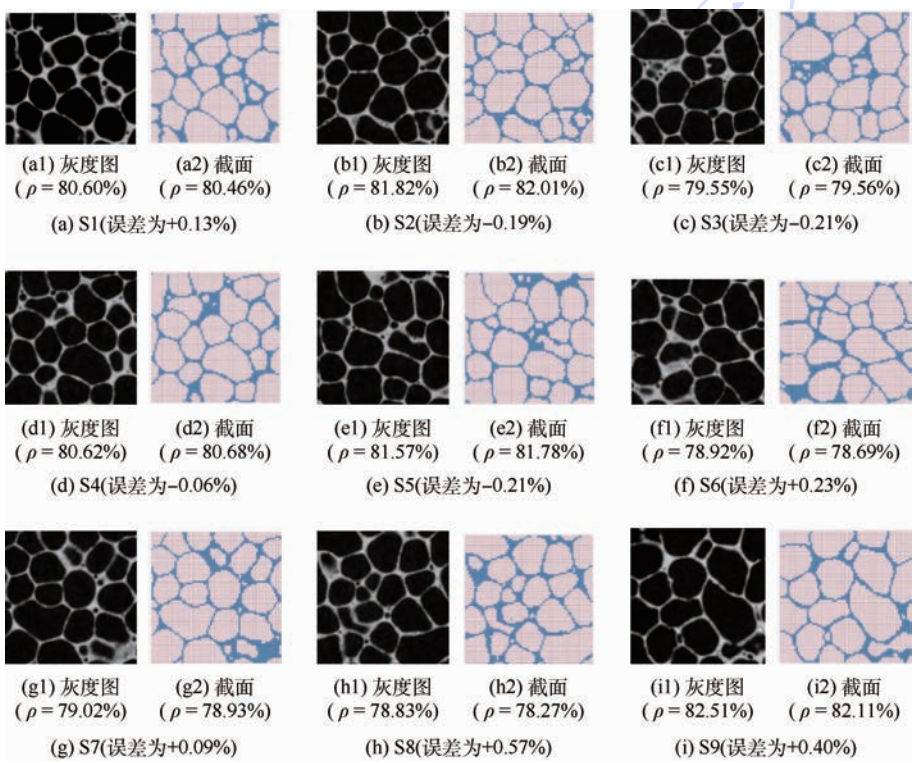


图 5　有限元模型不同截面和灰度图像对比

Fig. 5　Comparison of different section and gray images of finite element model

可以看出,模型中不同切片的孔隙率与灰度图像吻合良好,最大相差 0.57%,说明所建立的模型能准确地表达泡沫铝的细观结构特征,本文方法的重建结果可靠。

3 泡沫铝静、动态特性

3.1 静态力学特性

采用长春试验机研究所生产的 CSS-44200 电子万能试验机对 10 mm × 10 mm × 10 mm 试件进行准静态压缩试验。加载用位移控制,速度为 0.6 mm/min,名义应变率为 0.001 s^{-1}。

同时,利用 ANSYS/LS-DYNA 对重建后的模型进行有限元分析。其中基体材料的模型取为 * MAT_PLASTIC_KINEMATIC[26],材料密度为 2 680 kg/m^3,弹性模量为 70 GPa,泊松比为 0.31,屈服强度取 70 MPa,切线模量取 80 MPa。采用侵蚀算法模拟材料断裂,破坏准则取为最大主应变准则,破坏应变为 0.4。数值分析模型如图 6 所示,刚性加载板与试件之间设置接触,摩擦系数取 0.1。

图 7 给出了试验与数值模拟得到的名义应力-应变曲线。可以看出,两者总体吻合良好,都分为 3 个阶段:在应变小于 0.03 的初始加载阶段,应力水平较小,试件主要处于弹性加载状态,名义应力-应变成线性关系;随着进一步加载,泡沫铝的局部孔壁达到屈服强度,产生塑性变形,名义应变从 0.03 增加 0.51 时,名义应力的增加幅度变化很小,呈现为稳定的应力平台;当名义应变到达 0.51 之后,基体材料基本被压密实,此时随应变增加名义应力急剧上升。

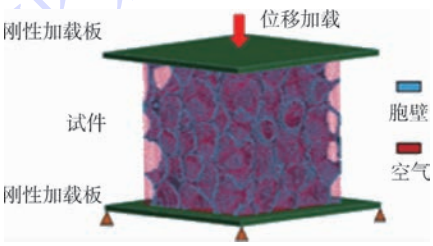


图 6 泡沫铝准静态压缩数值分析模型

Fig. 6 Numerical analysis model of aluminum foam under quasi-static compression

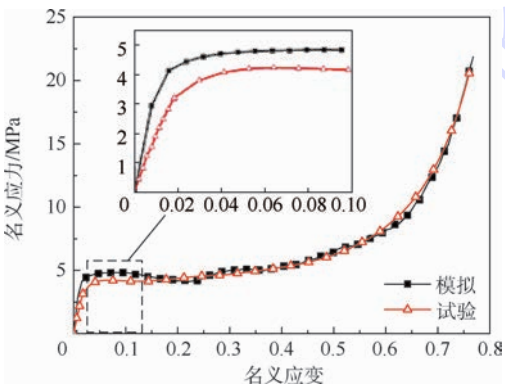


图 7 准静态压缩试验与模拟的名义应力-应变曲线

Fig. 7 Nominal stress-strain curves of test and simulation under quasi-static compression

将图 7 中名义应变 0 ~ 0.1 之间的部分放大观察发现,模拟所得曲线在名义应变小于 0.03 时的弹性模量大于试验值,这主要是由于实际材料中基体的力学曲线表现出非线性[27],而本文为了简化计算采用线性模型。尽管如此,二者对应的材料吸能效果只相差 1% 左右。图中还发现,在名义应变 0.03 ~ 0.2 之间时数值模拟得到的名义应力值略高于试验值。这可能是由于建模过程中忽略了胞壁中的微细孔洞,导致孔壁的强度增大所致。虽然如此,这一区间内的名义应力相差在 5% 以内,说明本文模型能够描述泡沫铝在静态荷载压缩下的基本力学性能特征。

为进一步观察泡沫铝压缩过程中内部变形破坏情况,选取 3 个纵向截面进行对比分析,如图 8 所示。其中 x 为截面距端面的距离,ε 为试件的名义应变。可以看到,试件在压缩的各个阶段,各部分同时产生变形,但各纵向截面都存在局部的不均匀变形,且不均匀变形出现的位置各不相同,破坏形态也有所区别,出现了随机的压溃带或剪切破坏带,这与文献[28]较为吻合。同时可明显看出,各纵向截面中的较大塑性变形并非首先出现在截面中最薄位置,而是存在一定的随机性,这表明泡沫铝胞壁发生变形直至坍塌破坏与整个试件的三维结构密切相关。

3.2 动态冲击特性

采用直径 37.5 mm 的霍普金森压杆(SHPB)设备进行动态冲击试验。子弹长度为 600 mm,入射杆和透射杆长度均为 2 000 mm,材料为 LC4 铝合金。圆柱体试件直径 30 mm,高 20 mm,子弹的冲击速度为 10.61 m/s。入射杆采用电阻应变片,透射杆采用半导体应变片进行测量,入射杆和透射杆中的应力-时间曲线如图 9 所示。试验前对泡沫铝试件进行 CT 扫描,利用第 2 节方法建立了试件的三维细观有限元模型,模型中同时包含了与试验尺寸相同的入射杆和透射杆,材料参数、破坏准则以及接触定义均与准静态加载条件下相同。利用所建立的模型进行模拟分析,结果示于图 9 中。可以看到,试验与数值模拟得到的入射波、反射波和透射波总体吻合良好,说明本文模型能有效地模拟泡沫铝试件的动态冲击过程。

图10给出了模拟得到的动态冲击加载作用

北航学报 赠阅

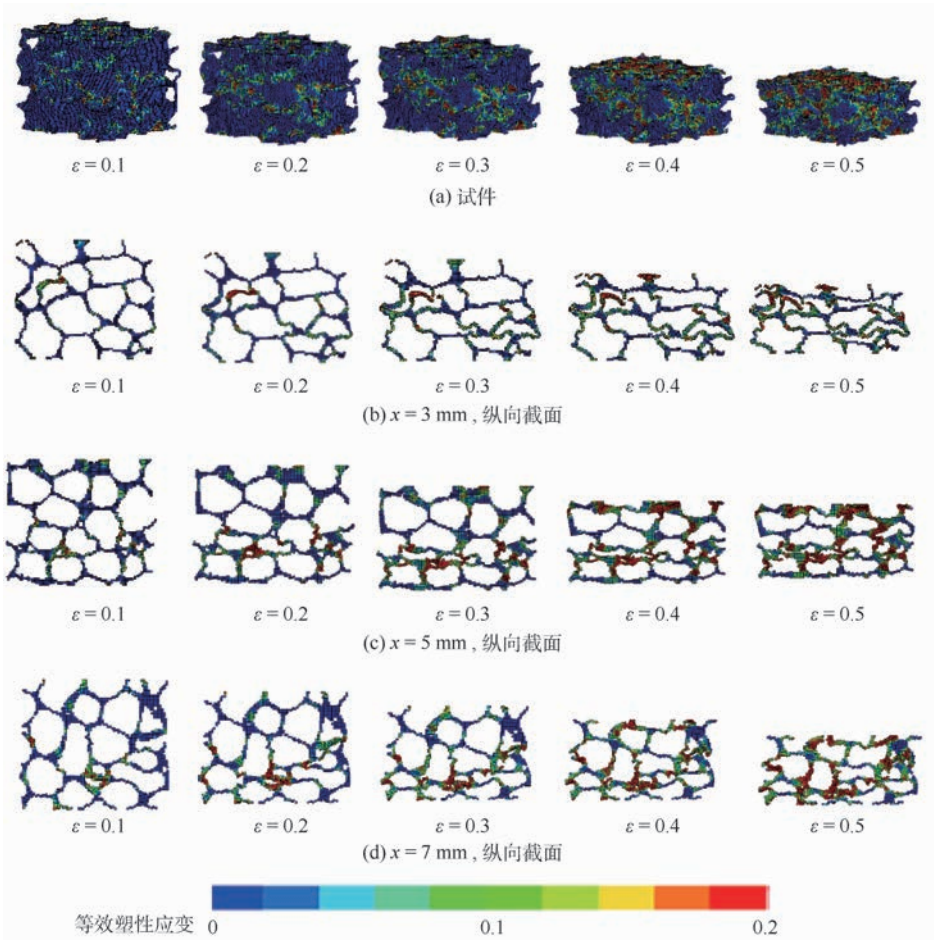


图 8　准静态加载下试件及胞壁的变形形态

Fig. 8　Deformation morphology of specimen and cell-wall under quasi-static loading

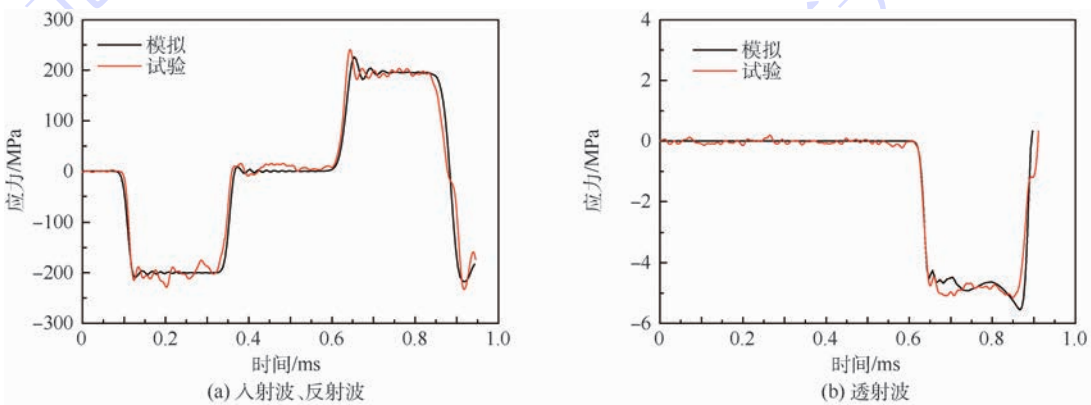


图 9　动态冲击试验与模拟的应力-时间曲线

Fig. 9　Comparison between test and simulation results under impact loading

下泡沫铝试件的截面应力状态和变形过程。可以看出，冲击荷载作用下泡沫铝的内部变形与准静态加载下有显著区别，变形首先发生在冲击端附近，并随着冲击过程逐层推进，表现出明显的局部效应。这与文献[28]中等加载速度下的变形与破坏形态相一致。图中胞壁的破坏变形呈多样化，有屈曲变形、剪切破坏、扭转变形等，表明了该种材料变形破坏过程的复杂性。

图 11 为试件两端的平均应力-时程关系曲线。结合图 10、图 11 发现，应力波在试件内的传播速度很快，大约 0.015 ms 就传到整个试件，表现为图 11 中试件两端平均应力-时程曲线中初始阶段的快速上升。图 11 表明，试件冲击端平均应力增长更快，首先达到材料的屈服强度，胞孔孔壁开始发生塑性变形。随后，试件两端的平均应力都呈现出平台阶段，孔壁开始大范围地发生变形，

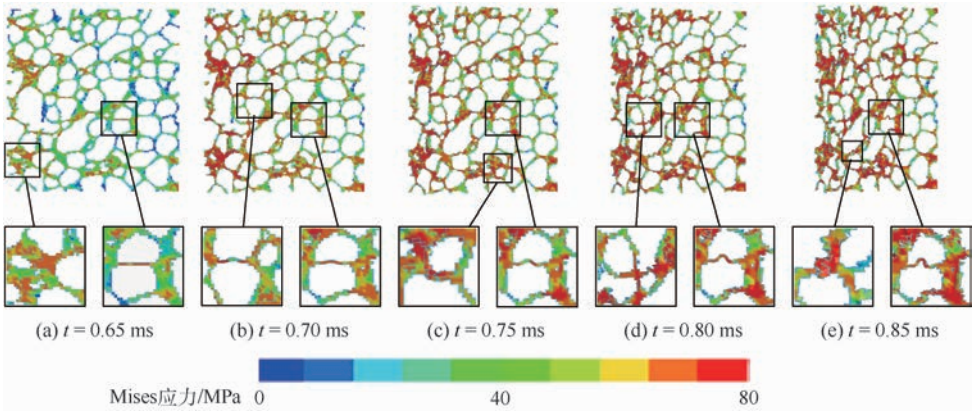


图 10 动态冲击加载下试件内部的 Mises 应力分布

Fig. 10 Mises stress distribution inside specimen under dynamic impact loading

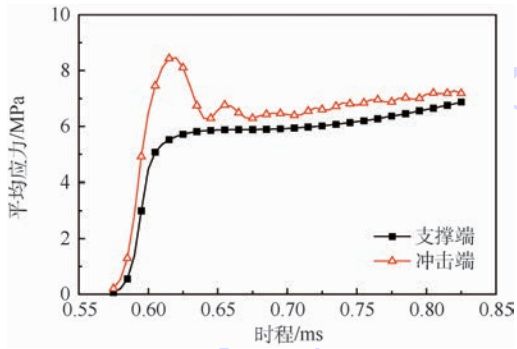


图 11 试件冲击端、支撑端的平均应力-时程曲线

Fig. 11 Average stress-time history curves at loading and supporting ends of specimen

直至最终压溃。另外,整个冲击过程中试件冲击端的平均应力始终高于支撑端,导致这一现象的可能原因:一是冲击过程中存在宏观惯性效应;二是冲击端局部压密的材料具有更高的屈服应力。

4 结 论

本文提出了一种基于 CT 的泡沫铝材料三维有限元分析模型重建新方法,基于此模型分析了泡沫铝试件在准静态压缩和动态冲击作用下的受力与变形特性,并与试验进行了对比。主要结论如下:

1)本文方法很大程度上简化了 CT 扫描图像转化为有限元模型的过程,重建过程中直接生成节点和单元,不存在冗余的点、线、面、体等数据;所采用的灰度阈值判定算法和映射网格技术,使得基体材料和空气单元能有效区分,生成的单元大小相同、形状统一,真实地还原出了泡沫铝的细观结构特性,并有利于有限元计算。

2)利用所建立的三维细观模型对泡沫铝试件在准静态压缩和动态冲击作用下的模拟分析与试验结果吻合良好,说明利用本文方法具有很高的可靠性和实用性。

3)准静态压缩作用下泡沫铝的变形在试件内部随机出现,且破坏的位置与材料的三维结构密切相关;在中等速度冲击作用下泡沫铝试件的破坏形态完全异于准静态加载作用,首先在冲击端附近开始压缩,并逐层压溃,表现出明显的变形局部化效应。

4)本文方法可用于其他具有细观结构特征材料的三维细观模型建模与研究。

参考文献(References)

[1] GIBSON L J, ASHBY M F. Cellular solids: Structure and properties[M]. Cambridge: Cambridge University Press, 1999: 6-11.

[2] BANHART J. Manufacture characterization and application of cellular metals and metal foams[J]. Progress in Materials Science, 2001, 46(6): 559-632.

[3] MAGNUCKA B E, MAGNUCKI K. Effective design of a sandwich beam with a metal foam core[J]. Thin-Walled Structures, 2007, 45(4): 432-438.

[4] YU J L, LI J R, HU S S. Strain-rate effect and micro-structural optimization of cellular metals[J]. Mechanics of Materials, 2006, 38(1): 160-170.

[5] DESHPANDE V S, FLECK N A. High strain rate compressive behavior of aluminum alloy foams[J]. International Journal of Impact Engineering, 2000, 24(3): 277-298.

[6] MEGUID S A, CHEON S S, EI-ABBASI N. FE modelling of deformation localization in metallic foams[J]. Finite Elements in Analysis and Design, 2002, 38(7): 631-643.

[7] NAMMI S K, MYLER P, EDWARDS G. Finite element analysis of closed-cell aluminum foam under quasi-static loading[J]. Materials & Design, 2010, 31(2): 712-722.

[8] ZHU H X, KNOTT J F, MILLS N J. Analysis of the elastic properties of open-cell foams with tetrakaidecahedral cells[J]. Journal of the Mechanics and Physics of Solids, 1997, 45(3): 319-343.

[9] DE GIORGI M, CAROFALO A, DATTOMA V, et al. Aluminum

foams structural modelling[J]. Computers & Structures,2010,88(1):25-35.

[10] 刘培生. 关于多孔材料的新模型[J]. 材料研究学报,2009,20(1):64-68.
LIU P S. A new model for porous materials[J]. Chinese Journal of Materials Research,2009,20(1):64-68(in Chinese).

[11] 刘培生,夏凤金,罗军. 多孔材料模型分析[J]. 材料工程,2009(7):83-87.
LIU P S,XIA F J,LUO J. Analyses of the classical model for porous materials[J]. Journal of Material Engineering,2009(7):83-87(in Chinese).

[12] ZHANG C Y,TANG L Q,YANG B,et al. Meso-mechanical study of collapse and fracture behaviors of closed-cell metallic foams[J]. Computational Materials Science,2013,79:45-51.

[13] ZHENG Z J,WANG C,YU J L,et al. Dynamic stress-strain states for metal foams using a 3D cellular model[J]. Journal of the Mechanics and Physics of Solids,2014,72:93-114.

[14] LI Z,ZHANG J,FAN J,et al. On crushing response of the three-dimensional closed-cell foam based on Voronoi model[J]. Mechanics of Materials,2014,68:85-94.

[15] CAROFALO A,DE GIORGI M,MORABITO A,et al. Geometric modelling of metallic foams[J]. Engineering Computations,2013,30(7):924-935.

[16] FANG Q,ZHANG J H,LIU J C,et al. Mesoscopic investigation of closed-cell aluminum foams on energy absorption capability under impact[J]. Composite Structures,2015,124:409-420.

[17] FANG Q,ZHANG J H,ZHANG Y D,et al. A 3D mesoscopic model for the closed-cell metallic foams subjected to static and dynamic loadings[J]. International Journal of Impact Engineering,2015,82:103-112.

[18] MAIRE E,FAZEKAS A,SALVO L,et al. X-ray tomography applied to the characterization of cellular materials related finite element modeling problems[J]. Composites Science and Technology,2003,63(16):2431-2443.

[19] MCDONALD S A,MMMMERY P M,JOHNSON G,et al. Characterization of the three-dimensional structure of a metallic foam during compressive deformation[J]. Journal of Microscopy,2006,223(2):150-158.

[20] JEON I,ASAHINA T,KANG K J,et al. Finite element simulation of the plastic collapse of closed-cell aluminum foams with X-ray computed tomography[J]. Mechanics of Materials,2010,42(3):227-236.

[21] SUN Y L,LI Q M,LOWE T,et al. Investigation of strain-rate effect on the compressive behavior of closed-cell aluminum foam by 3D image-based modelling[J]. Materials & Design,2016,89:215-224.

[22] JEON I,ASAHINA T. The effect of structural defects on the compressive behavior of closed-cell Al foam[J]. Acta Materialia,2005,53(12):3415-3423.

[23] TURBELL H. Cone-beam reconstruction using filtered back projection[D]. Linkping:Linkping University,2001.

[24] 韩思奇,王蕾. 图像分割的阈值法综述[J]. 系统工程与电子技术,2002,24(6):91-94.
HAN S Q,WANG L. A survey of thresholding methods for image segmentation[J]. Systems Engineering and Electronics,2002,24(6):91-94(in Chinese).

[25] MCDONALD S A,MMMMERY P M,JOHNSON G,et al. Characterization of the three-dimensional structure of a metallic foam during compressive deformation[J]. Journal of Microscopy,2006,223(Pt 2):150-158.

[26] HALLQUISTJ O. LS-DYNA keyword user's manual[Z]. Livermore:Livermore Software Technology Corporation,2015.

[27] JEON I,KATOU K,SONODA T,et al. Cell wall mechanical properties of closed-cell Al foam[J]. Mechanics of Materials,2009,41(1):60-73.

[28] 李妍妍,郑志军,虞吉林,等. 闭孔泡沫金属变形模式的有限元分析[J]. 爆炸与冲击,2014,34(4):464-470.
LI Y Y,ZHENG Z J,YU J L,et al. Finite element analysis on deformation modes of closed-cell metallic foam[J]. Explosion and Shock Waves,2014,34(4):464-470(in Chinese).

作者简介:

李侯贞强　男,硕士研究生。主要研究方向:材料细观力学。

张亚栋　男,教授,硕士生导师。主要研究方向:防护工程。

北京航空航天大学学报

北航学报 赠阅

Reconstruction and application of three-dimensional mesoscopic model of aluminum foam based on CT

LI Houzhenqiang[1,2], ZHANG Yadong[1,*], ZHANG Jinhua[1], JIANG Chunlin[1]

(1. State Key Laboratory of Disaster Prevention and Mitigation of Explosion and Impact, Army Engineering University of PLA, Nanjing 210007, China; 2. Taiyuan Satellite Launch Center, Taiyuan 036304, China)

Abstract: In order to obtain a more realistic mesoscopic analysis model of closed-cell aluminum foam, a new methodology for the finite element modeling based on computed tomography (CT) images is presented. First, the optimal threshold between base material and air was developed using Otsu algorithm by analyzing the images obtained from the CT scanning of closed-cell aluminum foam. Then, the mesoscopic finite element model was directly established based on the thought of mapping grid. As a result, the reconstruction of three-dimensional mesoscopic analysis model of metal foams is achieved. Finally, the numerical simulations of quasi-static compression and dynamic test of closed-cell foam are carried out respectively based on the mesoscopic analysis model. The results demonstrate that the internal deformation of closed-cell aluminum foam distributes throughout the whole specimen, which is closely bound up with their 3D structure under quasi-static compression, while it is close to the loading end and remarkably behaves with localization under dynamic compression. The methodology of modeling can describe mesoscopic structure realistically and provide a more detailed simulation analysis on the stress state, deformation and failure of closed-cell aluminum foams under quasi-static and dynamic loading.

Keywords: closed-cell aluminum foam; CT images; mesoscopic model; reconstruction; numerical simulation

北京航空航天大学学报

Received: 2016-12-21; **Accepted**: 2017-03-17; **Published online**: 2017-04-13 17:48

URL: kns.cnki.net/kcms/detail/11.2625.V.20170413.1748.007.html

Foundation items: National Key R&D Program of China (2014YFC0305200); National Natural Science Foundation of China (51478464, 51678566); National Key Scientific Instrument and Equipment Development Project of China (2014YQ24044509)

* **Corresponding author**. E-mail: zhydjs@139.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2016.0958

脉冲星角位置对脉冲模板的影响及其削弱策略

信世军，郑伟*，王奕迪

（国防科学技术大学 航天科学与工程学院，长沙 410073）

摘 要：构建脉冲模板是X射线脉冲星导航的一项关键技术，其精度与脉冲星角位置精度密切相关。基于脉冲模板构建的基本原理，推导得到了脉冲星角位置误差对脉冲模板构建影响的系统误差的年化平均值以及任意弧段积分的解析表达式，分析了脉冲星角位置误差对脉冲模板构建的影响。在此基础上，给出了削弱脉冲星角位置误差影响的方法。通过理论推导和仿真分析，验证了误差削弱方法的有效性，可以为优化X射线脉冲星的观测任务以及X射线脉冲模板的构建提供理论支持。

关 键 词：X射线脉冲星导航；脉冲模板；脉冲星角位置误差；周年特性；削弱策略
中图分类号：V448.22+4
文献标识码：A **文章编号**：1001-5965(2018)01-0169-07

随着航天事业的发展，在轨航天器日益增多。若仍采用依赖地面测控支持的导航模式，不仅会提高航天器的运营成本，而且无法确保航天器在特定情况下的生存能力。因此，自主导航是未来航天器发展的核心技术和关键技术[1]。目前，借助全球卫星导航系统(Global Navigation Satellite System，GNSS)低轨航天器已基本实现了自主导航，可完成如自主交会对接、短期自主运行等任务。局限于导航卫星的轨道高度和电磁信号的辐射角约束，GNSS难以应用于高轨航天器和深空探测器的自主导航任务[2]。研究适用于上述2种航天器的自主导航技术，是业界研究的热点。

X射线脉冲星导航是一种新兴的航天器自主导航方法，优势有：能提供高精度的参考时间基准，导航精度高，对导航探测器精度要求较低。同卫星导航相比，X射线脉冲星导航还具有可同时服务于近地航天器和深空探测器的优势[3]。由于其具有广阔的应用前景，X射线脉冲星导航技术得到了国内外的广泛关注[4-7]。经过近30年的发展，该导航技术已经逐步从概念研究阶段进入关键技术攻坚阶段。

脉冲星数据库的构建是X射线脉冲星导航的一项关键技术。构建脉冲星数据库涵盖了脉冲星的空间分布、周期分布和信号等参数的测定工作[8]。其中，脉冲星信号的脉冲模板是脉冲星信号处理的基本输入量，也是脉冲星数据库构建中的一项重要内容。将在轨恢复的轮廓与脉冲模板对比，可以估计出脉冲到达航天器的到达时间(Time of Arrival，TOA)[9]。TOA是X射线脉冲星导航的基本测量量。若脉冲模板存在偏差，则会导致TOA的估计结果出现偏差，从而降低了导航系统的性能。为此，分析影响脉冲模板的因素，并提出相应的削弱策略，可为提升脉冲星导航的性能提供有力支持。

本文首先简要介绍了脉冲星脉冲模板构建的基本原理。然后着重分析了脉冲星角位置误差对模板构建的影响，推导得到此系统误差的年化平均值以及任意弧段积分的解析表达式。最后提出

收稿日期：2016-12-21；**录用日期**：2017-01-20；**网络出版时间**：2017-03-08 09:29
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170308.0929.003.html
基金项目：国家自然科学基金（10973048）
* **通信作者**. E-mail：zhengwei@nudt.edu.cn

引用格式：信世军，郑伟，王奕迪. 脉冲星角位置对脉冲模板的影响及其削弱策略[J]. 北京航空航天大学学报，2018，44(1)：169-175. XIN S J，ZHENG W，WANG Y D. Impact of pulsar angular position on pulse template and its compensation method[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(1)：169-175(in Chinese).

一种系统误差削弱策略。

1 脉冲星脉冲模板构建的基本原理

在实际的空间 X 射线脉冲星观测中，脉冲星信号的观测量就是脉冲星辐射的 X 射线光子[10]。由于存在脉冲星自转频率变化、航天器轨道信息不准确等因素，难以在轨恢复出高精度的脉冲星脉冲模板。为此，将探测到的光子 TOA 序列转换到太阳系质心(Solar System Barycentre, SSB)以消除航天器运动、相对论效应等影响外，还需搜索信号周期，并将每个时段获得的轮廓进行相位对齐[11]。

1.1 光子到达时间转换

考虑到周年视差、Roemer 延迟、色散延缓和引力时延等效应，令光子到达探测器的时间为 t_{SC}，光子到达真空 SSB 的时间为 t_{SSB}，相应的时间转换模型可写为

$$t_{SSB}=t_{SC}+\frac{1}{c}\boldsymbol{n}\cdot\boldsymbol{r}_{SC}+\frac{1}{c}\left[-\frac{\|\boldsymbol{r}_{SC}\|^2}{2D_0}+\frac{(\boldsymbol{n}\cdot\boldsymbol{r}_{SC})^2}{2D_0}-\frac{\boldsymbol{b}\cdot\boldsymbol{r}_{SC}}{D_0}+\frac{(\boldsymbol{n}\cdot\boldsymbol{b})(\boldsymbol{n}\cdot\boldsymbol{r}_{SC})}{D_0}\right]+\sum_{i=1}^{9}\frac{2\mu_i}{c^3}\ln|\boldsymbol{n}\cdot\boldsymbol{p}_i+\|\boldsymbol{p}_i\||\tag{1}$$

式中：等号右边第 2 项为一阶多普勒延迟，第 3 项为由周年视差效应引起的，第 4 项为 Shapiro 延迟效应；$\boldsymbol{n}$ 为脉冲星位置矢量；$\boldsymbol{r}_{SC}$ 为航天器相对于 SSB 的位置矢量；c 为光速；D_0 为脉冲星在基准传播时间 T_0 时的位置；$\boldsymbol{b}$ 为 SSB 相对于太阳质心的位置矢量；$\boldsymbol{p}_i$ 为航天器相对于第 i 颗行星的位置矢量；μ_i 为第 i 颗行星的引力常数。若观测脉冲星的赤经和赤纬分别为 α 和 δ，则其矢量方向为

$$\boldsymbol{n}=\begin{bmatrix}\cos\alpha\cos\delta\\ \sin\alpha\cos\delta\\ \sin\delta\end{bmatrix}\tag{2}$$

1.2 周期搜索

假设在观测时间 t 内，探测器可接收到一段光子 TOA 序列。按照试验周期 P 进行历元折叠，即将该光子 TOA 序列分为 N_P 段。然后，将 N_P 段长度为试验周期 P 的光子 TOA 序列分为 N_b 份，即将一个周期的相位分为 N_b 份。各个试验周期的光子 TOA 依次累积按相位到第 1 段上，得到第 i 个相位的光子数为 c_i(见图 1)[12]。建立 χ^2 为

$$\chi^2=\sum_{i=1}^{N_b}(c_i-\bar{c})^2/\bar{c}\tag{3}$$

式中：$\bar{c}$ 为平均光子数。

一直变化试验周期 P，使 χ^2 达到最大值。χ^2 最大值对应的 P 即为脉冲星的自转周期。

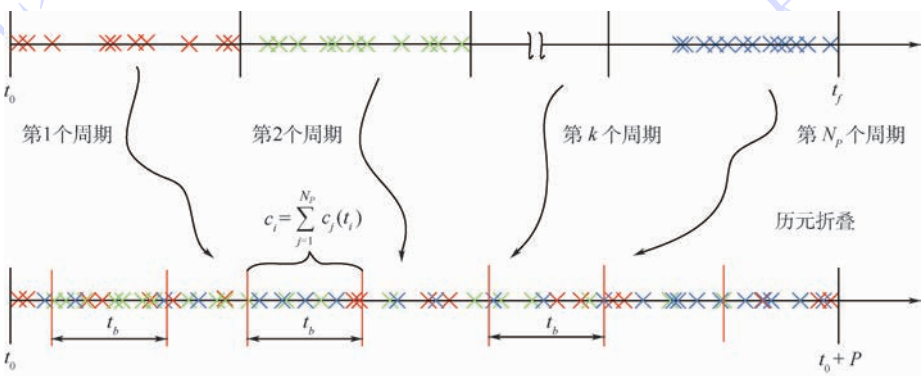


图 1 历元折叠过程

Fig. 1 Epoch folding process

1.3 相位对齐

通过 1.2 节周期寻找过程，可以获得最佳的折合周期 P_{best}。由于每段观测时间的脉冲星信号初相不同，需要把不同观测时段的初相对齐。因此，可通过交叉互相关、傅里叶变换法等将后续观测时段得到的轮廓与第 1 个观测时段得到的轮廓作比较，求得初相差。根据不同的初相差，就可将后续时段得到的恢复轮廓折合到第 1 个时段，从而获得高信噪比的信号轮廓。

2 脉冲星角位置误差的影响

目前通常采用甚长基线干涉测量(Very Long Baseline Interferometer, VLBI)来测定脉冲星的角位置。基于当前的 VLBI 测量水平，脉冲星角位置的精度可以达到 1 mas(毫角秒)附近[13]。

2.1 脉冲星角位置误差引起的系统误差

设脉冲星角位置的测量误差为($\Delta\alpha$, $\Delta\delta$)，则脉冲星指向误差为

北航学报 赠阅

$$\Delta \boldsymbol{n}_0 = \begin{bmatrix} -\sin\delta\cos\alpha \\ -\sin\delta\sin\alpha \\ \cos\delta \end{bmatrix}\Delta\delta + \begin{bmatrix} -\cos\delta\sin\alpha \\ \cos\delta\cos\alpha \\ 0 \end{bmatrix}\Delta\alpha \quad (4)$$

则有脉冲星指向为

$$\boldsymbol{n} = \boldsymbol{n}_0 + \Delta\boldsymbol{n}_0 \quad (5)$$

式中:$\boldsymbol{n}_0$ 为脉冲星的实际方向矢量。

从式(1)可知,脉冲星的角位置误差通过影响脉冲星的方向矢量来影响时间转换精度,从而影响脉冲模板构建的精度。因此,式(1)对脉冲星方向矢量的导数可用来衡量脉冲星角位置误差的影响。将 $\boldsymbol{p}_i = \boldsymbol{r}_{SC} - \boldsymbol{b}_i$ 代入式(1),得到其具体的误差传播函数为

$$\Delta t = \frac{1}{c}\boldsymbol{r}_{SC}\cdot\Delta\boldsymbol{n}_0 + \frac{1}{c}\left[\frac{(\boldsymbol{n}\cdot\boldsymbol{r}_{SC})(\boldsymbol{r}_{SC}\cdot\Delta\boldsymbol{n}_0)}{D_0} + \frac{(\Delta\boldsymbol{n}_0\cdot\boldsymbol{b})(\boldsymbol{n}\cdot\boldsymbol{r}_{SC}) + (\boldsymbol{n}\cdot\boldsymbol{b})(\Delta\boldsymbol{n}_0\cdot\boldsymbol{r}_{SC})}{D_0}\right] + \sum_{i=1}^{9}\left(\frac{2\mu_i}{c^3}\cdot\frac{\Delta\boldsymbol{n}_0\cdot(\boldsymbol{r}_{SC}-\boldsymbol{b}_i)}{\boldsymbol{n}\cdot(\boldsymbol{r}_{SC}-\boldsymbol{b}_i) + \|\boldsymbol{r}_{SC}-\boldsymbol{b}_i\|}\right) \quad (6)$$

式中:$\boldsymbol{b}_i$ 为第 i 颗行星相对 SSB 的位置矢量。航天器到 SSB 的矢量 $\boldsymbol{r}_{SC}$ 可表示为

$$\boldsymbol{r}_{SC} = \boldsymbol{r}_E + \boldsymbol{r}_{S/E} \quad (7)$$

式中:$\boldsymbol{r}_E$ 为地球到 SSB 的位置矢量;$\boldsymbol{r}_{S/E}$ 为航天器到地球的位置矢量。对于高轨航天器,$\|\boldsymbol{r}_{S/E}\| \ll \|\boldsymbol{r}_E\|$ 总成立,分析系统误差影响时可略去 $\boldsymbol{r}_{S/E}$ 的影响,即

$$\boldsymbol{r}_{SC} \approx \boldsymbol{r}_E \quad (8)$$

则式(6)可写为

$$\Delta t = \frac{1}{c}\boldsymbol{r}_E\cdot\Delta\boldsymbol{n}_0 + \frac{1}{c}\left[\frac{(\boldsymbol{n}\cdot\boldsymbol{r}_E)(\boldsymbol{r}_E\cdot\Delta\boldsymbol{n}_0)}{D_0} + \frac{(\Delta\boldsymbol{n}_0\cdot\boldsymbol{b})(\boldsymbol{n}\cdot\boldsymbol{r}_E) + (\boldsymbol{n}\cdot\boldsymbol{b})(\Delta\boldsymbol{n}_0\cdot\boldsymbol{r}_E)}{D_0}\right] + \sum_{i=1}^{8}\left(\frac{2\mu_i}{c^3}\cdot\frac{\Delta\boldsymbol{n}_0\cdot(\boldsymbol{r}_E-\boldsymbol{b}_i)}{\boldsymbol{n}\cdot(\boldsymbol{r}_E-\boldsymbol{b}_i) + \|\boldsymbol{r}_E-\boldsymbol{b}_i\|}\right) \quad (9)$$

记式(9)等式右侧的 3 项分别为 Δt_1、Δt_2、Δt_3,以 Crab 脉冲星为例,基于 DE405 进行仿真,取其角位置误差为 1 mas,分别绘制脉冲星角位置误差引起的总的系统误差与 Δt_1、Δt_2、Δt_3 随时间变化曲线,结果如图 2 所示。

目前对于简化的时间转换模型而言,其精度约是 1 μs,从图 2 可以看出,1 mas 的脉冲星角位置误差的影响在全年呈周期变化,且在全年大部分时间的影响大于 1 μs,这一误差的影响极为可观。另外,脉冲星角位置误差所引起时延主要是 Δt_1 所造成的。故分析脉冲星角位置误差对脉冲模板构建的影响时,其误差传播方程可简化为

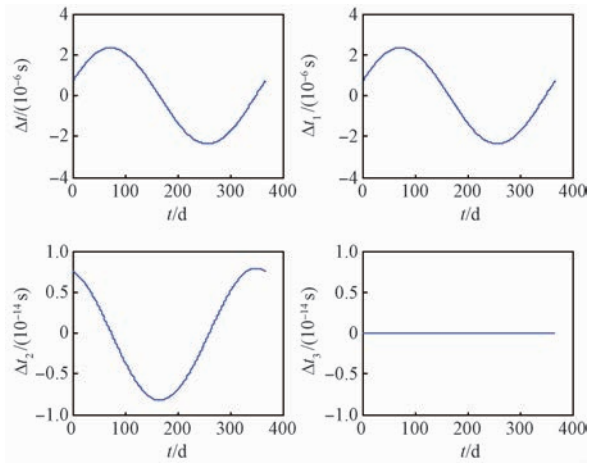


图 2　脉冲星角位置误差引起的系统误差

Fig. 2　Systom error caused by pulsar angular position error

$$\Delta t = \frac{1}{c}\boldsymbol{r}_E\cdot\Delta\boldsymbol{n}_0 \quad (10)$$

2.2　系统误差的特性分析

2.2.1　航天器到太阳系质心位置矢量表示

不考虑摄动影响以及地球轨道根数的不确定性,则地球相对 SSB 的位置矢量 $\boldsymbol{r}_E$ 可表示为

$$\boldsymbol{r}_E = \boldsymbol{M}_3(-\Omega)\boldsymbol{M}_1(-i)\boldsymbol{M}_3(-\omega)\boldsymbol{r}_{E/O} = \boldsymbol{M}\boldsymbol{r}_{E/O} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}\begin{bmatrix} a(\cos E - e) \\ a\sqrt{1-e^2}\sin E \\ 0 \end{bmatrix} = \begin{bmatrix} m_{11}a(\cos E - e) + m_{12}a\sqrt{1-e^2}\sin E \\ m_{21}a(\cos E - e) + m_{22}a\sqrt{1-e^2}\sin E \\ m_{31}a(\cos E - e) + m_{32}a\sqrt{1-e^2}\sin E \end{bmatrix} \quad (11)$$

式中:$\boldsymbol{M}$ 为地球轨道坐标系到太阳系质心惯性系的坐标转换矩阵;$\boldsymbol{M}_1$ 和 $\boldsymbol{M}_3$ 分别为绕 X 轴和 Z 轴旋转的方向余弦阵;a, e, i, Ω, ω 为地球绕 SSB 运动轨道的轨道根数;E 为偏近点角;$\boldsymbol{r}_{E/O}$ 为地球在轨道坐标系中表示的位置矢量,即

$$\boldsymbol{r}_{E/O} = \begin{bmatrix} a(\cos E - e) \\ a\sqrt{1-e^2}\sin E \\ 0 \end{bmatrix} \quad (12)$$

2.2.2　系统误差的周年特性

令

$$\Delta\tilde{t} = \frac{1}{T}\int_0^T \Delta t\,\mathrm{d}t \quad (13)$$

为系统误差的年化平均值。$T = 365.242\ 2$ d 为地球公转周期。在较大时间尺度上认为脉冲星光子到达为均匀分布。将式(10)代入式(13)得到

$$\Delta\tilde{t} = \frac{1}{T}\int_0^T \frac{1}{c}\boldsymbol{r}_E\cdot\Delta\boldsymbol{n}_0\,\mathrm{d}t \quad (14)$$

北航学报 赠阅

式中:$\Delta \boldsymbol{n}_0$ 为与时间无关的常矢量。将式(11)与式(12)代入式(14),考虑到

$$\mathrm{d}E = \frac{a}{r} n \mathrm{d}t \tag{15}$$

式中:$r = \|\boldsymbol{r}_{\mathrm{E}}\|$;$n = \sqrt{\mu / a^3}$ 为地球公转的平均角速度,μ 为地心引力常数。则式(14)的被积函数可以写为形如

$$F = k_1 + k_2 \cos E + k_3 \sin E \tag{16}$$

的形式,则有

$$\int_0^T (k_1 + k_2 \cos E + k_3 \sin E) \mathrm{d}t = \int_0^{2\pi} \frac{r}{an} (k_1 + k_2 \cos E + k_3 \sin E) \mathrm{d}E \tag{17}$$

将 $r = a(1 - e\cos E)$ 代入式(17),则可转化为求解以下 3 个积分:

$$\begin{cases} \int_0^{2\pi} \cos E \mathrm{d}E \\ \int_0^{2\pi} \cos^2 E \mathrm{d}E \\ \int_0^{2\pi} \sin E \cos E \mathrm{d}E \end{cases} \tag{18}$$

由于

$$\begin{cases} \int_0^{2\pi} \cos E \mathrm{d}E = 0 \\ \int_0^{2\pi} \cos^2 E \mathrm{d}E = \pi \\ \int_0^{2\pi} \sin E \cos E \mathrm{d}E = 0 \end{cases} \tag{19}$$

因此可以得出:①按年累积数据,系统误差的影响会大幅度下降。②角位置误差引起的系统误差中的周期项虽会被消除,但会留下一个长期项。

下面通过公式推导得到其周年特性的解析表达式,以便于仿真验证。

将 $r = a(1 - e\cos E)$ 代入式(15)得到

$$\mathrm{d}t = \frac{1 - e\cos E}{n} \mathrm{d}E \tag{20}$$

将式(11)、式(12)及式(20)代入式(14),进行推导,可得

$$\Delta \tilde{t} = \frac{1}{T}\int_0^T \frac{1}{c} \boldsymbol{r}_{\mathrm{E}} \cdot \Delta \boldsymbol{n}_0 \mathrm{d}t = \frac{1}{cTn}\int_0^{2\pi} (1 - e\cos E) \boldsymbol{r}_{\mathrm{E}} \cdot \Delta \boldsymbol{n}_0 \mathrm{d}E = \frac{ak_1}{cTn}\int_0^{2\pi} [-e\cos^2 E + (1 + e^2)\cos E - e] \mathrm{d}E + \frac{ak_2}{cTn}\int_0^{2\pi} (\sin E - e\sin E\cos E) \mathrm{d}E \tag{21}$$

式中:

$$\begin{cases} k_1 = \Delta \boldsymbol{n}_0 \cdot \begin{bmatrix} m_{11} \\ m_{21} \\ m_{31} \end{bmatrix} \\ k_2 = \sqrt{1 - e^2}\, \Delta \boldsymbol{n}_0 \cdot \begin{bmatrix} m_{12} \\ m_{22} \\ m_{32} \end{bmatrix} \end{cases} \tag{22}$$

又

$$\begin{cases} \int_0^{2\pi} \cos E \mathrm{d}E = 0 \\ \int_0^{2\pi} \cos^2 E \mathrm{d}E = \pi \\ \int_0^{2\pi} \sin E \cos E \mathrm{d}E = 0 \\ \int_0^{2\pi} \sin E \mathrm{d}E = 0 \end{cases} \tag{23}$$

则有

$$\begin{aligned} \Delta \tilde{t} = {} & \frac{1}{T}\int_0^T \frac{1}{c} \boldsymbol{r}_{\mathrm{E}} \cdot \Delta \boldsymbol{n}_0 \mathrm{d}t = \frac{1}{cTn}\int_0^{2\pi} (1 - e\cos E) \boldsymbol{r}_{\mathrm{E}} \cdot \Delta \boldsymbol{n}_0 \mathrm{d}E = \\ & -\frac{3a\pi e}{cTn} [m_{11}(-\Delta\delta \sin\delta\cos\alpha - \Delta\alpha\cos\delta\sin\alpha) + \\ & m_{21}(-\Delta\delta\sin\delta\sin\alpha + \Delta\alpha\cos\delta\cos\alpha) + \\ & m_{31}\Delta\delta\cos\delta] = -\frac{3a\pi e}{cTn} \{-(\cos\Omega\cos\omega - \\ & \sin\Omega\sin i\sin\omega)\sin\delta\cos\alpha - \\ & (\sin\Omega\cos\omega - \cos\Omega\cos i\sin\omega) \cdot \\ & \sin\delta\sin\alpha + \sin i\sin\omega\cos\delta]\Delta\delta + \\ & [(\sin\Omega\cos\omega - \cos\Omega\cos i\sin\omega) \cdot \\ & \cos\delta\cos\alpha - (\cos\Omega\cos\omega - \sin\Omega\sin i\sin\omega) \cdot \\ & \cos\delta\sin\alpha\}\Delta\alpha \end{aligned} \tag{24}$$

根据式(24)可知,若地球轨道根数和脉冲星的角位置误差确定,则角位置误差所引起系统误差 Δt 的年化平均值 $\Delta \tilde{t}$ 仅取决于脉冲星的角位置 (α, δ)。

根据文献[14],取地球公转轨道的轨道根数如表 1 所示。

取 1 mas 的脉冲星角位置误差,对文献[15]中提及的 10 个脉冲星,分别计算其 Δt 的最大值与 $\Delta \tilde{t}$,结果如表 2 所示。

表 1 地球轨道根数

Table 1 Orbit elements of the earth

轨道根数	数 值
半长轴/AU	0.999 985 8
偏心率	0.016 678 35
轨道倾角/(°)	23.44
升交点黄经/(°)	0
近日点黄经/(°)	103.302 75

北航学报 赠阅

表 2　脉冲星角位置误差引起的系统误差的最大值和年化平均值

Table 2　Maximum value and annual average of system error caused by pulsar angular position error

脉 冲 星	Max(Δt)/μs	$\Delta\tilde{t}$/(10^{-8} s)
J0534 +2200	2.3388	−2.0555
J1513 −5908	1.7748	3.8512
J1846 −0258	2.6979	−1.7737
J0540 −6919	2.5157	−6.4059
J0835 −4510	2.5553	−3.6324
J1824 −2452	2.2959	0.9289
J1959 −2048	2.6756	−1.6696
J0437 −4715	2.8082	−7.1214
J0218 +4232	2.6231	−4.2932
J1939 +2134	2.9625	−5.1393

由表 2 可以看出,对此项系统误差取年化平均值均可将其大幅度削弱,但仍存在长期项。从而验证了由式(13) ~ 式(19)的理论分析所得到的 2 个结论。

3　系统误差的削弱策略

由 2.2.2 节分析可知,脉冲星角位置误差的影响函数是一个以一年为周期的类似正弦函数。因此,为了削弱脉冲星角位置误差对脉冲模板构建的影响,在进行历元累积时,应尽量选择相对正弦曲线零点对称的时间段。

实际上,考虑到太阳规避等因素,无法对 X 射线脉冲星进行全年观测。为此,需考虑任意弧段下测量误差影响的特性。取弧段(E_1,E_2),对应时间段为(t_1,t_2),根据式(21) ~ 式(24)可得

$$\begin{aligned}\Delta\tilde{t} &= \frac{1}{t_2-t_1}\int_{t_1}^{t_2}\frac{1}{c}\boldsymbol{r}_{\mathrm{E}}\cdot\Delta\boldsymbol{n}_0\mathrm{d}t = \\ &\frac{1}{cn(t_2-t_1)}\int_{E_1}^{E_2}(1-e\cos E)\boldsymbol{r}_{\mathrm{E}}\cdot\Delta\boldsymbol{n}_0\mathrm{d}E = \\ &\frac{ak_1}{cn(t_2-t_1)}\int_{E_1}^{E_2}[-e\cos^2E+(1+e^2)\cos E- \\ &e]\mathrm{d}E+\frac{ak_2}{cn(t_2-t_1)}\int_{E_1}^{E_2}(\sin E-e\sin E\cos E)\mathrm{d}E\end{aligned} \tag{25}$$

由于

$$\begin{cases}\displaystyle\int_{E_1}^{E_2}\cos E\mathrm{d}E=\sin E_2-\sin E_1\\ \displaystyle\int_{E_1}^{E_2}\sin E\mathrm{d}E=\cos E_1-\cos E_2\\ \displaystyle\int_{E_1}^{E_2}\cos^2E\mathrm{d}E=\frac{1}{4}(\sin(2E_2)-\sin(2E_1))+\\ \qquad\displaystyle\frac{1}{2}(E_2-E_1)\\ \displaystyle\int_{E_1}^{E_2}\sin E\cos E\mathrm{d}E=\frac{1}{4}(\cos(2E_1)-\cos(2E_2))\end{cases} \tag{26}$$

将式(26)代入式(25)可得

$$\begin{aligned}\Delta\tilde{t}(E_1,E_2)=&\frac{ak_1}{cn(t_2-t_1)}\int_{E_1}^{E_2}[-e\cos^2E+(1+e^2)\cdot\\ &\cos E-e]\mathrm{d}E+\frac{ak_2}{cn(t_2-t_1)}\int_{E_1}^{E_2}(\sin E-\\ &e\sin E\cos E)\mathrm{d}E=\\ &\frac{ak_3}{cn(t_2-t_1)}[(-m_{11}\sin\delta\cos\alpha-\\ &m_{21}\sin\delta\sin\alpha+m_{31}\cos\delta)\Delta\delta+\\ &(-m_{11}\cos\delta\sin\alpha+m_{21}\cos\delta\cos\alpha)\Delta\alpha]+\\ &\frac{ak_4}{cn(t_2-t_1)}[(-m_{12}\sin\delta\cos\alpha-\\ &m_{22}\sin\delta\sin\alpha+m_{32}\cos\delta)\Delta\delta+\\ &(-m_{12}\cos\delta\sin\alpha+m_{22}\cos\delta\cos\alpha)\Delta\alpha]=\\ &\frac{a}{cn(t_2-t_1)}\{[-(k_3m_{11}+k_4m_{12})\sin\delta\cos\alpha-\\ &(k_3m_{21}+k_4m_{22})\sin\delta\sin\alpha+\\ &(k_3m_{31}+k_4m_{32})\cos\delta]\Delta\delta+\\ &[-(k_3m_{11}+k_4m_{12})\cos\delta\sin\alpha+\\ &(k_3m_{21}+k_4m_{22})\cos\delta\cos\alpha]\Delta\alpha\}\end{aligned} \tag{27}$$

式中:

$$\begin{cases}k_3=-e\left[\dfrac{1}{4}(\sin(2E_2)-\sin(2E_1))+\right.\\ \qquad\left.\dfrac{3}{2}(E_2-E_1)\right]+(1+e^2)(\sin E_2-\sin E_1)\\ k_4=\sqrt{1-e^2}[(\cos E_1-\cos E_2)-\\ \qquad\dfrac{e}{4}(\cos(2E_1)-\cos(2E_2))]\end{cases} \tag{28}$$

基于蒙特卡罗打靶,生成每组观测时长为 1 h 的 1000 组光子 TOA 序列,在 1 年中每 2 d 进行 1 h 的仿真模拟,图 3 给出了 1 mas 脉冲星角位置误差导致的系统误差在一年内的变化情况以及通过历元累积对此系统误差的削弱情况。

由图 3 可知,对于非对称时间段的系统误差进行历元累积,此误差不断增大,之后随着对称时间段数据的加入,可使其大幅度削弱。通过一年的观测,可将 1 mas 脉冲星角位置误差的影响从 1 ~ 2 μs 降低到 0.1 μs 附近。因此,若按周期性系统误差的周期对称选取时间段进行历元累积,可削弱该系统误差的影响。

在实际观测中,脉冲星并非全年可见,比如 Crab 脉冲星在 6 月份不可见。以 Crab 脉冲星为例,表 3 显示了考虑其不可见时段时选取对称时段与非对称时段进行历元累积的结果。由表 3 可以看出,当选取对称时段进行历元累积构建脉

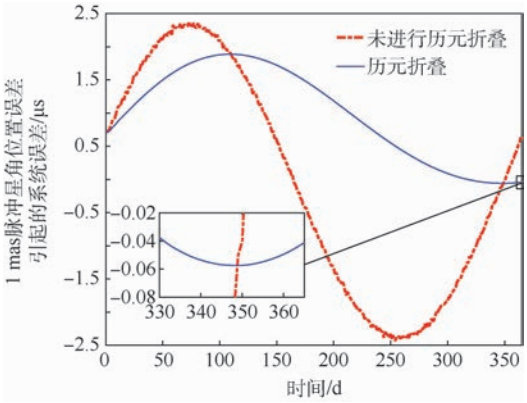


图 3 历元累积对脉冲星角位置误差的削弱情况

Fig. 3 Pulsar angular position error reduced by epoch accumulation

表 3 选取 Crab 脉冲星的对称时段进行历元累积与非对称情况对比

Table 3 Epoch accumulation results comparison for symmetric and asymmetric cases of Crab pulsar

时 段	所选具体时间段/d	脉冲最终系统误差/μs
对称时段	1 ~ 90, 241 ~ 330	−0.0398
非对称时段	1 ~ 120, 241 ~ 330	0.6100

冲模板时,其系统误差明显低于选取非对称时段的情况。

综上所述,在构建脉冲模板时,根据角位置误差所造成的系统误差的周期选取对称时段进行历元累积可以有效削弱此项系统误差,进而提高脉冲模板的精度。

4 结 论

1) 构建脉冲模板时,取 1 mas 的脉冲星角位置误差,其造成的系统误差在 1 年内的大部分时间大于 1 μs,不可忽略。

2) 经过公式推导,得到此系统误差的年化平均值的解析表达式,计算结果表明,对脉冲星角位置误差引起的系统误差取年化平均值可使其大幅度削弱,但仍存在长期项。

3) 以 Crab 脉冲星为例,按周期性系统误差的周期对称选取时间段进行历元累积,可将 1 mas 脉冲星角位置误差的影响从 1 ~ 2 μs 降低到 0.1 μs附近。

4) 在构建脉冲模板时,根据角位置误差所造成的系统误差的周期选取对称时段进行历元累积可以有效削弱此项系统误差,进而提高脉冲模板的精度。

5) 本研究可以为优化 X 射线脉冲星的观测任务以及 X 射线脉冲模板构建提供理论支持。

参考文献 (References)

[1] 郑伟,王奕迪,汤国建,等. X 射线脉冲星导航理论与应用[M]. 北京:科学出版社,2015:1-20.
ZHENG W, WANG Y D, TANG G J, et al. X-ray pulsar-based navigation: Theory and applications[M]. Beijing: Science Press, 2015:1-20(in Chinese).

[2] 何清举,孙前贵. 利用 GNSS 实现高轨卫星自主导航的新方案[J]. 飞行器测控学报, 2010,29(1):7-11.
HE Q J, SUN Q G. A new solution of autonomous navigation for GEO satellites based on GNSS[J]. Journal of Spacecraft TT&C Technology, 2010, 29(1):7-11(in Chinese).

[3] 帅平,李明,陈绍龙,等. X 射线脉冲星导航系统原理与方法[M]. 北京:中国宇航出版社,2009.
SHUAI P, LI M, CHEN S L, et al. X-ray pulsar-based navigation system: Theory and methods[M]. Beijing: China Astronautic Publishing House, 2009(in Chinese).

[4] SHEIKH S I. The use of variable celestial X-ray sources for spacecraft navigation[D]. Washington, D. C.: University of Maryland, 2005.

[5] WOODFORK D W. Use of X-ray pulsar for aiding GPS satellite orbit determination[D]. Alabama: Air University, 2005.

[6] 邓新坪. X 射线脉冲星自主导航研究[D]. 北京:中国科学院大学,2013.
DENG X P. Research on autonomous navigation using X-ray pulsars[D]. Beijing: University of Chinese Academy of Sciences, 2013(in Chinese).

[7] WANG Y D, ZHENG W, SUN S M, et al. X-ray pulsar-based navigation system with the errors in the planetary ephemerides for earth-orbiting satellite[J]. Advances in Space Research, 2013, 51(12):2394-2404.

[8] 周庆勇,刘思伟,郝晓龙,等. 空间 X 射线观测确定脉冲星星历表参数精度分析[J]. 物理学报,2016,65(7):368-377.
ZHOU Q Y, LIU S W, HAO X L, et al. Analysis of measurement accuracy of ephemeris parameters for pulsar navigation based on the X-ray space observation[J]. Acta Physica Sinica, 2016, 65(7):368-377(in Chinese).

[9] EMADZADEH A A, SPEYER J L, HADAEHG F Y. A parametric study of relative navigation using pulsars[C]// Proceeding of the 63rd Annual Meeting of the Institute Navigation, 2007: 454-459.

[10] LIU J, MA J, TIAN J, et al. X-ray pulsar navigation method for spacecraft with pulsar direction error[J]. Advances in Space Research, 2010, 46(11):1409-1417.

[11] 葛明玉. 脉冲星的 X 射线辐射特性研究[D]. 北京:中国科学院研究生院,2012.
GE M Y. The X-ray emission of pulsars[D]. Beijing: Graduate University of Chinese Academy of Sciences, 2012(in Chinese).

[12] EMADZADEH A A, SPEYER J L. Navigation in space by X-ray pulsars[M]. Berlin: Springer, 2010:13-33.

[13] 王奕迪. X 射线脉冲星信号处理与导航方法研究[D]. 长沙:国防科学技术大学,2016.
WANG Y D. X-ray pulsar-based navigation: Signal processing and positioning algorithms[D]. Changsha: National University

北航学报 赠阅

of Defense Technology,2016(in Chinese).

[14] 张洪波.航天器轨道力学理论与方法[M].北京:国防工业出版社,2015.
ZHANG H B. Theories and methods of spacecraft orbital mechanics[M]. Beijing: National Defend Industry Press,2015(in Chinese).

[15] 费保俊.相对论在现代导航中的应用[M].北京:国防工业出版社,2015:176-177.
FEI B J. Application of relativity in modern navigation[M]. Beijing: National Defend Industry Press,2015:176-177(in Chinese).

作者简介:

信世军　男,硕士研究生。主要研究方向:X 射线脉冲星导航。

郑伟　男,博士,教授,博士生导师。主要研究方向:飞行动力学与控制。

王奕迪　男,博士,讲师。主要研究方向:X 射线脉冲星导航。

Impact of pulsar angular position on pulse template and its compensation method

XIN Shijun, ZHENG Wei*, WANG Yidi

(College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: Recovering pulse template is one key technology of X-ray pulsar-based navigation system. Its precision is closely related to the pulsar angular position. First, this paper briefly introduced the theory of recovering standard profile. Then, the impact of pulsar angular position error on pulse template was analyzed and its analytical formula of annual mean value and the integral for any arc segment were derived. Finally, we proposed a feasible way that can significantly decrease the impact of pulsar angular position error. These research conclusions and simulation analysis verify the effectiveness of error compensation method, which could provide reference for optimizing observation task of X-ray pulsars and recovering X-ray pulse template.

Keywords: X-ray pulsar-based navigation; pulse template; pulsar angular position error; anniversary characteristic; compensation method

Received: 2016-12-21; **Accepted**: 2017-01-20; **Published online**: 2017-03-08 09:29

URL: kns.cnki.net/kcms/detail/11.2625.V.20170308.0929.003.html

Foundation item: National Natural Science Foundation of China (10973048)

* **Corresponding author**. E-mail: zhengwei@nudt.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2016.0965

无人机室内视觉/惯导组合导航方法

王亭亭[1], 蔡志浩[1,2,*], 王英勋[1,2]

(1. 北京航空航天大学 自动化科学与电气工程学院, 北京 100083;
2. 北京航空航天大学 飞行器控制一体化技术国防科技重点实验室, 北京 100083)

摘 要: 针对室内无卫星定位下的无人机自主导航问题,提出了一种融合惯导、光流和视觉里程计的组合导航方法。在速度估计上,采用基于 ORB 特征的光流法,该方法可以实时地估计出无人机的三轴线速度信息。方法采用基于特征点的稀疏光流,对金字塔 Lucas-Kanade 光流法进行了改进,采用前后双向追踪和随机采样一致的方法提高特征点追踪精度。在位置估计上,采用视觉/惯导融合的视觉里程计,以人工图标法为主,融合视觉光流信息和惯导数据实现无人机定位。通过与运动捕捉系统的定位信息、Guidance 和 PX4Flow 导航模块的测速信息进行对比,以及实际的飞行测试,验证本文方法的可行性。

关 键 词: 无人机; 视觉导航; 光流; ORB 特征; 多传感器融合
中图分类号: TP242.6; V249.32
文献标识码: A **文章编号**: 1001-5965(2018)01-0176-11

在传统的导航方法中,卫星导航是应用最广泛的导航方法。但是无人机在城市或者室内飞行时,由于楼宇、森林和墙壁等阻碍,卫星信号将不可用,加之普通的惯性测量元件存在精度低和严重漂移等问题,传统的导航方法将不能满足无人机的导航需求,因此基于视觉感知系统的无人机导航方法被广泛研究和应用。这些技术主要包括单目/双目视觉里程计(Visual Odometry, VO)[1-4]、单目/双目视觉同时定位与建图(Visual Simultaneous Localization and Mapping, V-SLAM)[5-6]、图像光流[7-10]等,且多采用组合导航方法。其中,文献[1]利用视觉、激光、惯导和 GPS 的组合导航实现多旋翼无人机室内/外自主飞行;文献[3-6]利用视觉/惯导组合导航实现无人机车/无人机位姿估计。部分消费级无人机,如大疆创新科技有限公司的精灵 4 利用 2 个双目视觉传感器和 1 个单目视觉传感器实现了自主避障、室内飞行和行人跟踪[9];零零无限科技有限公司的 Hover Camera 利用一个下视和前视视觉传感器实现室内指尖放飞、定点和跟拍飞行[10]。无人机是 1 个在三维空间运动,且只有 6 自由度的平台,对算法的实时性和鲁棒性要求较高,而一般的视觉导航方法普遍存在更新速率低和延迟大的缺点。

因此,本文提出了一种融合视觉传感器和惯性传感器的组合导航方法,采用基于 ORB(Oriented FAST and Rotated BRIEF)特征的光流法估计无人机速度信息,采用图标法里程计估计位置信息,并应用扩展卡尔曼滤波(Extended Kalman Filter, EKF)将惯导的角速度和加速度信息与视觉光流和里程计进行融合,利用单目/双目视觉系统和惯性传感器搭建导航平台,应用于小型无人机的室内导航。

收稿日期: 2016-12-23; **录用日期**: 2017-02-06; **网络出版时间**: 2017-03-23 16:44
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170323.1644.003.html
基金项目: 航空科学基金 (20135851043)
* **通信作者**. E-mail: czh@buaa.edu.cn

引用格式: *王亭亭, 蔡志浩, 王英勋. 无人机室内视觉/惯导组合导航方法[J]. 北京航空航天大学学报, 2018, 44(1): 176-186.*
WANG T T, CAI Z H, WANG Y X. Integrated vision/inertial navigation method of UAVs in indoor environment[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(1): 176-186 (in Chinese).

1　基于 ORB 特征的光流法的速度估计

1.1　ORB 特征提取与描述

ORB[11]是一种实时性较高的特征提取与描述方法，它融合了 FAST[12]和 BRIEF（Binary Robust Independent Elementary Features）[13]2 种描述子。FAST 具有很强的实时性，但其不具备尺度不变性，且不包含方向因子。BRIEF 用简单的二进制形式描述特征，实时性比较高，但其无法保证旋转和尺度不变性，对噪声比较敏感。ORB 基于 FAST 和 BRIEF 方法，具体实现如下。

1）建立图像高斯金字塔，对每一层图像提取 FAST 特征点。

2）为特征点添加方向因子。以特征点为中心建立一个图像块，定义图像块各像素点的矩为

$$m_{\alpha\beta}=\sum_{x,y}x^{\alpha}y^{\beta}I(x,y) \tag{1}$$

式中：α、β 为矩的阶次；$I(x,y)$ 为图像灰度表达式。利用零阶矩和一阶矩来计算质心的坐标，即 $C_0=\left(\frac{m_{10}}{m_{00}},\frac{m_{01}}{m_{00}}\right)$。建立一个从图像块的中心到质心的向量作为方向因子，向量的方向可简化为 $\theta=\arctan(m_{01},m_{10})$，这里忽略此特征是亮点还是暗点。

3）引入旋转因子。利用 BRIEF 方法描述特征点，并在此基础上采用方向因子驱动法实现二进制字符串向量的旋转。

在 BRIEF 方法中，对于一个 $N\times N$ 大小的图像块 P，有如下定义：

$$\tau(I;x,y):=\begin{cases}1 & I(x)<I(y)\\ 0 & \text{其他}\end{cases}$$

选取 n 个特征点得到的 n 维向量如下：

$$f_n(p):=\sum_{1\leqslant i\leqslant n}2^{i-1}\tau(I;x_i,y_i) \tag{2}$$

定义一个 $2\times n$ 的矩阵：$\boldsymbol{S}=\begin{bmatrix}x_1 & x_2 & \cdots & x_n\\ y_1 & y_2 & \cdots & y_n\end{bmatrix}$，利用方向因子 θ 和基础旋转矩阵 $\boldsymbol{R}_\lambda$，定义 $\boldsymbol{S}_\lambda=\boldsymbol{R}_\lambda\boldsymbol{S}$。得到最终的特征点描述子如式（3）所示。只要 λ 确定了，便可快速获得描述子。

$$g_n(p,\lambda):=f_n(p)\,|\,(x_i,y_i)\in\boldsymbol{S}_\lambda \tag{3}$$

本文对 ORB、SIFT[14]和 SURF[15]的特征提取时间开销进行了试验分析，如表 1 所示。

在对不同场景进行特征点提取时，ORB 比 SIFT 和 SURF 的耗时小，相差一个数量级。这是因为：①FAST角点提取时间复杂度低。②SIFT 特征为 128 维描述子，占据 512 Bytes 空间；SURF 特征为 64 维描述子，占据 256 Bytes 空间。而用 BRIEF 描述每个特征点仅需一个长度为 256 的向量（二进制形式向量），存储空间为 32 Bytes，降低了后续特征匹配的时间复杂度。因此将 ORB 特征提取方法用于稀疏光流场特征提取，可以减小时间开销。

表 1　SIFT、SURF 和 ORB 特征提取时间开销对比

Table 1　Comparison of time consumption of feature extraction among SIFT, SURF and ORB

方　法	特征点数	时间/ms
SIFT[14]	171	18.89
	253	18.99
	234	19.66
SURF[15]	86	12.44
	254	16.21
	187	13.86
ORB	168	2.82
	299	4.11
	251	4.7

1.2　改进的 Lucas-Kanade 光流提取

基于特征点的 Lucas-Kanade 算法（简称 LK 算法）是经典的光流提取算法，该算法基于以下 3 种假设：目标图像亮度一致、图像空间连续和图像变化在时间上连续。$I(u,v,t)$ 表示 t 时刻像素点处图像灰度值，基于以上假设可以得到像素点的亮度守恒方程为

$$\boldsymbol{I}_u\dot{u}+\boldsymbol{I}_v\dot{v}+\boldsymbol{I}_t=\boldsymbol{0} \tag{4}$$

式中：$\dot{u}$ 和 $\dot{v}$ 分别为图像横纵向光流值；$\boldsymbol{I}_u$ 和 $\boldsymbol{I}_v$ 分别为像素点亮度在 x 和 y 方向的偏导数；$\boldsymbol{I}_t$ 为像素点亮度值随时间的导数。这里待求的 $\dot{u}$ 和 $\dot{v}$ 将直接参与速度解算。

LK 算法的 3 个假设在实际场景中很难满足，传统的解决办法包括：①保证足够大的图像帧速率；②引入图像金子塔，通过对不同尺度的图像的光流提取保证连续性。为了提高前后帧特征点追踪准确性，本文在传统方法的基础上采用前后双向追踪策略和随机采样一致滤波（Random Sampling Consensus，RANSAC）提高光流场的提取精度。

1）前后双向追踪策略

如图 1 所示，针对前后 2 帧图像，首先提取第 1 帧图像的特征点，得到特征点集 A，然后通过金字塔 LK 算法在第 2 帧图像中得到对应特征点集 A 的特征点集 B，此为前向追踪。之后，针对特征点集 B，返回在第 1 帧追踪得到对应的特征点集，记为 C，此为后向追踪。

由于噪声影响，前后追踪的特征点并非完

北航学报 赠阅

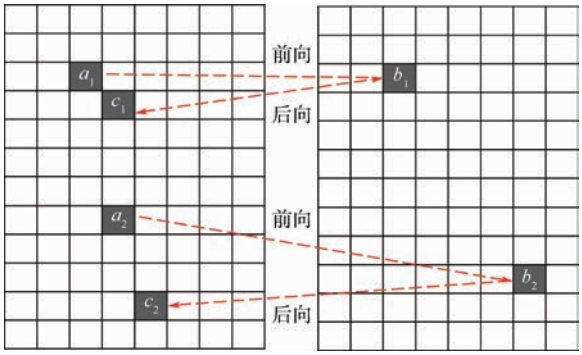


图 1 前后双向追踪示意图

Fig. 1 Schematic of forward-backward tracking

全一致。对此,选取误差和相似度 2 个标准对匹配点对进行滤波。其中,误差指 A、C 中对应点的距离,距离越小,误差越小。相似度指对 A、C 中对应的每一对点,分别选取 10×10 大小的邻域,对 2 个邻域进行模板匹配,选择归一化相关系数匹配法,将匹配的结果作为相似度。匹配值越大,表示越相似。最后选取相似度大且误差小的点对。

2)随机采样一致滤波

本文借助随机采样一致的思想,在前后双向追踪得到的特征点对中随机选取 8 对点对,算出基础矩阵,然后利用该基础矩阵对应的极性约束,对剩余的匹配进行测试。RANSAC 算法认为,支持集合越大,得到的矩阵正确的可能性越大。最终选取较大的支持集合的特征点对应的光流值来估计速度。

RANSAC[16]算法的实现如下:

步骤 1 初始:假设 D 是 N 个特征对应的集合。

步骤 2 开启循环:

1)随机选取集合 D 中的 S 个采样点;

2)为该 S 个采样点拟合一个模型;

3)计算 D 集合中的其他点与该模型的距离;

4)所有距离小于阈值的点构成一个内联集;

5)存储内联集;

6)返回 1)继续,直到迭代到最大迭代次数。

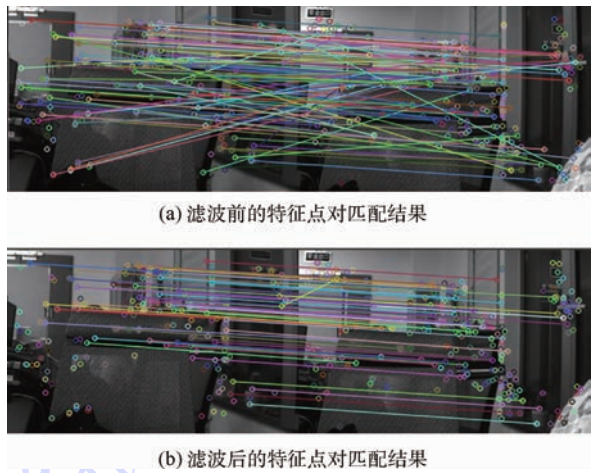
步骤 3 选择拥有最大内联集合的采样点 S^*,并用所有的内联集估计该模型。

图 2 显示了 RANSAC 算法前后的匹配效果。经过 RANSAC 算法,错误的匹配被有效滤除。

1.3 融合惯导角速度的速度解算

当相机没有水平移动而仅存在姿态变化时,依然会产生光流场,因此在实际由光流场解算速度时,需要考虑姿态变化的影响。

以下视安装的单目相机为例,如图 3 所示。



(a) 滤波前的特征点对匹配结果

(b) 滤波后的特征点对匹配结果

图 2 采用 RANSAC 算法对特征点对滤波

Fig. 2 Filtering features by RANSAC algorithm

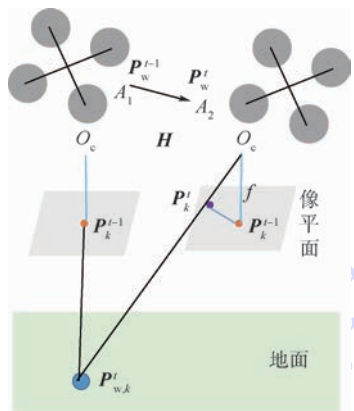


图 3 相机运动时空间点成像变化

Fig. 3 Point-mapping changing caused by moving camera

$t-1$ 到 t 时刻,相机从 A_1 点(坐标为 $\boldsymbol{P}_w^{t-1}$)运动到 A_2 点(坐标为 $\boldsymbol{P}_w^t$)。假设相机安装在飞机重心,像平面与机体 xy 平面平行,则相机到机体的旋转矩阵 $\boldsymbol{C}_c^k$ 可以求出,相机之间的移动可用单应矩阵 $\boldsymbol{H}$ 表示。令 t 时刻在图像中提取到 n 个特征点 $p_k^t, k=1,2,\cdots,n$,特征点在相机坐标系下的投影为 $\boldsymbol{P}_{c,k}^t$,对应于地面的点为 $\boldsymbol{P}_{w,k}^t$。已知相机内参矩阵为 $\boldsymbol{M}$,相机焦距为 f,f 和 $\boldsymbol{M}$ 可由标定相机得到,特征点在相机坐标系 O_c 轴上尺度为 Z_c。

鉴于图像更新频率很快(大于 80 Hz),认为连续两帧之间相机与地面的距离 h 未有明显的变化。则在 t 时刻有

$$\boldsymbol{P}_{c,k}^t = Z_c \boldsymbol{M}^{-1} p_k^t \tag{5}$$

$$\boldsymbol{P}_{w,k}^t = \boldsymbol{H} \boldsymbol{P}_{c,k}^t \tag{6}$$

$t-1$ 到 t 时刻 n 个特征点的稀疏光流为

$$\mathrm{OF}_k^t = (p_k^t - p_k^{t-1})/\Delta t \tag{7}$$

单位时间相机在空间的移动(v_{bx},v_{by})远远大于投影到像平面的像素的移动(Δu,Δv),相机距地面的高度 h 远远大于镜头的焦距 f。则光流与

北航学报 赠阅

无人机线运动、角运动的关系可以描述为

$$\mathrm{OF}_k^t = (V_\mathrm{b} f \sin\eta)/h - w \tag{8}$$

则有

$$\begin{cases} v_{\mathrm{b}x} = \dfrac{\Delta u}{f\Delta t}h - qh \\ v_{\mathrm{b}y} = \dfrac{\Delta v}{f\Delta t}h - \gamma h \end{cases} \tag{9}$$

式中：V_b 为无人机线速度；η 为无人机 x 轴与无人机-目标之间连线的夹角；w 为无人机角速率；$v_{\mathrm{b}x}$、$v_{\mathrm{b}y}$ 分别为 x、y 方向的线速度；γ、q 分别为滚转和俯仰角速率，可通过精度较高的 IMU 测量得到。高度数据 h 可用超声传感器测量，对于前视双目相机的光流解算则需要获取每一对特征的深度 d，用 d 代替 h。如式(10)所示，(u_1, v_1) 和 (u_2, v_2) 分别为左相机和右相机中的一对匹配点的图像层的坐标，以像素为单位。$\boldsymbol{M}_{\mathrm{iL}}$ 和 $\boldsymbol{M}_{\mathrm{iR}}$ 分别为左右相机的内参数，通过相机标定获得。(X, Y, Z) 为该对匹配点对应的外部空间(即世界坐标系)点的实际坐标。在此，令世界坐标系与左相机的相机坐标系重合(见式(11))，求解出图像点对应的场景中的点的坐标(X, Y, Z)，即场景特征到相机的相对位置。工程应用中采用最小二乘法求解，以降低因数据噪声带来的解算误差。最后，采用卡尔曼滤波处理速度信息。

$$\begin{cases} Z_{c1}\begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^1 & m_{12}^1 & m_{13}^1 & m_{14}^1 \\ m_{21}^1 & m_{22}^1 & m_{23}^1 & m_{24}^1 \\ m_{31}^1 & m_{32}^1 & m_{33}^1 & m_{34}^1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \boldsymbol{M}_{\mathrm{iL}}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \\ Z_{c2}\begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^2 & m_{12}^2 & m_{13}^2 & m_{14}^2 \\ m_{21}^2 & m_{22}^2 & m_{23}^2 & m_{24}^2 \\ m_{31}^2 & m_{32}^2 & m_{33}^2 & m_{34}^2 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \boldsymbol{M}_{\mathrm{iR}}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \end{cases} \tag{10}$$

$$\begin{cases} \boldsymbol{M}_{\mathrm{L}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \\ \boldsymbol{M}_{\mathrm{R}} = [\boldsymbol{R}_{\mathrm{r,l}} \quad \boldsymbol{T}_{\mathrm{r,l}}] \end{cases} \tag{11}$$

手持相机做平稳往复运动，图 4 显示了本文方法与 Guidance 和 PX4Flow[17] 等的测速性能对比。大的姿态角变化会引起较大的速度测量误差，使速度曲线产生大的“尖峰”，从图中可以看出(箭头指示)，因为引入了惯导角速率，本文方法能较好地降低姿态角变化带来的干扰。

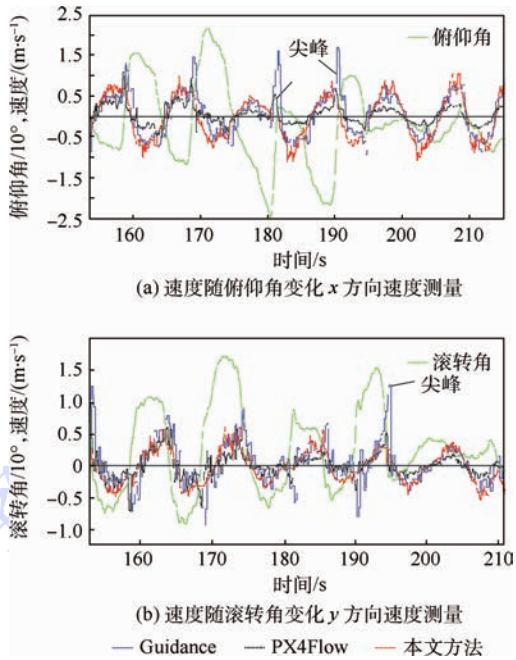


图 4　姿态变化时测速对比

Fig. 4　Velocity comparison with changing attitude

2　图标法里程计

针对室内地板格环境，本文采用图标法里程计，以地板格直线的交点为图标，通过追踪图标进行定位。算法基本流程如图 5 所示。

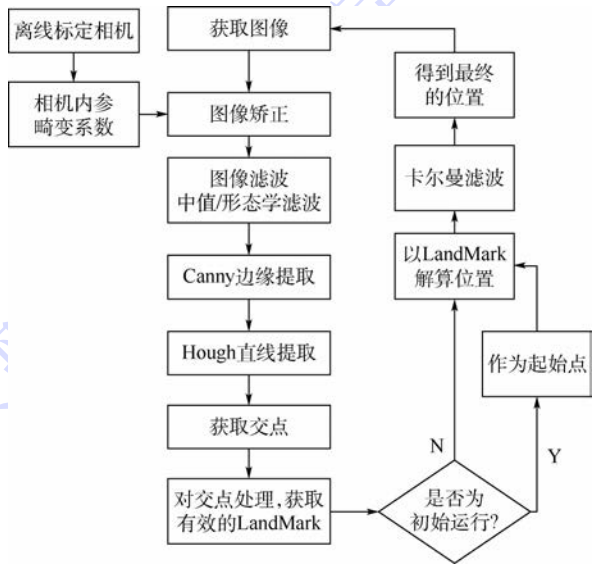


图 5　图标法里程计算法基本流程

Fig. 5　Algorithm flowchart of icon odometry

首先对摄像头获取的地板灰度图像进行滤波，利用 Canny 算子进行边缘提取(见图 6)。然后对获取的边缘图像利用 Hough 变换提取线段。根据以下原则对线段滤波：①合并“同线”的线段；②保留“水平”和“垂直”的线段；③滤除“较短”的线段。最后，求取线段所在直线的交点。

北航学报 赠阅

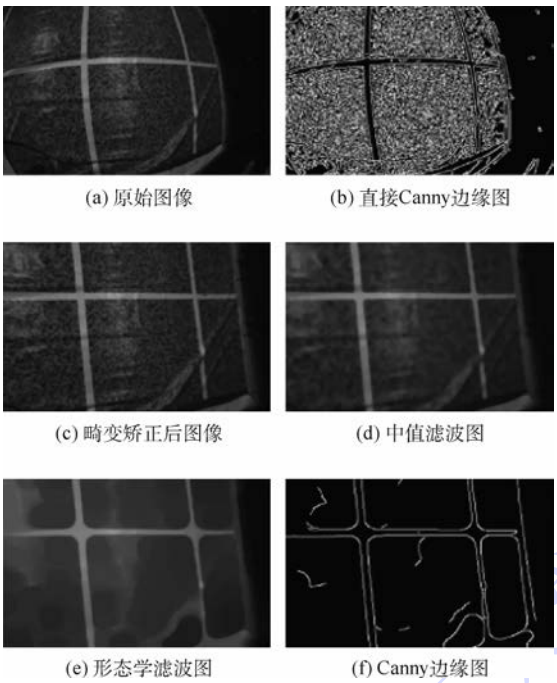


图 6 图像滤波与边缘提取

Fig. 6 Image filtering and edge extraction

地板直线的交点表现在图像中是一个图像块,取图像块的几何中心的像素值(u^*,v^*)用于二维图像平面到三维空间的映射[18]。

$$z_c\begin{bmatrix}u^*\\v^*\\1\end{bmatrix}=[\boldsymbol{A}\quad\boldsymbol{0}]\begin{bmatrix}\boldsymbol{R}&\boldsymbol{t}\\\boldsymbol{0}&1\end{bmatrix}\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix}\tag{12}$$

式中:$\boldsymbol{A}$ 为相机的内参矩阵;$\boldsymbol{R}$ 和 $\boldsymbol{t}$ 分别为相机坐标系与导航坐标系的旋转矩阵和平移向量。鉴于飞行器飞行高度远远大于相机的焦距 $H\gg f$,因此可以令 $Z\approx H$,这样 Z 可以通过超声传感器获得。

相机与飞机固连且垂直向下安装。如图 7 所示,假设飞机在 1 个直线交点的附近起飞,其相机定位到的第 1 个交点作为原点。则定义交点 0 为原点,即导航坐标系的(0,0,0),飞机此时的位置为(0,0,H)。假设飞机沿着红色箭头的方向直线飞行到交点 3 的上空。在这个过程中,椭圆曲线范围内的点在不断变化,利用椭圆内点的定位平均值作为当前飞机的位置估计。之所以定义一个椭圆的范围,主要是因为考虑图像边缘畸变大,定位误差大,该椭圆的中心与图像的中心重合。而利用椭圆包围的点进行定位,误差小,只有当椭圆内搜索不到点时,才考虑椭圆外的点。已知地板格为边长为 1 m 的正方形,因此在飞机从 0 点到 3 点运动过程中,y 方向的值从 0 变为 1。

利用图标法里程计,只要保证视野里始终有交点,便可以实现对飞机的有效定位。

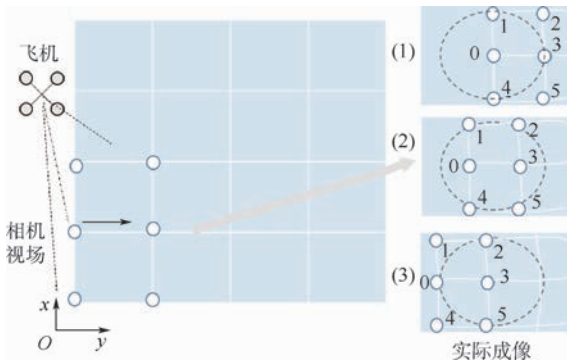


图 7 图标法里程计定位

Fig. 7 Positioning based on icon odometry

3 基于 EKF 的数据融合

第 1 节和第 2 节所述的光流法和图标法里程计都依赖于视觉传感器和外界特征(特征点和交点)。但在实际飞行中,由于种种原因,会出现视觉传感器失效的状况。为此,考虑视觉和惯导融合进行速度和位置的估计。

3.1 全状态下的视觉/惯导融合

假设图标法里程计和图像光流场比较稳定,将里程计、光流和惯导做数据融合,本文将这种情况定为全状态。

假设在极小的 Δt 时间内系统为非线性的匀变速过程。定义系统的状态量为导航坐标系下的飞机的三维全局位置,三维速度和三维加速度和飞机的俯仰角、滚转角、偏航角:

$$\boldsymbol{X}_k=[x,y,z,v_x,v_y,v_z,a_x,a_y,a_z,\theta,\phi,\psi]^{\mathrm{T}}\in\mathbf{R}^{12}$$

系统的观测量为视觉里程计测量的 x,y,z 方向的位置,光流法获得的速度,以及 MTI 的三轴线性加速度信息和姿态角:

$$\boldsymbol{Y}_k=[x_{\mathrm{vo}},y_{\mathrm{vo}},z_{\mathrm{vo}},v_{\mathrm{oF},x},v_{\mathrm{oF},y},v_{\mathrm{oF},z},a_{\mathrm{gyro},x},a_{\mathrm{gyro},y},a_{\mathrm{gyro},z},\theta_{\mathrm{gyro}},\phi_{\mathrm{gyro}},\psi_{\mathrm{gyro}}]^{\mathrm{T}}\in\mathbf{R}^{12}$$

定义过程模型为

$$\begin{cases}x_k=x_{k-1}+tv_{x,k-1}+\omega_{1,k-1}\\y_k=y_{k-1}+tv_{y,k-1}+\omega_{1,k-1}\\z_k=z_{k-1}+tv_{z,k-1}+\omega_{1,k-1}\\v_{x,k}=v_{x,k-1}+a_{x,k-1}t+\omega_{2,k-1}\\v_{y,k}=v_{y,k-1}+a_{y,k-1}t+\omega_{2,k-1}\\v_{z,k}=v_{z,k-1}-a_{z,k-1}t+\omega_{2,k-1}\\a_{x,k}=a_{x,k-1}+\omega_{3,k-1}\\a_{y,k}=a_{y,k-1}+\omega_{3,k-1}\\a_{z,k}=a_{z,k-1}+\omega_{3,k-1}\\\theta_k=\theta_{k-1}+\omega_{4,k-1}\\\phi_k=\phi_{k-1}+\omega_{4,k-1}\\\psi_k=\psi_{k-1}+\omega_{4,k-1}\end{cases}\tag{13}$$

系统的观测模型为

$$\begin{cases} x_{\mathrm{vo},k}=x_k+\delta_{1,k} \\ y_{\mathrm{vo},k}=y_k+\delta_{1,k} \\ z_{\mathrm{vo},k}=z_k+\delta_{1,k} \\ v_{\mathrm{oF}x,k}=v_{x,k}+\delta_{2,k} \\ v_{\mathrm{oF}y,k}=v_{y,k}+\delta_{2,k} \\ v_{\mathrm{oF}z,k}=v_{z,k}+\delta_{2,k} \\ \begin{bmatrix} a_{\mathrm{gyrox},k} \\ a_{\mathrm{gyroy},k} \\ a_{\mathrm{gyroz},k} \end{bmatrix}=\boldsymbol{C}_{\mathrm{n}}^{\mathrm{b}}\begin{bmatrix} a_{x,k} \\ a_{y,k} \\ a_{z,k} \end{bmatrix}-g\boldsymbol{e}+\boldsymbol{C}_{\mathrm{n}}^{\mathrm{b}}\delta_{3,k} \\ \theta_{\mathrm{gyro},k}=\theta_k+\delta_{4,k} \\ \phi_{\mathrm{gyro},k}=\phi_k+\delta_{4,k} \\ \psi_{\mathrm{gyro},k}=\psi_k+\delta_{4,k} \end{cases} \tag{14}$$

式中：ω 为观测噪声；δ 为量测噪声；$\boldsymbol{C}_{\mathrm{n}}^{\mathrm{b}}$ 为导航-机体坐标系转换矩阵；g 为重力加速度；$\boldsymbol{e}$ 为单位向量。观测方程为非线性模型，根据 EKF 公式将 f_k 在最优估计点处展开，将观测方程线性化。

3.2　非全状态下的视觉/惯导融合

1）图标法里程计失效的情况

当地面没有可用的边线用于图标法里程计时，里程计的定位将失效。具体表现为视野内没有可用的交点。此时，置 VO_KEY 为 False，仅利用光流和惯导组合，估计定位信息。此时观测量为

$$\boldsymbol{Y}_k=[v_{\mathrm{oF},x},v_{\mathrm{oF},y},v_{\mathrm{oF},z},a_{\mathrm{gyro},x},a_{\mathrm{gyro},y},a_{\mathrm{gyro},z},\theta_{\mathrm{gyro}},\phi_{\mathrm{gyro}},\psi_{\mathrm{gyro}}]^{\mathrm{T}}\in\mathbf{R}^9$$

状态方程参考式(11)，观测模型为

$$\begin{cases} v_{\mathrm{oF}x,k}=v_{x,k}+\delta_{2,k} \\ v_{\mathrm{oF}y,k}=v_{y,k}+\delta_{2,k} \\ v_{\mathrm{oF}z,k}=v_{z,k}+\delta_{2,k} \\ \begin{bmatrix} a_{\mathrm{gyrox},k} \\ a_{\mathrm{gyroy},k} \\ a_{\mathrm{gyroz},k} \end{bmatrix}=\boldsymbol{C}_{\mathrm{n}}^{\mathrm{b}}\begin{bmatrix} a_{x,k} \\ a_{y,k} \\ a_{z,k} \end{bmatrix}-g\boldsymbol{e}+\boldsymbol{C}_{\mathrm{n}}^{\mathrm{b}}\delta_{3,k} \\ \theta_{\mathrm{gyro},k}=\theta_k+\delta_{4,k} \\ \phi_{\mathrm{gyro},k}=\phi_k+\delta_{4,k} \\ \psi_{\mathrm{gyro},k}=\psi_k+\delta_{4,k} \end{cases} \tag{15}$$

2）视觉传感器失效的情况

如果视觉传感器失效，则利用惯导积分做短时的导航，直到光流速度或者里程计的测量恢复。这里判断光流速度失效的原则为连续图像光流点数量小于 4 个，或连续出现野值。

3.3　视觉、惯导的多速率问题

惯性传感器数据更新速率快，视觉传感器的数据更新速率慢，因此其中涉及了多速率融合的问题。对此的解决方法为：保证状态和噪声的定义与之前的定义相同，修改传感器的观测。系统的基本采样周期为 T_0，惯导和视觉传感器的采样周期分别为 T_1 和 T_2。这里保证 T_1 和 T_2 为 T_0 的整数倍，即 $T_1=n_1T_0$，$T_2=n_2T_0$。令 κT_0 为各个观测数据的采样周期的最小公倍数，则定义观测矩阵和观测噪声协方差如下：

$$\boldsymbol{H}_k=\begin{cases} H_{ik} & \mathrm{mod}(k,n_i)=0\ \&\ \mathrm{mod}(k,\kappa)\neq 0 \\ \begin{bmatrix} H_{1k} \\ H_{2k} \end{bmatrix} & \mathrm{mod}(k,\kappa)=0 \\ \boldsymbol{0} & \text{其他} \end{cases} \tag{16}$$

$$\boldsymbol{R}_k=\begin{cases} R_{ik} & \mathrm{mod}(k,n_i)=0\ \&\ \mathrm{mod}(k,\kappa)\neq 0 \\ \mathrm{diag}(R_{1k},R_{2k}) & \mathrm{mod}(k,\kappa)=0 \\ \boldsymbol{I} & \text{其他} \end{cases} \tag{17}$$

式中：mod(·)为求余函数。

已知试验所用的视觉传感器的数据更新速率约为 50 Hz，惯导的更新速率为 100 Hz，令 T_0 = 10 ms，则 $T_1=20$ ms，$T_2=10$ ms，$n_1=2$，$n_2=1$。在 ROS 操作系统中，多线程采集到的数据可以通过时间戳进行数据对准，并可以根据时间戳来判断是否采样得新数据，这样根据式(16)和式(17)便可以方便地编程实现多速率数据的融合。

图 8 显示了“地板格”从有(视觉里程计可用、下视光流可用)到无(纯白色地面、视觉里程计失效、下视相机失效)时的位置估计，包括三维轨迹及高度、速度和位置随时间的变化。手持相机运动到约 2 m 的位置时，开始移动出地板(朝着位置增加的方向)。在 3 次往返后，终点和起点近似重合，依然可以实现有效定位。

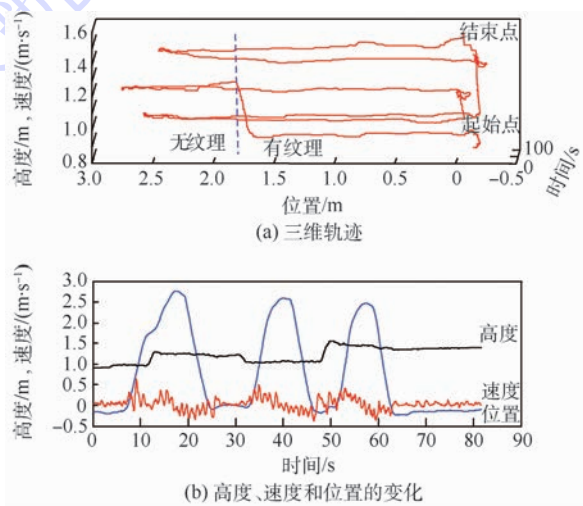


图 8　定位测试

Fig. 8　Localization test

北航学报 赠阅

4 试验与分析

鉴于运动捕捉系统 OptiTrack 的数据延迟小(低于 5 ms),定位误差小(约 0.1 mm),可以 OptiTrack的定位数据作为基准,衡量本文方法的精度和延迟。本文采用美国 OptiTrack Prime 41 相机搭建运动捕捉系统进行试验。

图 9 ~ 图 11 分别为将本文方法与运动捕捉系统获得的位置和速度信息进行对比结果,可得本文方法延迟约 50 ms,位置误差约 0.15 m,速度误差约 0.05 m/s,基本满足室内环境下小型无人机的飞行需求。

本文搭建四轴八桨多旋翼无人机平台(见图 12),以手控的方式操纵飞机飞行(见图 13),实时记录导航数据,并将本文方法的测速数据与 Guidance 和 PX4Flow 导航模块的测速数据进行了对比。

图 14 显示了不同速度下本文方法与 Guidance 导航模块测速情况对比。无人机的飞行速度基本在 0 ~ 2.5 m/s 的范围。在此速度范围内本文方法可以有效实现对无人机的导航。当速度大于 2 m/s 时,有效特征点数下降,提到的光流场数量降低,当速度大于 2.5 m/s 时,有效光流场数量频繁出现小于 8 的状态,单纯的视觉信息将不再可靠。

图 15 和图 16 显示了某次飞行测试的导航数据。其中导航速度信息与 PX4Flow 的测速信息进行了对比。从图中的数据可以看出,本文方法在实际的无人机平台上比较稳定,可以尝试用于自动(闭环)飞行。

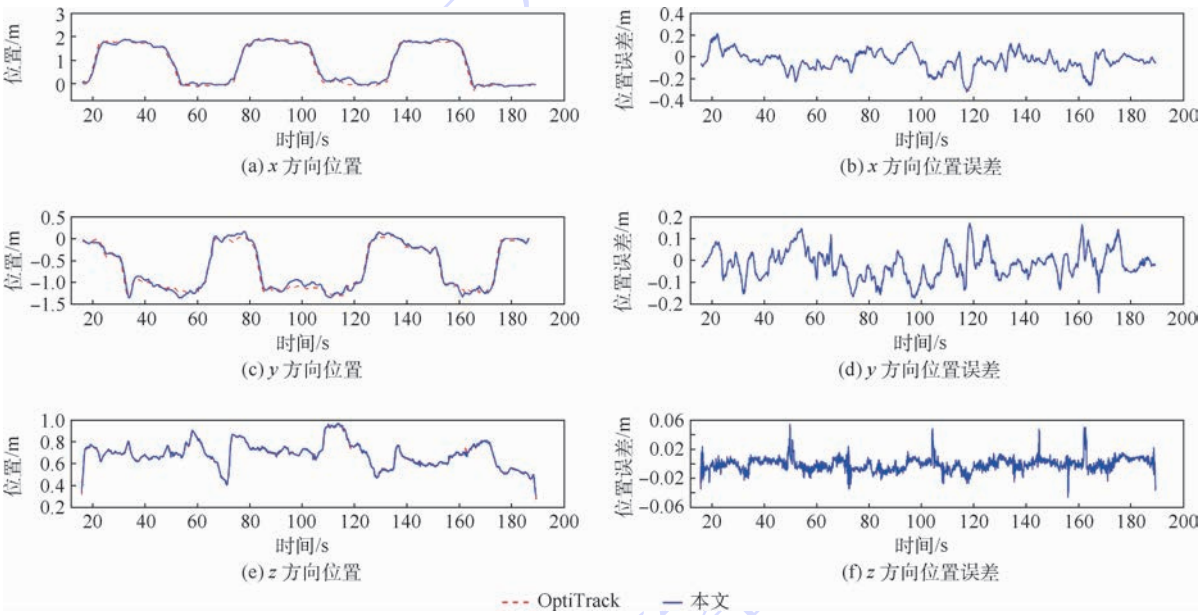


图 9 三维位置估计与位置误差

Fig. 9 Localization estimation and positioning error in 3D-directions

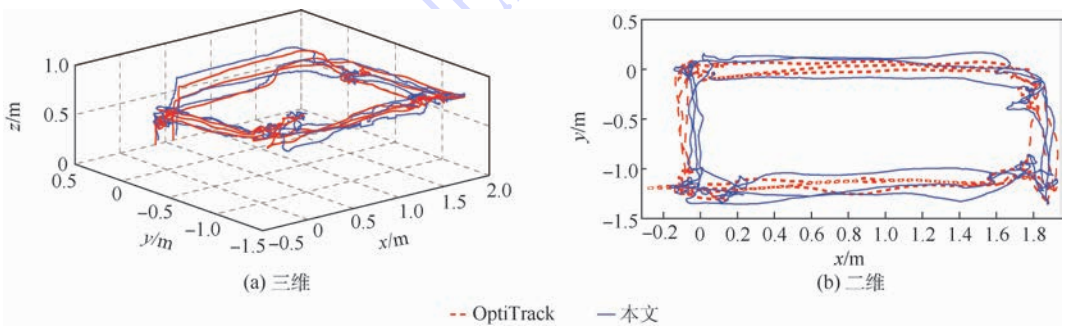


图 10 三维/二维运动轨迹

Fig. 10 Three/two-dimensional motion trajectory

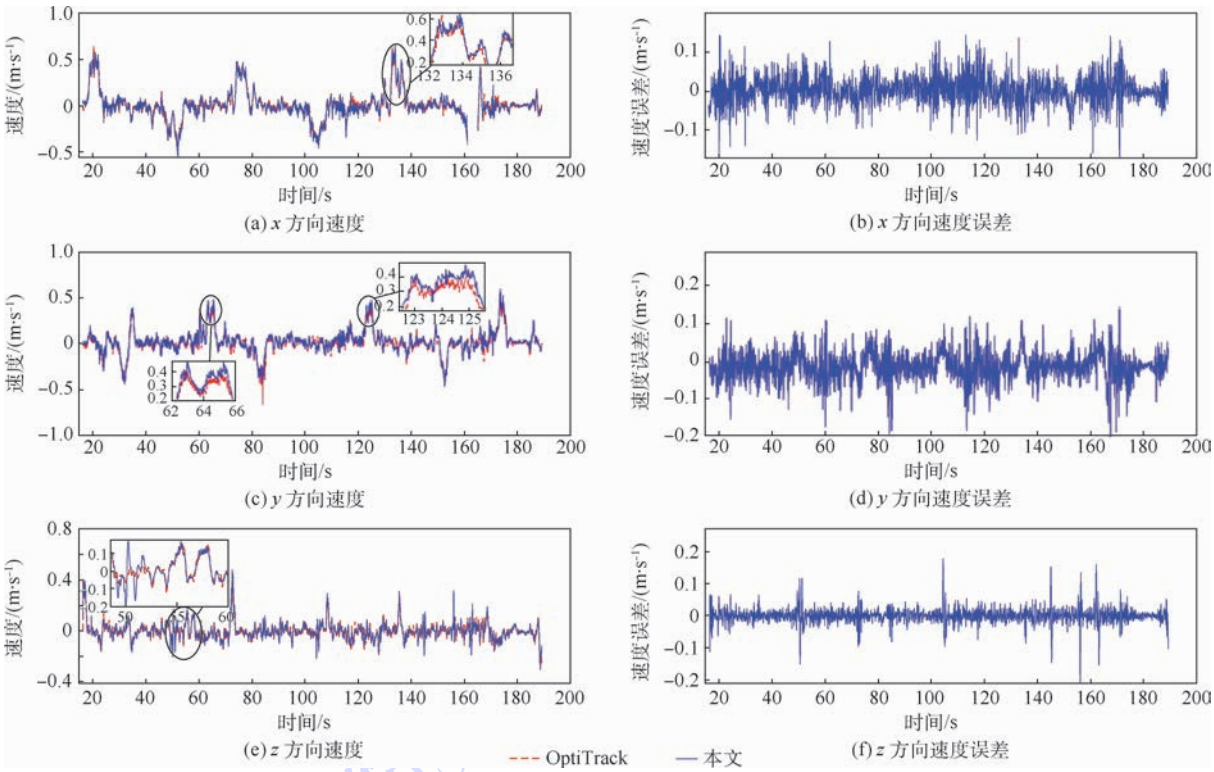


图 11　三维速度估计与速度误差

Fig. 11　Velocity estimation and velocity error in 3D-directions

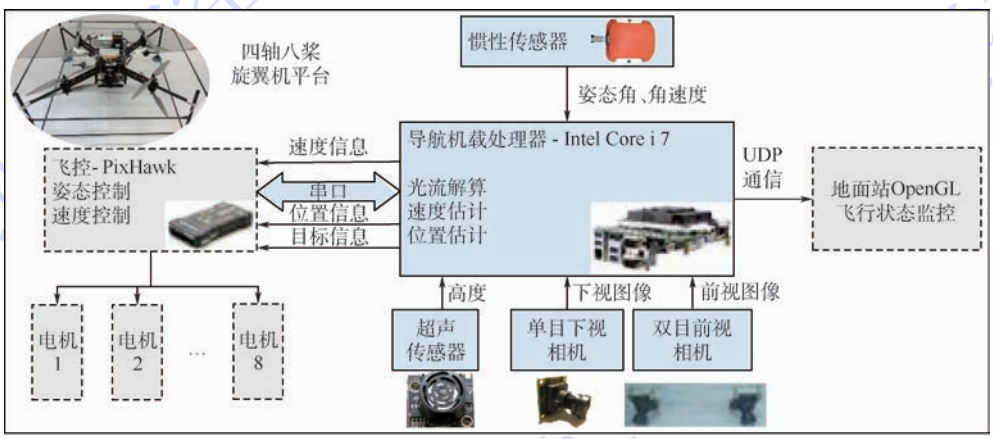


图 12　无人机平台及系统结构

Fig. 12　UAV platform and system structure



图 13　飞行视频截图

Fig. 13　Screenshots of flight video

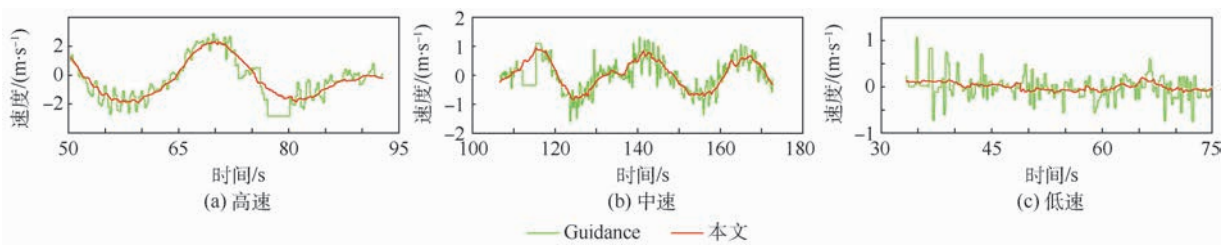


图 14　手控飞行速度测试

Fig. 14　Flight velocity test via manual control

北航学报 赠阅

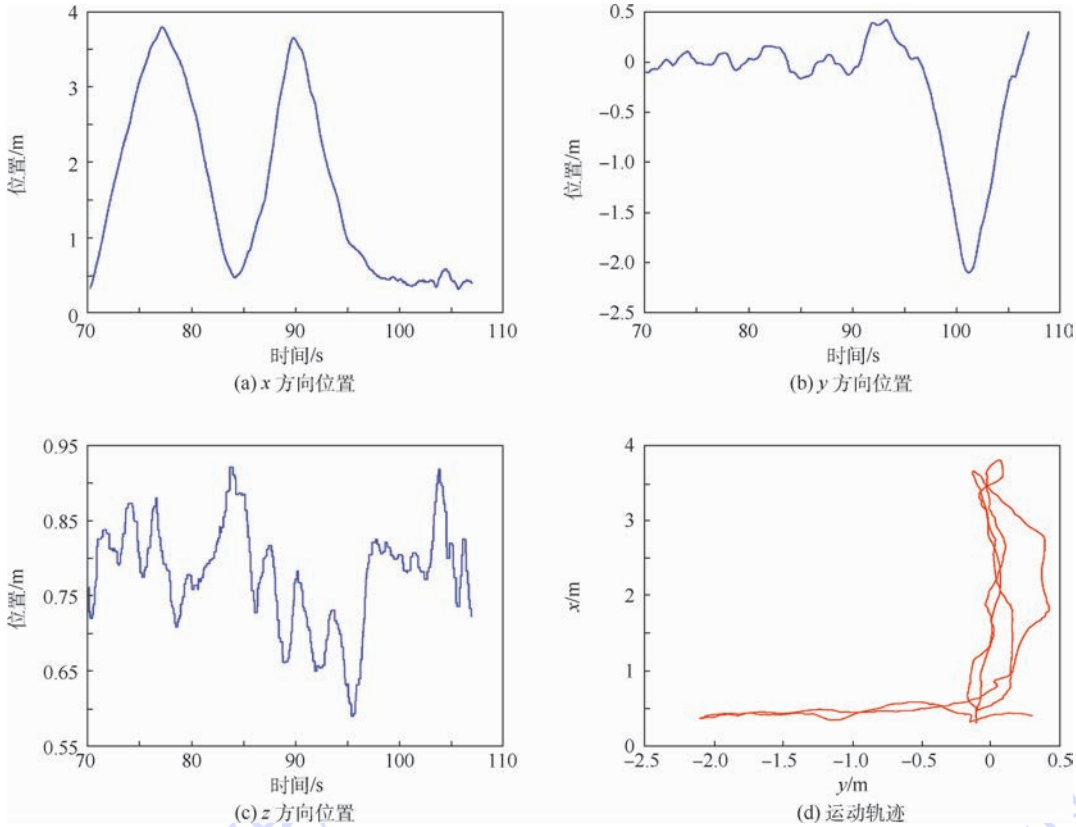


图 15 位置估计

Fig. 15 Position estimation

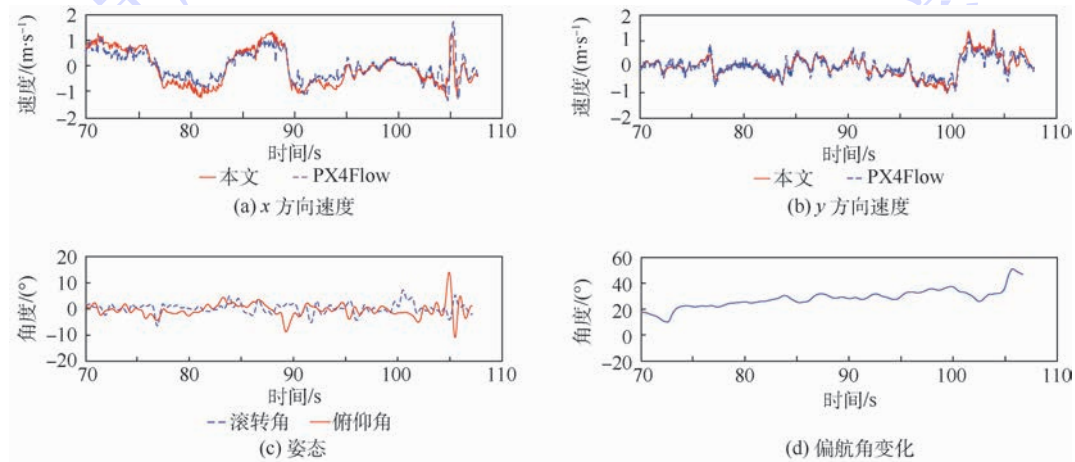


图 16 速度估计与飞机姿态

Fig. 16 Velocity estimation and aircraft attitude

5 结 论

本文提出了一种视觉和惯导融合用于估计小型无人机速度和位置的方法,可得到以下结论:

1) 基于 ORB 特征点的前后向追踪的光流提取算法,提高了光流场的提取精度,降低了算法的时间开销。

2) 基于图标法里程计的快速定位算法,可以有效应对带地板格的室内环境定位。该算法简单,实时性好。

3) 利用 EKF 将图标法里程计、光流速度和惯导信息进行融合,降低了方法对环境的依赖。

4) 以延迟小于 5 ms,精度小于 0.01 mm 的运动捕捉系统为参考基准,与本文提出的运动估计方法进行对比验证,得到本文方法延迟约 50 ms,定位误差约 0.15 m,速度误差约 0.05 m/s。同时搭建了多旋翼平台进行飞行验证。

参考文献（References）

[1] SHEN S J. Autonomous navigation in complex indoor and outdoor environments with micro aerial vehicles[D]. Philadelphia: University of Pennsylvania, 2014.

[2] 吴琦，蔡志浩，王英勋. 用于无人机室内导航的光流与地标融合方法[J]. 控制理论与应用，2015，32(11)：1511-1517.
WU Q, CAI Z H, WANG Y X. Optical flow and landmark fusion method for UAV indoor navization[J]. Control Theory & Applications, 2015, 32(11): 1511-1517(in Chinese).

[3] LI P, LAMBERT A. A monocular odometer for a quadrotor using a homogra-phy model and inertial cues[C]// IEEE Conference on Robotics and Biomimetics. Piscataway, NJ: IEEE Press, 2015: 570-575.

[4] 叶长春. IARC 第 7 代任务中定位与目标跟踪方法研究[D]. 杭州：浙江大学，2016.
YE C C. Research on localization and object tracking for the IARC mission7[D]. Hangzhou: Zhejiang University, 2016(in Chinese).

[5] MUR-ARTAL R, MONTIEL J M M, TARDOS J D. ORB-SLAM: A versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.

[6] LEUTENEGGER S, FURGALE P, RABAUD V, et al. Keyframe-based visual-inertial SLAM using nonlinear optimization[C]// Robotics: Science and Systems, 2013: 789-795.

[7] CHAO H, GU Y, GROSS J, et al. A comparative study of optical flow and traditional sensors in UAV navigation[C]// American Control Conference(ACC). Piscataway, NJ: IEEE Press, 2013: 3858-3863.

[8] MAMMARELLA M, CAMPA G, FRAVOLINI M L, et al. Comparing optical flow algorithms using 6-dof motion of real-world rigid objects[J]. IEEE Transactions on, Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2012, 42(6): 1752-1762.

[9] DJI Innovations. PHANTOM 4 user's manual V1.2[EB/OL]. (2016-12-23) https://dl.djicdn.com/downloads/phantom_4/cn/Phantom_4_User_Manual_cn_v1.2_160328.pdf.

[10] Hover Camera 2016[EB/OL]. (2016-12-23) http://gethover.com.

[11] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF[J]. Proceedings, 2011, 58(11): 2564-2571.

[12] ROSTEN E, DRUMMOND T. Machine learning for high-speed corner detection[C]// European Conference on Computer Vision. Berlin: Springer-Verlag, 2006: 430-443.

[13] CALONDER M, LEPETIT V, STRECHA C, et al. BRIEF: Binary robust independent elementary features[C]// European Conference on Computer Vision. Berlin: Springer-Verlag, 2010: 778-792.

[14] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[15] BAY H, TUYTELAARS T, GOOL L V. SURF: Speeded up robust features[J]. Computer Vision & Image Understanding, 2006, 110(3): 404-417.

[16] CHUM O, MATAS J, KITTLER J. Locally optimized RANSAC[J]. Lecture Notes in Computer Science, 2003, 2781: 236-243.

[17] HONEGGER D, MEIER L, TANSKANEN P, et al. An open source and open hardware embedded metric optical flow cmos camera for indoor and outdoor applications[C]// International Conference on Robotics and Automation. Piscataway, NJ: IEEE Press, 2013: 1736-1741.

[18] HARTLEY R I, ZISSERMAN A. Multi-view geometry in computer vision[M]. Cambrige: Cambridge University Press, 2004: 239-247.

作者简介：

王亭亭　女，硕士研究生。主要研究方向：机器视觉、无人机视觉导航。

蔡志浩　男，副教授，硕士生导师。主要研究方向：无人机自主控制与导航、多机协同与训练。

王英勋　男，研究员，博士生导师。主要研究方向：系统设计、自主控制与模拟训练。

北京航空航天大学学报

北航学报 赠阅

Integrated vision/inertial navigation method of UAVs in indoor environment

WANG Tingting[1], CAI Zhihao[1,2,*], WANG Yingxun[1,2]

(1. School of Automation Science and Electrical Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China;
2. Aircraft Control Integration National Defense Key Laboratory, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: A new integrated navigation method based on inertial sensor, optical flow and visual odometry is proposed for self-navigation indoor in GPS-denied environment. An ORB optical flow based method is also proposed for estimating real-time three-axis velocity of the UAV. The algorithm improves the traditional pyramid Lucas-Kanade method using sparse optical flow based on feature points. The tracking of feature points is made more accurate by applying forward-backward tracking and random sampling consensus strategies. For position estimation, a visual odometry method with integrated vision/inertial navigation is adopted, which uses the artificial icon method, visual optical flow information and inertial navigation data. Finally, the velocity and position estimations from the proposed method are validated via actual flight test and via comparison with velocity measurement information from a PX4Flow module and a Guidance module and with locating information from movement capture system.

Keywords: UAV; vision navigation; optical flow; ORB features; multi-sensor fusion

北京航空航天大学学报

Received: 2016-12-23; **Accepted**: 2017-02-06; **Published online**: 2017-03-23 16:44

URL: kns.cnki.net/kcms/detail/11.2625.V.20170323.1644.003.html

Foundation item: Aeronautical Science Foundation of China (20135851043)

* **Corresponding author**. E-mail: czh@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2016.0985

孔边倒角裂纹当量初始缺陷尺寸的确定方法

孙晓娜[1], 贺小帆[1,*], 李玉海[2]

(1. 北京航空航天大学 航空科学与工程学院, 北京 100083; 2. 中国航空工业集团有限公司, 北京 100022)

摘 要: 孔边倒角裂纹是含孔下陷细节的常见裂纹形态,为进行含孔下陷细节的经济寿命评定,需要确定倒角裂纹的原始疲劳质量(IFQ)。首先,为探究倒角对裂纹前缘应力强度因子的影响,进行了有、无倒角2种情况下应力强度因子的有限元分析。计算表明倒角对相对小裂纹的应力强度因子影响显著。其次,为合理表征该种裂纹的IFQ,将初始缺陷当量为萌生于倒角和试件表面交点,前缘为圆弧的初始裂纹,以萌生点到裂纹前缘沿孔径向的距离作为裂纹特征尺寸。最后,采用相对小裂纹扩展方程描述倒角裂纹的扩展规律,反推得到倒角裂纹的当量初始缺陷尺寸(EIFS)分布。统计分析表明,采用本文定义的裂纹特征尺寸得到的EIFS分布参数与应力水平无关。

关 键 词: 原始疲劳质量(IFQ); 应力强度因子; 有限元; 裂纹扩展方程; 当量初始缺陷尺寸(EIFS)

中图分类号: V215.5
文献标识码: A **文章编号**: 1001-5965(2018)01-0187-09

概率断裂力学方法(PFMA)是结构耐久性分析的重要方法,采用该方法进行耐久性分析的关键是确定结构的原始疲劳质量(IFQ),通常用当量初始缺陷尺寸(EIFS)描述[1]。EIFS是假想的初始裂纹尺寸,用来表征材料加工以及结构制造引入的初始缺陷对疲劳寿命的影响。Rudd和Gray[2-3]首次引入EIFS的概念,Yang、Manning和Molent等对此进行了进一步研究[4-8],其成为最常用的表征结构(IFQ)的方法。并发展了基于EIFS的损伤容限分析方法[9-10]和基于EIFS的考虑腐蚀环境影响的结构寿命预测方法[11]。

EIFS一般通过断口数据反推得到,它不仅与材料特性、加工工艺质量有关,还与细节形式和断口裂纹形态有关。紧固孔是飞机结构中数量最多的细节,其疲劳强度直接影响着飞机的服役寿命。针对紧固孔细节,Fawaz[12]研究了飞机蒙皮搭接和对接2种情况下紧固孔的EIFS分布,Wang[13]探究了多种因素对紧固孔EIFS的影响,证明EIFS是初始表面质量的函数。文献[14]给出各种类型紧固孔的耐久性试验数据并确定了EIFS分布。文献[11-13]均把紧固孔细节的当量初始裂纹形态假设为孔边角裂纹或孔壁半椭圆形裂纹,定义孔壁到裂纹前缘的径向最远距离为裂纹尺寸。若裂纹在孔壁或孔角起裂,采用这种方法是合理的。但是在实际结构中,为了提高细节的抗疲劳性能,方便部件装配,孔边往往要进行倒角处理。文献[14]试验表明,此类细节在疲劳载荷作用下会出现倒角裂纹,即裂纹在倒角与试件表面交点处萌生的情况。结构形式及尺寸的变化(如梁缘条及蒙皮厚度渐变),会形成下陷含孔细节,在拉、弯复合载荷作用下,更容易在倒角处起裂。在这种情况下,若假定初始裂纹为孔壁裂纹或孔角裂纹,以孔壁到裂纹前缘的最远距离作为裂纹长度,会造成如下问题:

收稿日期: 2017-01-03; **录用日期**: 2017-03-10; **网络出版时间**: 2017-04-13 15:05
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170413.1505.002.html
* **通信作者**. E-mail: xfhe@buaa.edu.cn

引用格式: 孙晓娜, 贺小帆, 李玉海. 孔边倒角裂纹当量初始缺陷尺寸的确定方法[J]. 北京航空航天大学学报, 2018, 44(1): 187-195. SUN X N, HE X F, LI Y H. Determination method of equivalent initial flaw size for crack initiated at hole chamfering[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(1): 187-195 (in Chinese).

北航学报 赠阅

1）由于裂纹在倒角和试件表面交点处萌生，沿试件表面及倒角表面扩展。若以孔壁为裂纹起点，裂纹长度必然大于倒角尺寸，造成萌生寿命过短，与实际情况不符，过于保守。

2）对于含孔细节，常用的修理方式是扩孔。如果按裂纹前缘到孔壁的距离定义裂纹尺寸，扩孔量会偏大，这会造成修理后剩余寿命降低。

3）采用反推法确定 EIFS 时，需要建立裂纹长度随寿命的变化规律，如果裂纹尺寸定义不合理，裂纹扩展规律可能难以描述，从而造成 EIFS 估计及经济寿命预测结果不准确。

Molent 和 Yang 等指出 EIFS 与裂纹萌生位置有关[8,15]。然而对于萌生于孔边倒角的裂纹，尚未建立合理的 EIFS 确定方法。为此，本文进行孔边倒角的含孔下陷细节试件耐久性试验，研究裂纹萌生和扩展过程，基于断口分析结果，计算分析倒角对裂纹前缘应力强度因子的影响，确定裂纹特征尺寸，并建立孔边倒角裂纹 EIFS 确定方法。

1 耐久性试验

1.1 试 件

试件材料为国产 7B04-T74 铝合金板材，材料力学性能如表 1 所示，取样方向为 L-S 向。试件的 2 个孔位于下陷区，模拟某型飞机某关键结构的细节形式。试件的几何形状如图 1 所示，孔边倒角尺寸为 0.3 mm。

表 1 7B04-T74 铝合金力学性能

Table 1 Mechanical properties of 7B04-T74 aluminum alloy

参 数	弹性模量 E/GPa	泊松比 ν	屈服强度 σ_s/MPa	抗拉强度 σ_b/MPa
数 值	70	0.33	450	518

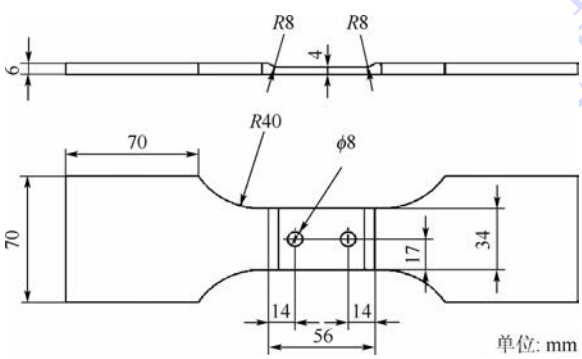


图 1 试件几何形状

Fig. 1 Geometry of specimen

1.2 载 荷 谱

基于某型飞机实测过载谱，通过适当调整，形成了可以在断口上留下标识的耐久性载荷谱，一个基本周期代表 1 057.998 fh（fh 为飞行小时）。经试验验证，调整后谱的损伤与原谱损伤相当。采用耐久性谱，进行了 2 种应力水平下的耐久性试验，单位过载的净截面应力水平分别取为 33.5 MPa（应力水平 1）、31.5 MPa（应力水平 2）。

1.3 试验结果

试验在 Instron8801-100 kN 电液伺服疲劳试验机上进行。试件直接夹持在试验机夹头上，在室温大气环境下沿轴向加载，试验频率为 6 Hz。2 种应力水平下有效试件数分别为 7 件和 6 件。

试件典型断口如图 2 所示，断口上留下了较清晰的标识线。

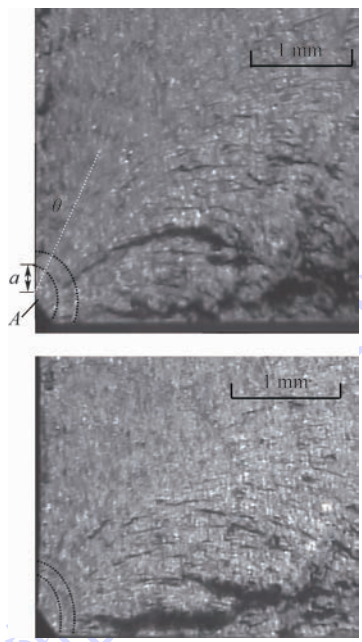


图 2 试件典型断口

Fig. 2 Typical fracture surfaces of specimen

1.4 裂纹形态分析

如图 2 所示，裂纹均在倒角处萌生，裂纹萌生位置为倒角和试件表面的交点，用 A 点表示。在裂纹相对较小时，裂纹前缘近似为以 A 为圆心的圆弧形，沿表面和厚度方向的尺寸基本相同。随着裂纹扩展，裂纹形态逐渐发展为 1/4 椭圆裂纹，沿孔径方向的尺寸明显大于沿试件厚度方向的尺寸。

2 裂纹特征尺寸的定义

为计算孔边倒角裂纹形态的 EIFS，首先应该定义合理的裂纹特征尺寸。裂纹特征尺寸应反映裂纹几何，通常以裂纹前缘到基准线（点）的最长距离作为裂纹特征尺寸。在确定裂纹特征尺寸时，不仅要考虑裂纹萌生位置和前缘形状，还要考

北航学报 赠阅

虑拟采取的维修方式。

孔角或孔壁萌生的裂纹常表现为孔壁半椭圆形裂纹或孔角椭圆裂纹,通常采用扩孔方式进行修理,因此以孔壁作为基准线,以孔壁到裂纹前缘的最远距离作为裂纹特征尺寸。但是对于倒角处萌生的裂纹,如引言所述,若采用孔壁作为基准线会造成耐久性分析结果偏离实际情况。因此,本节基于 1.4 节的裂纹形态分析结果,从裂纹前缘的应力强度因子分布及修理时能否去除裂纹前缘塑性区等角度探讨倒角处萌生裂纹的特征尺寸定义方法。

2.1　倒角对裂纹前缘应力强度因子的影响

紧固孔的 EIFS 常规计算方法通常忽略倒角的影响,将其假设为孔角椭圆裂纹,该裂纹前缘到孔壁和试件表面的距离与倒角裂纹相同。本节基于断口判读得到的裂纹形态和几何尺寸,将图 1 所示试件作为分析对象,采用有限元方法计算倒角裂纹与假想的孔角椭圆裂纹 2 种裂纹形态的裂纹前缘应力强度因子分布,对比分析倒角对裂纹前缘应力强度因子 K_{I} 的影响。

裂纹形态如图 3 所示,针对存在倒角的情况,模拟倒角圆弧裂纹。对无倒角情况,模拟孔角椭圆裂纹,椭圆长轴为孔壁到倒角裂纹前缘的最大长度,即倒角裂纹径向尺寸与倒角尺寸之和,短轴为倒角裂纹沿厚度方向的尺寸。

采用 ABAQUS 有限元软件提供的相互作用积分法进行应力强度因子的求解,由于试件的对称性,采用1/2有限元模型,如图4所示。在试验条件下,试验机夹头相对试件刚度无穷大,试件只能沿轴向移动。因此在试件夹持端模拟夹持边界条件,将试件端部刚化,只保留试件沿加载方向的自由度[16]。

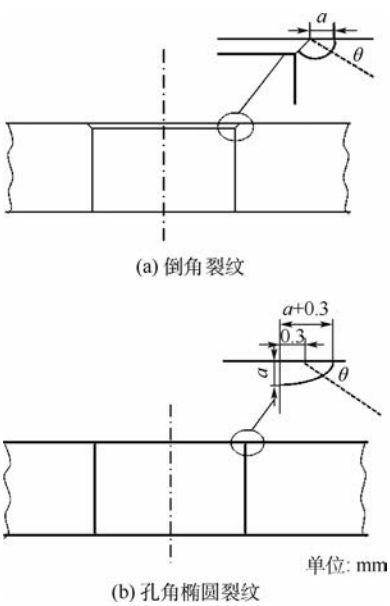


图 3　有无倒角下的 2 种裂纹形态

Fig. 3　Shape of two types of cracks initiated at holes with and without chamfering

整体的网格划分如图 5 所示。裂纹前缘区域采用二次楔形单元(C3D15),如图 6 区域 1 所示。裂纹前缘区域以外的其他区域采用二次六面体单元(C3D20),如图 6 区域 2 所示。本文中奇异单元长度为 0.05 mm,角度为 18°,积分路径数目为 5,计算表明,当积分路径数目大于 2(即积分区域直径大于 0.08 mm)时,沿不同积分路径的应力强度因子的计算结果相对误差小于 1%。

远端拉伸应力取 100 MPa,不同裂纹尺寸 a 下,倒角裂纹和孔角椭圆裂纹(见图 3)的应力强度因子 K_{I} 如图 7 所示。为方便比较,对于倒角裂纹和孔角椭圆裂纹,选取相同坐标系,坐标原点为倒角与试件表面的交点 A(见图 2),横坐标为裂纹前缘和 A 点的连线与试件上表面的夹角 θ,如图 3 所示。考虑到 2 种裂纹形态下裂纹长度的定义不同,故未对应力强度因子进行无量纲化,纵坐标直接取应力强度因子值。对比结果如图 7 所示,可以看出:

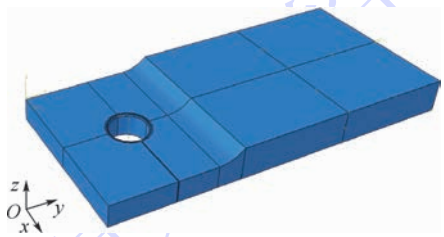


图 4　几何模型

Fig. 4　Geometry model

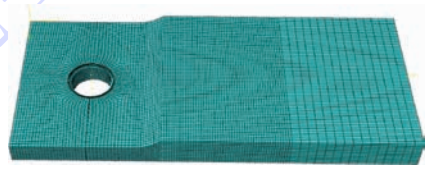


图 5　总体网格

Fig. 5　Mesh of whole model

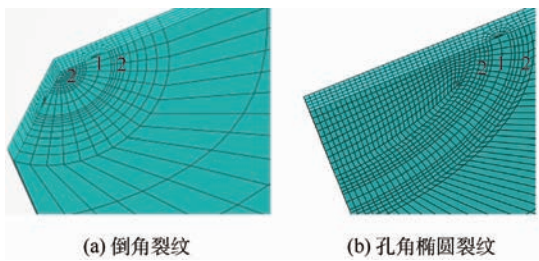


图 6　裂纹前缘附近网格

Fig. 6　Mesh near crack tip

北航学报 赠阅

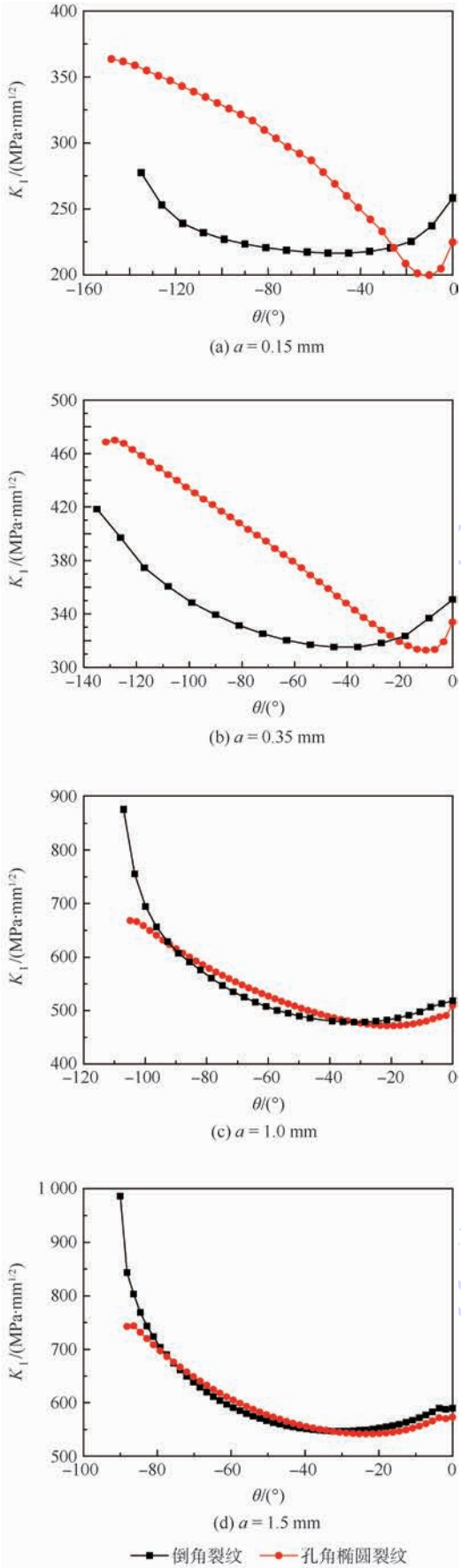


图 7 2 种裂纹形态下应力强度因子的对比

Fig. 7 Comparison of stress intensity factors of two types of cracks

1）当裂纹尺寸较短（a < 1.5 mm）时，2 种裂纹形态的应力强度因子分布规律和取值均存在较大的差别。在孔壁附近，孔角椭圆裂纹的应力强度因子值明显高于倒角裂纹的应力强度因子值，最大相差 31%；在试件表面附近，倒角裂纹的应力强度因子值略大于孔角椭圆裂纹的应力强度因子值，最大相差 13.2%。显然不能将倒角裂纹等效为孔角椭圆裂纹。

2）当裂纹较长时（a = 1.5 mm），2 种裂纹形态下应力强度因子的分布比较一致，可以将倒角裂纹等效为孔角椭圆裂纹。

2.2 裂纹前缘塑性区尺寸

实际材料不是理想线弹性的，裂纹前缘存在塑性区，在进行结构修理时，必须去除裂纹及裂纹前缘塑性区。从 2.1 节的应力强度因子计算结果可以看出，若将倒角裂纹等效为孔角椭圆裂纹，在最为关注的试件表面方向应力强度因子计算值小于实际裂纹应力强度因子计算值。由此估计得到的裂纹前缘前缘塑性区尺寸必然小于实际塑性区尺寸，造成偏危险的结果。因此进行塑性区尺寸估计时，应依据实际裂纹形态。

对于倒角裂纹，修理方式有 2 种。第 1 种是直接通过扩孔将裂纹及前缘塑性区尺寸去除，然后倒角，如图 8 所示（点划线为塑性区范围，虚线为修理方法示意图），此时需要关注尺寸 r_1。这种修理方式下扩孔量较大，导致剩余强度及剩余寿命低，甚至存在难以实施的可能性。第 2 种修理方式是先扩孔后倒角，通过扩孔和倒角的综合处理将裂纹及裂纹前缘塑性区去除，如图 9 所示。此时，关注的尺寸为r_2。显然此种修理方式

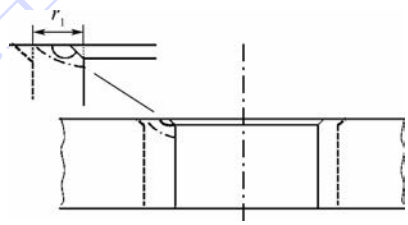


图 8 修理方式 1

Fig. 8 Repair scheme 1

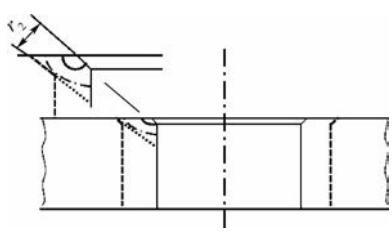


图 9 修理方式 2

Fig. 9 Repair scheme 2

北航学报 赠阅

所需要的扩孔量小于图 8 中的做法。若采用此种修理方式,则需计算裂纹前缘塑性区尺寸和范围。塑性区尺寸和范围一般依据裂纹的一个或多个特征尺寸给出,裂纹特征尺寸的选取应参考裂纹前缘塑性区形状。下面进行倒角裂纹的塑性区形状估计。

对于Ⅰ型裂纹问题,根据裂纹尖端塑性区的二级估计[17],塑性区尺寸可通过式(1)计算:

$$\begin{cases} R_{0,\mathrm{stress}} = \dfrac{K_{\mathrm{I}}^2}{\pi\sigma_{\mathrm{s}}^2} & \text{平面应力} \\ R_{0,\mathrm{strain}} = (1-2\nu)^2\dfrac{K_{\mathrm{I}}^2}{\pi\sigma_{\mathrm{s}}^2} & \text{平面应变} \end{cases} \tag{1}$$

式中:σ_{s} 为屈服强度,本次试件材料的实测屈服强度为 450 MPa。

对于三维裂纹,裂纹前缘不同位置的应力状态不同,试件表面($\theta = 0°$)、倒角表面($\theta = -135°$)以及孔壁均可近似为平面应力状态,除此之外的区域处于复杂应力状态。参考文献[17],裂纹前缘塑性区尺寸可按式(2)计算:

$$R_0 = \begin{cases} R_{0,\mathrm{stress}} - (R_{0,\mathrm{strain}} - R_{0,\mathrm{stress}})\sin\left(\dfrac{4\theta}{3}\right) & R_0 + a \leqslant 0.3\sqrt{2} \\ R_{0,\mathrm{stress}} - (R_{0,\mathrm{strain}} - R_{0,\mathrm{stress}}) \cdot \sin\left(\dfrac{\pi}{\pi - \arccos(0.3/a)}\theta\right) & R_0 + a > 0.3\sqrt{2} \end{cases} \tag{2}$$

式中:$R_{0,\mathrm{stress}}$ 和 $R_{0,\mathrm{strain}}$ 由式(1)计算。

在孔壁附近,由于几何形状复杂,通过该种方法计算的塑性区尺寸有一定误差,但在试件表面附近较为准确。进行倒角时,主要关注试件表面附近的塑性区形态,因此使用该计算方法是可行的。以远端拉应力 $\sigma = 100$ MPa 为例,在极坐标下绘制裂纹前缘塑性区形状如图 10 所示。裂纹前缘的塑性区形状如图 10 所示。计算表明,塑性区前缘任一点的切线斜率小于等于 1,因此修理时只需在图中点划线位置所示进行倒角,即可完全去除塑性区。此时需要关注的是试件表面的塑性区尺寸。

2.3　裂纹特征尺寸确定

对于倒角裂纹,采用图 9 所示的经济修理方式时,需关注试件表面的塑性区尺寸。同时在进行断口判读时,沿试件表面方向(孔径向)的裂纹尺寸更容易读取。综上所述,在进行 EIFS 分析时,应将初始裂纹假设为倒角裂纹,以萌生点到裂纹前缘的径向距离为裂纹特征尺寸,如图 3(a)所示。

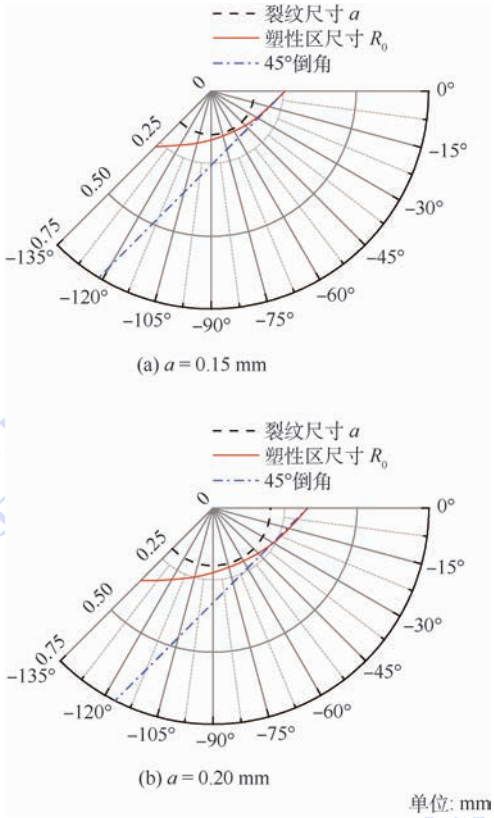


图 10　塑性区形状

Fig. 10　Plastic zone shape

3　EIFS 分析

在倒角裂纹尺寸定义下,计算得到 EIFS 是否能真实反映结构的 IFQ 有待进一步验证。本节在倒角裂纹尺寸定义基础上,计算得到不同应力水平下的 EIFS 分布,并进行统计对比,检验其是否与应力水平无关,从而间接说明本文定义方式的合理性。并在此基础上,计算得到含孔下陷细节的 EIFS 分布。

根据文献[18],EIFS 分布的确定一般有 2 种方法,一种是 TTCI 反推法,首先计算每个试件达到指定裂纹尺寸的时间(Time to Crack Initiation, TTCI),假设 TTCI 服从某一特定分布,通过 TTCI 分布反推得到 EIFS 分布。另一种是 EIFS 拟合法,即直接反推得到 EIFS 样本,选定合适的分布类型,通过统计处理得到分布参数。本文采用后者。

3.1　EIFS 控制曲线的建立

首先,通过对试件进行断口判读得到(a, t)数据,绘制 a-t 曲线,如图 11 所示。可以看出 a-t 曲线在对数坐标下呈分段线性,裂纹较短和较长时 a-t 曲线的斜率不同。考虑到耐久性分析的范围,取 1.2 mm 以下[18] 的 a-t 数据进行分析,如

北航学报 赠阅

图 12所示,在对数坐标系中相对小裂纹的 a-t 曲线近似为一条直线。对于相对小裂纹,通常采用式(3)描述其裂纹扩展规律:

$$\mathrm{d}a/\mathrm{d}t = Qa^b \tag{3}$$

式中:Q 为裂纹扩展速度参数;b 为裂纹扩展方程指数。

式(3)是基于描述裂纹扩展过程的 Paris 公式形式,将三维裂纹应力强度因子计算公式作级数展开,取低阶项得到的近似表达式[19]。当裂纹长度较小时,该式具有很好的适用性。由于该细节应力集中区的应力状态复杂,该式的适用性需要通过建立试件表面裂纹前缘应力强度因子与 a 的变化关系曲线来验证。因此,基于试验断口判读的裂纹形态和尺寸,进行了相对小裂纹情况下裂纹特征尺寸对应的试件表面裂纹前缘应力强度因子随特征尺寸变化的有限元模拟。对应力强度因子进行无量纲化,令 $\beta = K_{\mathrm{I}}/(\sigma\sqrt{\pi a})$,得到无量纲化应力强度因子随特征尺寸变化关系,如图 13所示。用幂函数拟合结果为 $\beta = 2.92a^{-0.13}$,相关系数为 0.993,线性相关度较好。因此用式(3)描述裂纹扩展规律是合理的。

对式(3)两边取对数,得到 $\ln(\mathrm{d}a/\mathrm{d}t) = \ln Q + b\ln a$。根据文献[18]中的方法,计算得到 2 种应力水平下,b 分别为 1.066 和 0.949,因此可近似取 $b = 1$。

根据文献[18],每个试件的裂纹扩展速度参数 Q_{ik} 由该试件的断口数据决定。采用最小二乘法,由 a-t 数据,直接拟合得到 Q 值,第 i 种应力水平,第 k 个试件的裂纹扩展方程参数 Q_{ik} 由该试件的 m_{ik} 个断口数据 (a_{ikj}, t_{ikj}) $(j = 1, 2, \cdots, m_{ik})$ 数据通过式(4)计算:

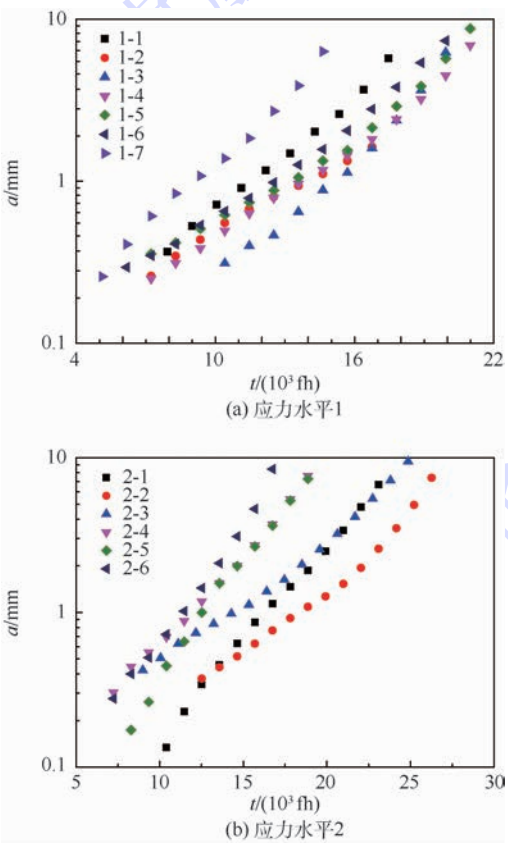


图 11 裂纹扩展数据

Fig. 11 Crack growth data

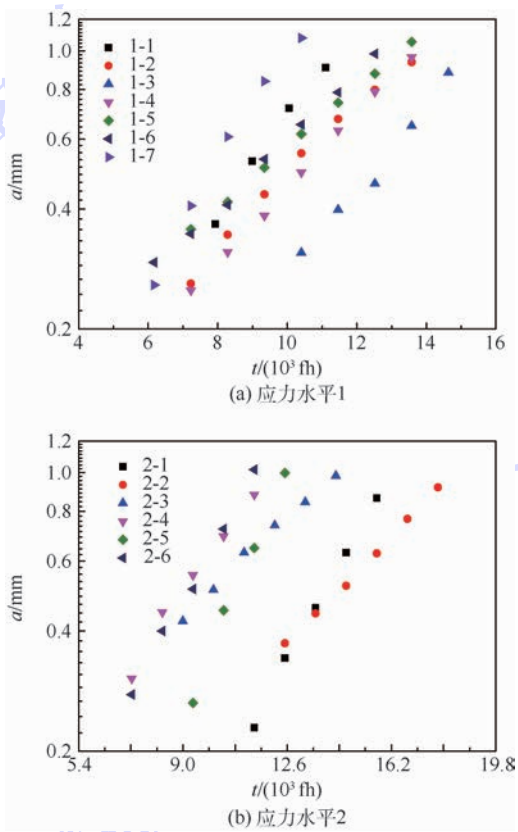


图 12 相对小裂纹 a-t 数据

Fig. 12 a-t data for relatively small crack

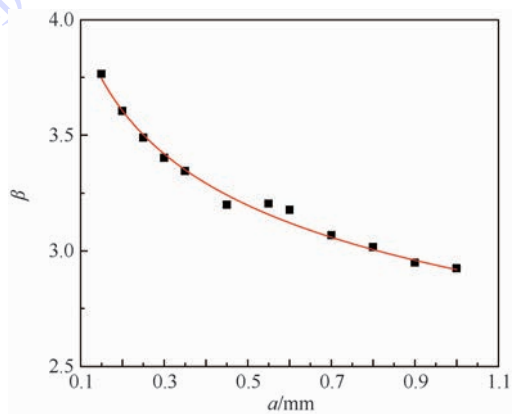


图 13 无量纲化应力强度因子随裂纹尺寸变化曲线

Fig. 13 Changing curve of non-dimensional stress intensity factor with crack size

$$Q_{ik}=\frac{m_{ik}\sum_{j=1}^{m_{ik}}t_{ikj}\ln a_{ikj}-\sum_{j=1}^{m_{ik}}t_{ikj}\sum_{j=1}^{m_{ik}}\ln a_{ikj}}{m_{ik}\sum_{j=1}^{m_{ik}}t_{ikj}^2-\left(\sum_{j=1}^{m_{ik}}t_{ikj}\right)^2} \tag{4}$$

指定参考裂纹尺寸 a_r 下 EIFS 控制方程如式(5)所示。取 $a_r=0.5$ mm 计算得到试件的 EIFS 如图 14 所示。

$$x_{ik}=a_r\exp(-Q_{ik}t_{ik}) \tag{5}$$

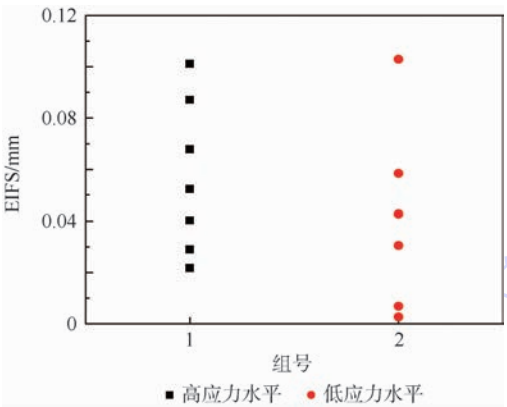


图 14　EIFS 值

Fig. 14　EIFS value

3.2　2 种应力水平下的 EIFS 的对比

假定 EIFS 服从对数正态相容分布,引入随机变量 $Z=\ln\ln(a_r/X)=\ln(QT)$,其中 X 代表 EIFS 总体,T 为裂纹萌生时间总体,Z 满足正态分布。EIFS 的分布函数如式(6)所示。分布参数可通过式(7)进行估计。在 $a_r=0.5$ mm 时,得到 2 种应力水平下的 Z 的分布概率函数,分布参数如表 2 所示。

$$F_X(x)=1-\Phi\left(\frac{\ln\ln(a_r/x)-\mu_z}{\sigma_z}\right) \tag{6}$$

式中:x 为 EIFS 值。

$$\begin{cases}\bar{z}=\hat{\mu}_z=\dfrac{1}{n}\sum_{i=1}^{n}z_i\\ s=\hat{\sigma}_z=\sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(z_i-\hat{\mu}_z)^2}\end{cases} \tag{7}$$

表 2　2 种应力水平下 Z 分布参数估计

Table 2　Z distribution parameter estimation for two stress levels

应力水平	$\hat{\mu}_z(\bar{z})$	$\hat{\sigma}_z(s)$
应力水平 1	0.806	0.250
应力水平 2	1.043	0.443

1) 方差齐性检验

采用 F 检验法,取显著度水平 $\alpha=0.05$,对 2 种应力水平下 Z_1(应力水平 1),Z_2(应力水平 2)进行方差齐性检验,根据参考文献[20],F 值如式(8)所示。因此,可认为 Z_1,Z_2 具有方差齐性,其综合标准差如式(9)所示。

$$F=\left(\frac{\max(s_1,s_2)}{\min(s_1,s_2)}\right)^2=3.14<F_{0.975}(5,6)=5.99 \tag{8}$$

$$s_{W,12}=\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}=0.351 \tag{9}$$

2) 差异性检验

采用 t 检验法,取显著度水平为 $\alpha=0.05$,根据参考文献[20],t 值如式(10)所示,因此可认为在这 Z_1,Z_2 中值寿命没有显著差异。因此 2 个样本来自同一母体,即 EIFS 与应力水平无关。

$$t=\frac{|\bar{z}_1-\bar{z}_2|}{s_{W,12}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}=1.21<t_{0.95}(n_1+n_2-2)=t_{0.95}(11)=2.201 \tag{10}$$

3.3　通用 EIFS 估计

在 2 种应力水平下,通过式(5)得到指定 a_r 下 EIFS 分布的一组大样本,其母体的分布参数可通过式(7)进行估计。对$(\hat{\mu}_z,\hat{\sigma}_z)$进行参数优化,在 0.5 ~ 1.0 mm 之间,间隔为 0.05 选择若干个 a_r,选择使得偏差平方和 SSE(a_r)最小的 a_r,此时对应的 a_r,$\hat{\mu}_z$,$\hat{\sigma}_z$ 则作为分布参数的估计值。当 $a_r=0.5$ mm 时,SSE 最小。此时 EIFS 的分布函数如式(11)所示,分布概率曲线与平均秩法估计值如图 15所示。

$$F_X(x)=1-\Phi\left(\frac{\ln\ln(0.5/x)-0.915}{0.358}\right) \tag{11}$$

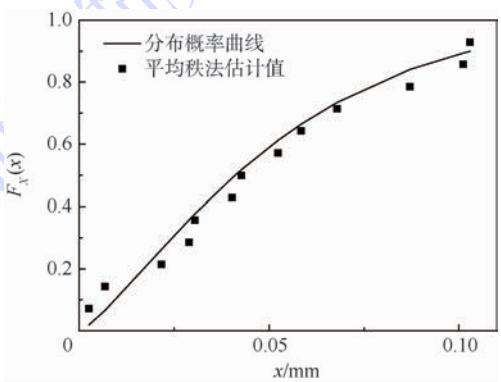


图 15　分布概率函数曲线

Fig. 15　Probability distribution function curve

4　讨　论

1) 在实验室条件下,试验机夹持边界条件可等效为远场均布拉伸载荷和试件端部除轴向位移之外的所有约束的共同作用。文献[16]表明对于不对称裂纹,在裂纹扩展过程中,试件刚度发生

北航学报 赠阅

变化,导致远场弯矩发生变化,因此在裂纹扩展过程中,Q 有一定变化。由于含孔下陷试件偏弯现象比较明显,这种影响更为显著。3.1 节为得到 EIFS 控制曲线,进行了倒角裂纹应力强度因子的有限元计算。计算结果表明应力强度因子与裂纹尺寸的关系可用幂函数表示。这说明弯矩的变化对裂纹扩展方程的影响较小,可忽略。因此,选取相对小裂纹扩展方程来描述裂纹扩展规律是合理的。

2) 假设检验表明在本文定义的裂纹特征尺寸下,2 种应力水平下的 EIFS 分布无显著差异,说明本文计算得到的 EIFS 与应力水平无关,可以作为试件初始缺陷的表征。但是,一方面 EIFS 是一个当量化的数学概念,并非实际的材料初始缺陷,另一方面,很难通过断口判读直接得到 0 时刻的裂纹长度[2],因此并不能直接通过试验结果验证本文中计算得到的 EIFS 分布的合理性。

3) 由图 15 可看出,该种裂纹形态的 EIFS 值在 0.1 mm 以内,中值约为 0.05 mm。若使用孔角椭圆裂纹来定义初始裂纹计算 EIFS,中值约为 0.11 mm,部分 EIFS 值甚至达到 0.17 mm。根据文献[8],铝合金含孔细节不同加工工艺下的 EIFS 均在 0.1 mm 以下,因此使用本文定义的初始裂纹和特征尺寸计算的 EIFS 更为合理。

5 结 论

本文针对孔边倒角处萌生裂纹的含孔下陷细节,选取合适的裂纹特征尺寸。并基于该尺寸,由耐久性试验断口数据反推得到 EIFS 的分布。有以下结论:

1) 在裂纹相对较小(1.5 mm 以下)情况下,孔角椭圆裂纹和倒角裂纹的应力强度因子分布存在很大差异,进行塑性区估计时应采用真实的裂纹形态。

2) 对于倒角处萌生的裂纹,应将裂纹特征尺寸定义为倒角与试件表面的交点到裂纹前缘沿孔径向的距离。

3) 采用反推法得到的倒角裂纹 EIFS 分布与应力水平无关。说明该特征尺寸下的 EIFS 可以合理表征结构的原始疲劳质量。

参考文献 (References)

[1] TONG Y C. Literature review on aircraft structural risk and reliability analysis: DSTO-TR-1110 [R]. Melboure: DSTO Aeronautical and Maritime Research Labratory, 2001.

[2] RUDD J L, GRAY T D. Quantification of fastener-hole quality [J]. Journal of Aircraft, 1978, 15(3): 143-147.

[3] RUDD J L. Applications of the equivalent initial quality method: AFFDL-TM-76-83-FBE [R]. Dayton: Wright-Patterson AFB, 1977.

[4] MANNING S D, YANG J N, SHINOZUKA M. USAF durability design handbook: Guidelines for the analysis and design of durable aircraft structures: AFFDL-TR84-3027 [R]. Dayton: Wright-Patterson AFB, 1984.

[5] YANG J N. Statistical estimation of economic life for aircraft structures [J]. Journal of Aircraft, 1980, 17(7): 528-535.

[6] BARTER S A, MOLENT L. Fatigue cracking from a corrosion pit in an aircraft bulkhead [J]. Engineering Failure Analysis, 2014, 39(4): 155-163.

[7] MOLENT L. A review of equivalent pre-crack sizes in aluminium alloy 7050-T7451 [J]. Fatigue & Fracture of Engineering Materials & Structures, 2014, 37(10): 1055-1074.

[8] MOLENT L, SUN Q, GREEEN A J. Characterisation of equivalent initial flaw sizes in 7050 aluminium alloy [J]. Fatigue & Fracture of Engineering Materials & Structures, 2006, 29(11): 916-937.

[9] MATTRAND C, BOURINET J M, THÉRET D. Analysis of fatigue crack growth under random load sequences derived from military in-flight load data [C] // ICAF 2011 Structural Integrity: Inuence of Eciency and Green Imper-atives. Berlin: Springer-Verlag, 2011: 399-413.

[10] MATTRAND C, BOURINET J M. Random load sequences and stochastic crack growth based on measured load data [J]. Engineering Fracture Mechanics, 2011, 78(17): 3030-3048.

[11] XIANG Y, LIU Y. EIFS-based crack growth fatigue life prediction of pitting-corroded test specimens [J]. Engineering Fracture Mechanics, 2010, 77(8): 1314-1324.

[12] FAWAZ S A. Equivalent initial flaw size testing and analysis of transport aircraft skin splices [J]. Fatigue & Fracture of Engineering Materials & Structures, 2003, 26(3): 279-290.

[13] WANG D Y. An investigation of initial fatigue quality: STPZ8860S [R]. West Conshohocken: ASTM Special Technical Publication, 1982.

[14] 航空航天部 AFFD 系统办公室. 美国空军耐久性手册背景材料: 第Ⅷ卷 [M] 西安: 航空航天部 AFFD 系统办公室, 1989.
AFFD System Engineering Office of the Department of Aeronautics and Astronautics. The background information of USAF durability design handbook: Vol. Ⅷ [M]. Xi' an: AFFD System Engineering Office of the Department of Aeronautics and Astronautics, 1989 (in Chinese).

[15] YANG J, MANNING S, RUDD J, et al. Effect of specimen size and notch geometry on equivalent initial flaw size distribution [C] // AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and Exhibit. Reston: AIAA, 1996: 385-397.

[16] 曹淑森, 贺小帆, 杨博霄, 等. 夹持边界条件下表面裂纹应力强度因子求解 [J]. 北京航空航天大学学报, 2014, 40(11): 1637-1642.
CAO S S, HE X F, YANG B X, et al. Solution of stress intensity factor of surface cracked geometry with clamped ends [J]. Jour-

nal of Beijing University of Aeronautics and Astronautics,2014,40(11):1637-1642(in Chinese).

[17] 张行.断裂力学[M].北京:中国宇航出版社,1990:72-169.
ZHANG X. Fracture mechanics[M]. Beijing: China Aerospace Press,1990:72-169(in Chinese).

[18] 刘文珽.概率断裂力学与概率损伤容限/耐久性[M].北京:北京航空航天大学出版社,1999:144-181.
LIU W T. Probabilistic fracture mechanics and probability damage tolerance/durability [M]. Beijing: Beihang University Press,1999:144-181(in Chinese).

[19] PROVAN J W. Probabilistic fracture mechanics and reliability [M]. Leiden: Martinus Nijhoff Publishers,1987:70-72.

[20] 高镇同.疲劳应用统计学[M].北京:国防工业出版社,1986:277-295.
GAO Z T. Fatigue application statistics[M]. Beijing: National Defence of Industry Press,1986:277-295(in Chinese).

作者简介:

孙晓娜　女,硕士研究生。主要研究方向:结构耐久性。

贺小帆　男,博士,副教授。主要研究方向:疲劳/断裂、结构可靠性与耐久性、腐蚀条件下使用寿命评定。

Determination method of equivalent initial flaw size for crack initiated at hole chamfering

SUN Xiaona[1], HE Xiaofan[1,*], LI Yuhai[2]

(1. School of Aeronautic Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China;
2. Aviation Industry Corporation of China, Ltd., Beijing 100022, China)

Abstract: Cracks often initiate at the chamfering for sagging holes. The initial fatigue quality (IFQ) of this type of crack needs to be determined prior to economic life assessment. Firstly, finite element models with and without chamfering are developed so as to investigate the effect of chamfering on the stress intensity factors at the crack front. The results show that the chamfering imposes great influence on the stress intensity factors of relatively small crack. Secondly, in order to characterize the IFQ of sagging holes, a circular-front crack initiated at the intersection of the chamfering and specimen surface is taken as the initial flaw, with the radial distance of crack front from the initiation site as the crack size. Finally, the crack growth equation for relatively small cracks is employed to characterize the crack growth behavior, and the equivalent initial flaw size (EIFS) distribution is obtained through back-extrapolation. Statistical analyses show that the EIFS distribution parameters obtained by using the proposed crack size definition are independent of stress levels.

Keywords: initial fatigue quality (IFQ); stress intensity factor; finite element; crack growth equation; equivalent initial flaw size (EIFS)

Received: 2017-01-03; **Accepted**: 2017-03-10; **Published online**: 2017-04-13 15:05
URL: kns.cnki.net/kcms/detail/11.2625.V.20170413.1505.002.html

* **Corresponding author**. E-mail: xfhe@buaa.edu.cn

北京航空航天大学学报

北航学报 赠阅

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2016.0941

基于空间混联机构的人体肩部骨骼运动模型

全下文载

聂超[1], 宋智斌[1,*], 戴建生[1,2]
(1. 天津大学 机械工程学院, 天津 300350; 2. 伦敦大学 国王学院, 伦敦 WC2R 2LS)

摘 要: 为了描述人体肩部骨骼系统的运动特征,将肩胛骨与胸廓的相对运动关系定义为类似于圆柱-平面副的运动约束,建立了肩部骨骼系统的空间混联机构模型。首先定义了肩部复合体各关节的类型,并完成了肩带部分和整个肩部机构的自由度分析。然后通过定义附着于各骨骼上的局部坐标系,以齐次坐标变换矩阵和矢量法建立机构的运动分析方程,求得其关节位置的闭合解。最后为了验证该模型,以获得自肩部运动实验的骨骼姿态数据反向驱动该机构模型,从而得到肩胛骨姿态的计算结果,并与测量结果进行对比。结果表明:该机构模型能够反映肩部骨骼的运动约束关系。同时,该模型可以通过缩放处理从而用于适应不同个体的骨骼几何特征。

关 键 词: 肩部复合体; 肩带; 骨骼运动; 齐次坐标变换; 空间混联机构
中图分类号: TH112; Q66
文献标识码: A **文章编号**: 1001-5965(2018)01-0196-09

人体肩部的骨骼、肌肉、韧带和周围皮肤构成的肩部复合体(shoulder complex)闭环系统,是人体最为复杂的运动功能部位之一[1]。研究肩关节复合体运动特征,特别是骨骼系统的运动特征对于人体体能训练,肩部功能障碍诊断以及康复等具有重要意义。同时,正确描述肩部骨骼的姿态和运动行为是建立肩部复合体模型的关键[2]。

为了准确描述肩部骨骼的运动特征,国内外学者在肩部骨骼的机构学建模方面做了大量的研究工作,这些机构模型主要分为 2 类:开式链机构和闭式链机构[1]。在开式链机构方面,从 1965 年 Dempster[3] 建立了第一个简单的肩部二维串联连杆模型开始,Engin 和 Chen[4] 于 1986 年又提出了一种将肱骨相对于人体躯干的运动描述成三维 6 自由度的刚体模型,之后 Engin 和 Tumer[5] 进一步将该肩部机构模型的自由度增加到 9。Dempster[3] 和 Engin 等[4-5] 虽然建立了肩部骨骼的三维机构模型,但是这些机构模型都是从仿人机器人领域的需求出发而建立的,其目的是使模型具有与人体肱骨相近的活动度和运动范围,而非重现肩带内部骨骼的运动约束关系,即没有系统描述人体肩部复合体所有骨骼的运动本质特征。

进入 20 世纪 90 年代之后,随着光标位置检测技术[6-7]、电磁跟踪技术[8] 以及三维 X 射线照相定位[7,9-10] 等技术的发展,肩部骨骼的姿态测量成为了可能,同时肩部机构模型也朝着更贴近于实际解剖结构的并联机构方向发展。Garner 和 Pandy[11] 于 2001 年以美国国家医学图书馆的核磁共振影像资料数据库为依据建立了一个具有 13 自由度的人体上肢骨骼模型,其中胸锁关节(Sternoclavicular joint, SC)、肩锁关节(Acromioclavicular joint, AC)以及盂肱关节(Glenohumeral joint, GH)皆被视为球形铰链,并首次将胸廓建模为椭球,同时用肩胛骨在椭球表面的相对滑动

收稿日期: 2016-12-15; **录用日期**: 2017-01-13; **网络出版时间**: 2017-03-08 09:29
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20170308.0929.002.html
基金项目: 国家自然科学基金 (51475322,51535008)
* **通信作者**. E-mail: songzhibin@tju.edu.cn

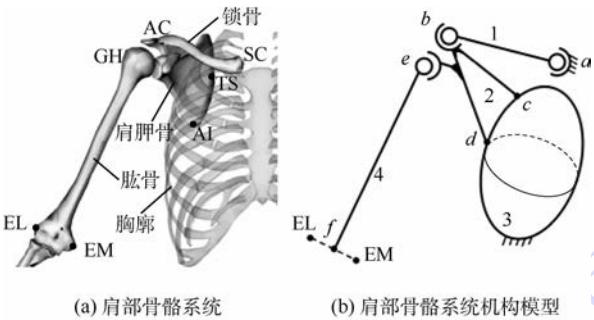
引用格式: 聂超, 宋智斌, 戴建生. 基于空间混联机构的人体肩部骨骼运动模型[J]. 北京航空航天大学学报, 2018, 44(1): 196-204. NIE C, SONG Z B, DAI J S. A shoulder skeletal kinematic model based on spatial hybrid mechanism[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(1): 196-204 (in Chinese).

描述肩胛胸廓关节(Scapulathoracic articulation, ST)的运动。该模型虽未从机构学角度精确定义 ST 几何约束,也未对该模型进行运动分析,但是却提供了一套较为完整的人体结构数据可供他人使用。基于此模型的概念,Maurel 等[12-13]进一步将 ST 视作单个点与面接触而形成的约束,但是该研究同样没有对机构的整体进行分析,而是将闭式链拆开,以 ST 处的虚拟接触点的位置为输入,以此来模拟和分析肩胛骨的运动。

综上所述,在建立基于肩部骨骼解剖结构的骨骼机构模型时,肩带中的 3 个关节(SC、AC、ST)以及 3 块骨骼(锁骨、肩胛骨、胸骨)的运动必须予以充分考虑。因此,本文基于肩部骨骼的解剖结构,在 Garner 和 Pandy[11]、Maurel 等[12-13]的基础上提出了一个基于空间混联机构的描述肩带内部骨骼运动约束关系的运动模型,并建立了其位置闭环方程,得到了位置分析的闭合解,并最终通过运动实验数据验证了该模型的正确性。

1　骨骼系统机构学模型的建立

人体肩部可认为是由肩带(shoulder girdle)和肱骨(humerus)2 部分构成的骨骼系统,其中肩带由胸廓(thorax)、锁骨(clavicle)、肩胛骨(scapula) 3 块骨骼组成,可视为一个闭环骨骼系统。与肩部运动相关的关节包括:胸锁关节、肩锁关节、肩胛胸廓关节和盂肱关节,如图 1 所示。



(a) 肩部骨骼系统　　(b) 肩部骨骼系统机构模型

图 1　肩关节机构模型简图

Fig. 1　Schematic of shoulder joint mechanism model

本文使用一空间混联机构模型来模拟肩部骨骼系统,相比于传统的骨骼模型能够更真实地的反映肩部复合体的内部骨骼运动情况,如图 1 所示。其中,胸锁关节 a、肩锁关节 b、盂肱关节 e 皆被视为理想 3 自由度球形铰链关节。与 Maurel 模型[12-13]和 Tondu 模型[14]中关于 ST 的定义不同的是,本文将胸廓视为一个椭球体,而肩胛骨通过其肩胛冈三角 c(Trigonum spinae Scapulae, TS)与肩胛骨下角 d(Angulus Inferior, AI)与胸廓椭球表面实现两点接触,其可等效于一个圆柱-平面副,具有 4 个自由度。其中,构件 1 表示锁骨,构件 2 表示肩胛骨,构件 3 表示胸廓表面,构件 4 表示上肢肱骨。由于本文的研究对象仅为肩部,因此假设上肢处于伸直状态,且上臂和前臂之间无相对运动,以肱骨长轴代表构件 4,即 GH 中心 e 至肘关节中心 f 的连线。其中 f 定义为外上踝(Lateral Epicondyle, EL)和内上踝(Medial Epicondyle, EM)连线的中点。关节 a 与 b 中心距离设为 l_1, b 与 c 距离为 l_2,b 与 d 的距离为 l_3,肱骨长度为 l_4。

2　混联机构模型的自由度分析

要分析混联机构模型混联机构的运动特征,首先要分析其自由度。由 Kutzbach-Grübler 公式可得机构的自由度为

$$M = 6(n - g - 1) + \sum_{i=1}^{g} f_i \tag{1}$$

式中:n 为机构构件数目;g 为运动副数目;f_i 为第 i 个运动副的相对自由度数。模型共有 4 个构件,4 个运动副,其中 SC、AC 和 GH 为球副,自由度皆为 3,ST 为圆柱-平面副,自由度为 4,因此肩部机构总的自由度为 7。然而在肩部运动过程中,锁骨绕自身轴线方向的转动非常微小,可忽略不计[15-17]。同时从机构角度可知,对于关节 a 而言,其绕锁骨轴线方向的转动为虚自由度。因此整个肩部机构自由度为 6,其中对应于肩带的并联部分自由度为 3。

3　混联机构的运动学

3.1　坐标系的建立

为了对机构进行位置分析,定义附着于单个骨骼的坐标系如图 2 所示。假定胸廓固定不动,定义附着于其上的坐标系 $\boldsymbol{S}_0\{a\text{-}x_0y_0z_0\}$ 为固定系,其原点位于 a 点,x_0、y_0、z_0 轴分别平行于人体的解剖学主轴——冠状轴、矢状轴、垂直轴。$\boldsymbol{S}_1\{a\text{-}x_1y_1z_1\}$ 为锁骨系,其原点位于胸锁关节 a 处,z_1 轴沿锁骨轴线方向,由 a 指向 b,由于锁骨只有 2 个骨骼标记点,因此定义其 x_1 轴位于水平面且垂直于 z_1,y_1 则由右手定则确定。$\boldsymbol{S}_2\{b\text{-}x_2y_2z_2\}$ 为肩胛骨系,其原点位于肩锁关节 b 点,其 z_2 轴为点 b 指向点 c,x_2 垂直于 b、c、d 3 点所确定的肩胛骨平面并指向前方,且垂直于 z_2 轴,y_2 轴则由右手定则确定。$\boldsymbol{S}_3\{O\text{-}x_3y_3z_3\}$ 为胸廓系,其原点位于胸廓椭球中心,其坐标轴分别与椭球的3个

北航学报 赠阅

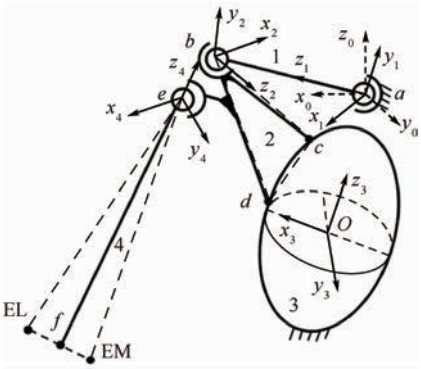


图 2 肩部骨骼机构坐标系定义

Fig. 2 Definition of related reference frames of shoulder skeleton mechanism

主轴重合,且同时平行于人体解剖主轴,即 $\boldsymbol{S}_3 /\!/ \boldsymbol{S}_0$。$\boldsymbol{S}_4\{e\text{-}x_4y_4z_4\}$ 为肱骨坐标系,其原点位于 e 点处,z_4 沿肱骨长轴方向且由 f 指向 e,y_4 轴垂直于点 e、EL、EM 所确定的平面且指向内侧,x_4 由右手定则确定。

3.2 运动学分析

在肩带中,胸廓、锁骨、肩胛骨互相之间通过关节 SC、AC、ST 相互约束,从而构成了肩部混联机构的并联部分。由于肩带机构的自由度为 3,因此只需任意 3 个关节变量作为输入,就可确定整个肩带的姿态。本文进行运动学分析的目的在于以解析解的方式建立肩带各骨骼的运动约束关系,并验证所提出的空间混联机构模型对实际肩部运动预测的准确性。

从系 $\boldsymbol{S}_0\{a\text{-}x_0y_0z_0\}$ 到 $\boldsymbol{S}_1\{a\text{-}x_1y_1z_1\}$,可认为是先绕 z 轴旋转 θ_1,再绕 x 轴旋转 θ_2,再绕 z 轴旋转 θ_3 而得到。从系 $\boldsymbol{S}_1$ 到 $\boldsymbol{S}_2$,可认为是先沿 z 轴平移 l_1,再绕 y 轴旋转 θ_4,再绕 x 轴旋转 θ_5,最后绕 y 轴旋转 θ_6 而得到。因此有变换矩阵

$$
{}_1^0\boldsymbol{T} = \mathrm{Rot}(z,\theta_1)\mathrm{Rot}(x,\theta_2)\mathrm{Rot}(z,\theta_3) =
\begin{bmatrix}
\cos\theta_1\cos\theta_3 - \sin\theta_1\cos\theta_2\sin\theta_3 & -\cos\theta_1\sin\theta_3 - \sin\theta_1\cos\theta_2\cos\theta_3 & \sin\theta_1\sin\theta_2 & 0\\
\sin\theta_1\cos\theta_3 + \cos\theta_1\cos\theta_2\sin\theta_3 & -\sin\theta_1\sin\theta_3 + \cos\theta_1\cos\theta_2\cos\theta_3 & -\cos\theta_1\sin\theta_2 & 0\\
\sin\theta_2\sin\theta_3 & \sin\theta_2\cos\theta_3 & \cos\theta_2 & 0\\
0 & 0 & 0 & 1
\end{bmatrix} \tag{2}
$$

$$
{}_2^1\boldsymbol{T} = \mathrm{Trans}(z,l_1)\mathrm{Rot}(y,\theta_4)\mathrm{Rot}(x,\theta_5)\mathrm{Rot}(y,\theta_6) =
\begin{bmatrix}
\cos\theta_4\cos\theta_6 - \sin\theta_4\cos\theta_5\sin\theta_6 & \sin\theta_4\sin\theta_5 & \cos\theta_4\sin\theta_6 + \sin\theta_4\cos\theta_5\cos\theta_6 & 0\\
\sin\theta_5\sin\theta_6 & \cos\theta_5 & -\sin\theta_5\cos\theta_6 & 0\\
-\sin\theta_4\cos\theta_6 - \cos\theta_4\cos\theta_5\sin\theta_6 & \cos\theta_4\sin\theta_5 & -\sin\theta_4\sin\theta_6 + \cos\theta_4\cos\theta_5\cos\theta_6 & l_1\\
0 & 0 & 0 & 1
\end{bmatrix} \tag{3}
$$

假设椭球中心点 O 在固定系 $\boldsymbol{S}_0$ 中的位置为 ${}^0\boldsymbol{p}_O = [x_{O,0} \quad y_{O,0} \quad z_{O,0} \quad 1]^{\mathrm{T}}$,则椭球系 $\boldsymbol{S}_3$ 相对于固定系 $\boldsymbol{S}_0$ 的变换矩阵为

$$
{}_3^0\boldsymbol{T} = \begin{bmatrix} 1 & 0 & 0 & x_{O,0}\\ 0 & 1 & 0 & y_{O,0}\\ 0 & 0 & 1 & z_{O,0}\\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}
$$

设胸廓椭球曲面的半轴分别为 c_1、c_2、c_3,而点 c 和点 d 位于胸廓椭球表面上,其在系 $\boldsymbol{S}_3$ 的位置可由椭球的参数方程表示如下:

$$
{}^3\boldsymbol{p}_c = [c_1\sin\varphi_c\cos\phi_c \quad c_2\sin\varphi_c\sin\phi_c \quad c_3\cos\varphi_c \quad 1]^{\mathrm{T}} \tag{5}
$$

$$
{}^3\boldsymbol{p}_d = [c_1\sin\varphi_d\cos\phi_d \quad c_2\sin\varphi_d\sin\phi_d \quad c_3\cos\varphi_d \quad 1]^{\mathrm{T}} \tag{6}
$$

式中:φ_c、ϕ_c、φ_d、ϕ_d 为分别描述点 c 与点 d 位置的参数。如图 3 所示,c、d 同时也位于肩胛骨上。则 c、d 在系 $\boldsymbol{S}_2$ 中的位置分别为

$$
{}^2\boldsymbol{p}_c = [0 \quad 0 \quad l_2 \quad 1]^{\mathrm{T}} \tag{7}
$$

$$
{}^2\boldsymbol{p}_d = [0 \quad l_3\sin\gamma \quad l_3\cos\gamma \quad 1]^{\mathrm{T}} \tag{8}
$$

式中:γ 为 $\overrightarrow{bc}$ 与 $\overrightarrow{bd}$ 的夹角,由于 b、c、d 皆位于肩胛骨上,因此 l_2、l_3、γ 皆为恒定值。综上,利用 c、d 两点既位于肩胛骨上又位于胸廓椭球上这一性质,可建立肩带机构的闭环回路方程如下:

$$
{}_1^0\boldsymbol{T}\,{}_2^1\boldsymbol{T}\,{}^2\boldsymbol{p}_c - {}_3^0\boldsymbol{T}\,{}^3\boldsymbol{p}_c = \boldsymbol{0} \tag{9}
$$

$$
{}_1^0\boldsymbol{T}\,{}_2^1\boldsymbol{T}\,{}^2\boldsymbol{p}_d - {}_3^0\boldsymbol{T}\,{}^3\boldsymbol{p}_d = \boldsymbol{0} \tag{10}
$$

式(9)与式(10)即为肩带机构的关节位置求解约束方程,令其各项分量分别为0可得到6个

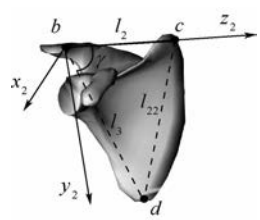


图 3 点 c、d 在肩胛骨系中的位置

Fig. 3 Position of points c and d in scapula frame

约束方程，其中含有 θ_1、θ_2、θ_3、θ_4、θ_5、θ_6、φ_c、ϕ_c、φ_d、ϕ_d 共 10 个关节变量。而锁骨绕自身轴线方向的转动为虚自由度，若令 $\theta_3=0$，则只需以任意 3 个关节变量作为输入就可求解其余 6 个关节变量，本文以 θ_1、θ_2 和 ϕ_d 作为输入变量。

由锁骨系的定义可知 ${}^1\boldsymbol{p}_b=[0\quad 0\quad l_1\quad 1]^{\mathrm{T}}$，若已知 θ_1、θ_2，则对于点 b 有

$$ {}^3\boldsymbol{p}_b=({}^0_3\boldsymbol{T})^{-1}{}^0_1\boldsymbol{T}\,{}^1\boldsymbol{p}_b=[x_{b,3}\quad y_{b,3}\quad z_{b,3}\quad 1]^{\mathrm{T}} \tag{11}$$

图 4 为肩胛骨与胸廓椭圆接触示意图。为了简化计算，首先分析肩胛骨上的 b、c、d 3 点，由于 $\triangle bcd$ 所代表的肩胛骨被认为是刚体，因此有

$$\left\|{}^3\boldsymbol{p}_b-{}^3\boldsymbol{p}_c\right\|=l_2 \tag{12}$$

$$\left\|{}^3\boldsymbol{p}_b-{}^3\boldsymbol{p}_d\right\|=l_3 \tag{13}$$

$$\left\|{}^3\boldsymbol{p}_c-{}^3\boldsymbol{p}_d\right\|=l_{22} \tag{14}$$

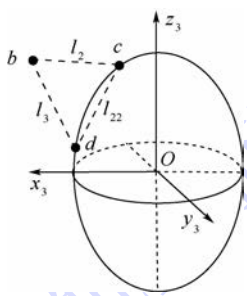


图 4　肩胛-胸廓椭圆约束示意图

Fig. 4　Schematic of the constraint between scapula and thorax ellipsoid

将式(5)、式(6)、式(11)代入式(12)～式(14)即可得到 3 个含有变量 ϕ_d、φ_d、φ_c、ϕ_c 的约束方程。若 ϕ_d 为已知，则可求解得到 φ_d、φ_c、ϕ_c。再由式(5)和式(6)可计算得到 ${}^3\boldsymbol{p}_c$ 和 ${}^3\boldsymbol{p}_d$ 的值。同时可得

$${}^0\boldsymbol{p}_c={}^0_3\boldsymbol{T}\,{}^3\boldsymbol{p}_c=[x_{c,0}\quad y_{c,0}\quad z_{c,0}\quad 1]^{\mathrm{T}} \tag{15}$$

$${}^0\boldsymbol{p}_d={}^0_3\boldsymbol{T}\,{}^3\boldsymbol{p}_d=[x_{d,0}\quad y_{d,0}\quad z_{d,0}\quad 1]^{\mathrm{T}} \tag{16}$$

再将 ${}^0\boldsymbol{p}_c$ 和 ${}^0\boldsymbol{p}_d$ 代入回路方程式(9)和式(10)，则可推导得到中间变量 A、B、C、D 如下：

$$A=\sin\theta_5\cos\theta_6=(x_{c,0}\sin\theta_1\cos\theta_2-y_{c,0}\cos\theta_1\cos\theta_2-z_{c,0}\sin\theta_2)/l_2 \tag{17}$$

$$B=\sin\theta_4\sin\theta_5=[l_2(x_{d,0}\cos\theta_1+y_{d,0}\sin\theta_1)-l_3\cos\gamma(x_{c,0}\cos\theta_1+y_{c,0}\sin\theta_1)]/(l_2l_3\sin\gamma) \tag{18}$$

$$C=\cos\theta_4\sin\theta_5=[l_3\cos\gamma(l_1\cos\theta_2-l_2\sin\theta_2-z_{c,0})-l_2\cos\theta_2(y_{c,0}\cos\theta_1\sin\theta_2+l_1-x_{d,0}\sin\theta_1\sin\theta_2-z_{d,0}\cos\theta_2)]/(l_2l_3\sin\gamma\cos\theta_2) \tag{19}$$

$$D=\cos\theta_5=(Al_3\cos\gamma-x_{d,0}\sin\theta_1\cos\theta_2+y_{d,0}\cos\theta_1\cos\theta_2+z_{d,0}\sin\theta_2)/(l_3\sin\gamma) \tag{20}$$

由此可得到 θ_4、θ_5、θ_6 的封闭解如下：

$$\theta_4=\arctan(B/C) \tag{21}$$

$$\theta_5=\arctan(C/(D\cos\theta_4)) \tag{22}$$

$$\theta_6=\arccos((A\cos\theta_4)/C) \tag{23}$$

4　运动实验

4.1　实验数据来源

本文以 Bolsterlee 和 Veeger[15] 于 2014 年发表的上肢运动实验数据库(Shoulder Database V1.1)作为机构运动分析的数据来源。该运动数据库记录了人体上肢在完成一定的动作任务时，相关人体骨骼标记点在实验室坐标系中的空间位置坐标。所有相关数据可以从美国国家康复仿真研究中心(NCSRR)的网站中获得[18]。

该实验利用一套运动捕捉系统(optotrak system)检测分别固定于胸廓、肩胛骨、肱骨、前臂和手掌上的 5 个标记群的空间位置，如图 5 所示[12]。所有骨骼标记点的位置按照 Wu 等[19]提出的 ISB (International Society of Biomechanics)标准布置，如图 6 所示。图中骨骼标记点已经由 ISB 标准详细定义[19]。

图 6 中坐标系 $\boldsymbol{S}_{\mathrm{t}}$、$\boldsymbol{S}_{\mathrm{c}}$、$\boldsymbol{S}_{\mathrm{s}}$、$\boldsymbol{S}_{\mathrm{h}}$ 为 ISB 标准中的

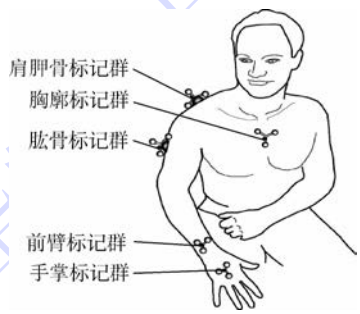


图 5　标记群的布置[12]

Fig. 5　Placement of marker clusters[12]

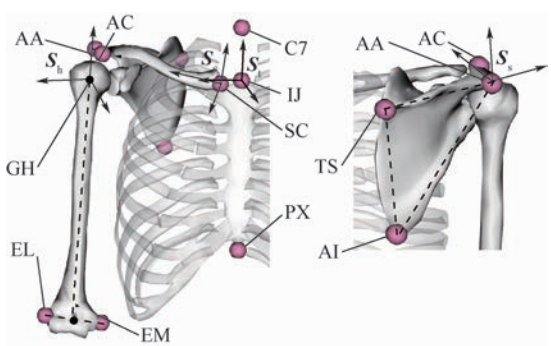


图 6　ISB 标准中的骨骼标记点与局部坐标系

Fig. 6　Skeletal landmarks and local frames defined in ISB standard

北航学报 赠阅

各骨骼的局部坐标系,分别对应于计算系中的 $\boldsymbol{S}_0$、$\boldsymbol{S}_1$、$\boldsymbol{S}_2$、$\boldsymbol{S}_4$。标记点的位置数据采集频率为 100 Hz。本文从该数据库中选取一个男性对象(年龄 29 岁,身高 186 cm,体重 109 kg)的运动数据作为参考。

在实验过程中,受测对象被要求完成运动范围任务,包括上肢前屈(FLEX)、外展(ABD)和肩胛平面运动(SCAP)。

4.2 机构尺寸的确定

Garner 和 Pandy[11,20] 依据获得自美国国家医学图书馆的人体影像数据建立了具有较高尺寸精度的人体上肢的骨骼模型。本文据此确定肩部机构模型中的各杆件尺寸,如表 1 和表 2 所示。表 3为 Garner 标准模型与实验对象的骨骼特征尺寸在系 $\boldsymbol{S}_t$ 中的坐标值的对比,实验对象缩放系数为(1.20,1.05,1.14),平均系数为 1.13。

由于 Garner 骨骼模型具有较高的尺寸精度且包含胸廓椭球的尺寸,因此将其作为标准模型,而实验对象的骨骼尺寸则通过对标准模型的缩放而得到。相应的缩放系数可通过对比标准模型与实验对象各骨骼标记点的位置而确定[15]。

实验对象的机构模型中各杆件尺寸可由标准模型的相应尺寸乘以平均缩放系数而确定。胸廓椭球的位置及 3 个半轴的长度则由标准模型的椭球尺寸乘以相应方向的缩放系数而确定。

表 1 标准模型的各构件尺寸参数

Table 1 Size parameters of each link in generic model

参 数	l_1/mm	l_2/mm	l_3/mm	l_{22}/mm	l_4/mm	γ/(°)
数 值	182.5	129.1	185.9	117.6	301.6	38.8

表 2 标准模型中胸廓椭球的几何尺寸

Table 2 Size parameters of thorax ellipsoid in generic model mm

几何尺寸	c_1	c_2	c_3	O 点位置 ${}^0\boldsymbol{p}_0$	e 点位置 ${}^2\boldsymbol{p}_e$
数 值	144.6	95.6	211.7	[0 −62.1 −152.1 1]$^{\mathrm{T}}$	[28.1 −43.9 −164.5 1]$^{\mathrm{T}}$

表 3 实验对象的骨骼特征尺寸参数

Table 3 Skeletal size parameters of experimental subject mm

骨骼标记点	标准模型	实验对象
IJ	(0,0,0)	(0,0,0)
PX	(31.9,−132.7,−9.8)	(56.0,−163.0,−2.4)
C7	(−124.2,54.1,0)	(−123.1,80.7,−4.7)
T8	(−156.61,−171.5,0)	(−179.1,−133.2,11.2)
AA	(−105.6,7.5,182.6)	(−73.5,9.3,205.1)
TS	(−156.0,−11.7,75.0)	(−136.3,5.1,121.8)
AI	(−156.7,−126.2,101.9)	(−141.6,−126.5,116.2)
AC	(−71.8,26.6,165.1)	(−37.0,40.4,167.2)
SC	(−2.8,−15.2,1.4)	(15.3,−31.0,−6.6)
椭球中心	(−62.1,−152.1,0)	(−61.6,−148.1,0)
椭球半轴	(95.6,211.7,144.6)	(114.4,222.7,164.8)

4.3 ISB 系与计算系的转换关系

运动实验所采集的标记点位置及各骨骼的姿态是在 Wu 等[19] 推荐的 ISB 标准系中记录和描述的,而 ISB 标准同时也是生物医学工程领域广泛采用的标准,本文亦将骨骼姿态的模型计算结果和实验测量结果统一放在 ISB 标准中来对比衡量。因此必须建立本文计算坐标系与 ISB 坐标系之间的转换与对应关系。

依据各局部坐标系的定义,以及表 3 中各标记点的坐标,3.1 节中用于计算的坐标系 $\boldsymbol{S}_0$、$\boldsymbol{S}_1$、$\boldsymbol{S}_2$、$\boldsymbol{S}_4$ 与 ISB 标准坐标系 $\boldsymbol{S}_t$、$\boldsymbol{S}_c$、$\boldsymbol{S}_s$、$\boldsymbol{S}_h$ 的关系可分别表示为如下旋转变换矩阵:

$$ {}^t_0\boldsymbol{R} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \tag{24} $$

$$ {}^c_1\boldsymbol{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{25} $$

$$ {}^s_2\boldsymbol{R} = \begin{bmatrix} 0.959\,9 & 0.037\,3 & -0.277\,8 \\ -0.013\,1 & -0.984\,1 & -0.177\,3 \\ -0.280\,0 & -0.173\,8 & -0.944\,1 \end{bmatrix} \tag{26} $$

$$ {}^h_4\boldsymbol{R} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \tag{27} $$

5 基于混联机构的运动模型验证

5.1 模型驱动参数的确定

通过对实验中测得的各标记点空间坐标的处理,得到关节变量 θ_1、θ_2、ϕ_d,以此 3 个独立关节变量为输入,则可确定肩带的运动姿态。

以 FLEX 运动为例,图 7 为测得的实验对象的锁骨姿态,并以 YXZ 序列下的 3 个欧拉角 γ_c、β_c、α_c 来描述[19]。

由此可得在任意时刻,锁骨系 $\boldsymbol{S}_c$ 相对于实验固定系 $\boldsymbol{S}_t$ 的姿态矩阵为

$$ {}^t_c\boldsymbol{R} = \mathrm{Rot}(y,\gamma_c)\,\mathrm{Rot}(x,\beta_c)\,\mathrm{Rot}(z,\alpha_c) \tag{28} $$

将式(24)和式(25)代入式(28),可得

$$ {}^0_1\boldsymbol{R} = ({}^t_0\boldsymbol{R})^{-1}\,{}^t_c\boldsymbol{R}\,{}^c_1\boldsymbol{R} = \begin{bmatrix} u_z & v_z & w_z \\ u_x & v_x & w_x \\ u_y & v_y & w_y \end{bmatrix} \tag{29} $$

令式(29)的矩阵的第 3 列元素与式(2)的矩阵的第 3 列前 3 个元素对应相等,则可得

$$\theta_1 = \arctan\left(-\frac{w_z}{w_x}\right) \tag{30}$$

$$\theta_2 = \arctan\left(\frac{w_z}{w_y \sin\theta_1}\right) \tag{31}$$

由此获得了输入关节变量 θ_1、θ_2，如图 8 所示。

另一个输入关节变量 ϕ_d 则由 AI 点相对于胸廓椭球系的空间位置而得到。在 FLEX 运动中，AI 在椭球系 $\boldsymbol{S}_3$ 中的位置 ${}^3\boldsymbol{p}_{\mathrm{AI}}(X_{\mathrm{AI}}, Y_{\mathrm{AI}}, Z_{\mathrm{AI}})$ 如图 9 所示。

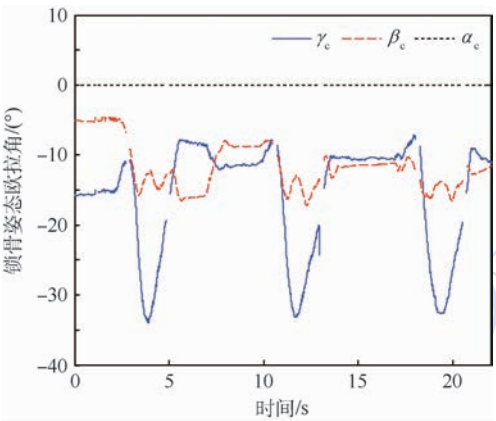


图 7　锁骨的姿态（FLEX）

Fig. 7　Posture of clavicle（FLEX）

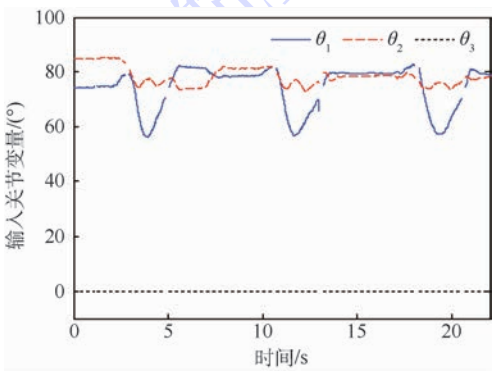


图 8　用于机构位置分析的输入关节变量 θ_1、θ_2（FLEX）

Fig. 8　Joint variables θ_1, θ_2 inputted for mechanism position analysis（FLEX）

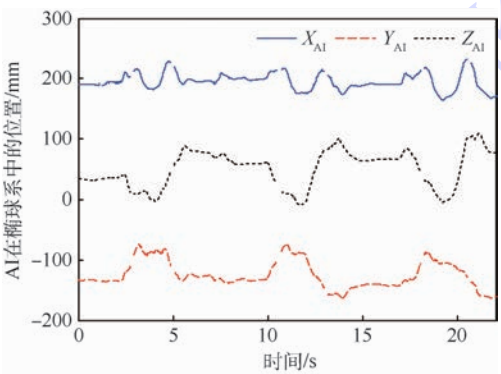


图 9　肩胛骨下角 AI 在椭球系 $\boldsymbol{S}_3$ 中的位置（FLEX）

Fig. 9　Position of angulus inferior AI described in ellipsoid frame $\boldsymbol{S}_3$（FLEX）

虽然 AI 非常接近于椭球表面，然而 AI 点始终位于肩胛骨上，与胸廓椭球表面之间存在一定间隙，与机构模型中所构想的点 d 并不完全一致。本文通过将肩胛下角 AI 沿椭球中心线方向的投影点视为等效的 d 点。如图 10 所示，AI 与椭球中心 O 的连线交椭球表面于 d。

由此得到了 ${}^3\boldsymbol{p}_d = [\,x_{d,3} \quad y_{d,3} \quad z_{d,3} \quad 1\,]^{\mathrm{T}}$，再依据式(6)，有

$$\phi_d = \arctan\left(\frac{c_1}{c_2} \cdot \frac{y_{\mathrm{AI},3}}{x_{\mathrm{AI},3}}\right) \tag{32}$$

由此则可得到 ϕ_d 的值，如图 11 所示。

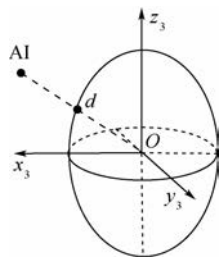


图 10　AI 投影于胸廓椭圆表面示意图

Fig. 10　Schematic of point AI projected on thorax elliptical surface

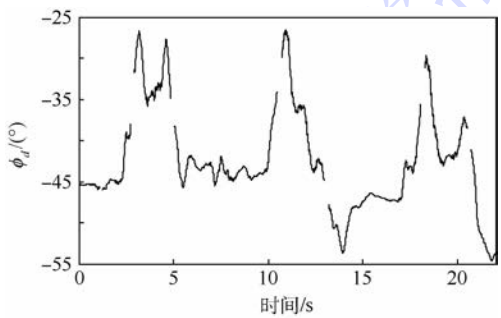


图 11　关节变量 ϕ_d（FLEX）

Fig. 11　Joint variable ϕ_d（FLEX）

至此，完整地得到了 FLEX 运动相应的 3 个输入关节变量 θ_1、θ_2、ϕ_d。同样地可以对 ABD 和 SCAP 运动进行分析，在此不再赘述。

5.2　肩胛骨姿态验证

确定了输入关节变量 θ_1、θ_2、ϕ_d 后，将其代入式(12)～式(23)中，则可求解得到其余关节变量 φ_d、φ_c、ϕ_c 与 θ_4、θ_5、θ_6 的值，进而确定整个肩带的运动。将各关节变量代入式(1)和式(2)，则可得

$${}^0_2\boldsymbol{T} = {}^0_1\boldsymbol{T}\,{}^1_2\boldsymbol{T} = \begin{bmatrix} [\,{}^0_2\boldsymbol{R}\,]_{3\times3} & [\,{}^0\boldsymbol{p}_b\,]_{4\times1} \\ 0 \quad 0 \quad 0 & \end{bmatrix} \tag{33}$$

由此可得系 $\boldsymbol{S}_2$ 相对于系 $\boldsymbol{S}_0$ 的姿态矩阵 ${}^0_2\boldsymbol{R}$，再依据式(24)和式(26)可得

$${}^{\mathrm{t}}_{\mathrm{s}}\boldsymbol{R} = {}^{\mathrm{t}}_0\boldsymbol{R}\,{}^0_2\boldsymbol{R}\,({}^{\mathrm{s}}_2\boldsymbol{R})^{-1} \tag{34}$$

同时，在肩带机构中假设以肩胛骨为末端执

北航学报 赠阅

行件。在 *YXZ* 序列下其姿态可用 3 个欧拉角 γ_s、β_s、α_s 用以描述[19]。则肩胛骨 $\boldsymbol{S}_s$ 相对于固定系 $\boldsymbol{S}_t$ 的姿态可表示为

$$ {}^{t}_{s}\boldsymbol{R} = \mathrm{Rot}(y,\gamma_s)\,\mathrm{Rot}(x,\beta_s)\,\mathrm{Rot}(z,\alpha_s) \tag{35} $$

因此，将姿态矩阵 ${}^{t}_{s}\boldsymbol{R}$ 按照式(35)的 *YXZ* 旋转序列进行分解，可得到 γ_s、β_s、α_s 3 个关节角度值。

图 12 为 3 组运动相应的肩胛骨姿态的模型预测结果(γ_{mod}、β_{mod}、α_{mod})与实验测量结果(γ_{meas}、β_{meas}、α_{meas})的对比，从其中可以看出利用本文所提出的机构模型对于肩带运动姿态的描述与实验测量结果吻合较好，其基本趋势一致。

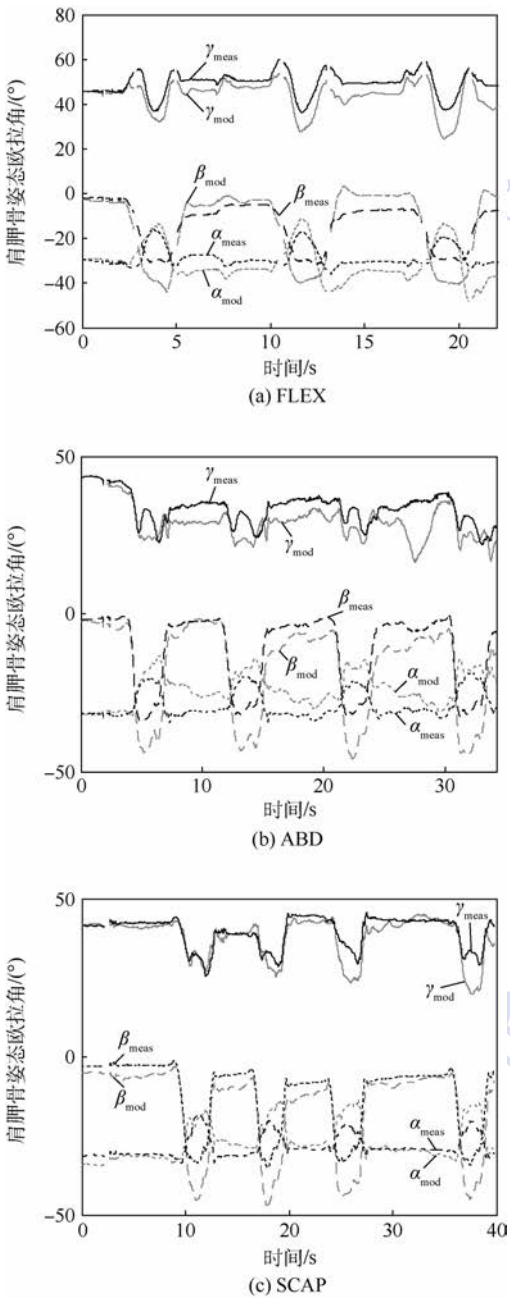


图 12 肩胛骨相对于胸廓的运动姿态(FLEX、ABD、SCAP)

Fig. 12 Movement posture of scapula with respect to thorax (FLEX, ABD, SCAP)

对于每一个运动任务，每一运动时刻，肩胛骨姿态的模型预测值与实验测量值之间的平均偏差为

$$ R = \frac{1}{3}\left(|\gamma_{mod}-\gamma_{meas}| + |\beta_{mod}-\beta_{meas}| + |\alpha_{mod}-\alpha_{meas}|\right) \tag{36} $$

R 值越小，则肩胛骨姿态的模型预测值与实验测量值的相近程度越好。平均偏差的计算结果如图 13 所示，$\overline{R}$为 R 的平均值。

从图 13 中可以看出，对于 3 组运动而言，模型预测的肩胛骨姿态欧拉角与实际测量值之间的偏差较小。最小的$\overline{R}$存在于 SCAP 运动中，为 3.56°，最大的$\overline{R}$存在于 ABD 运动中，仅为 5.71°。

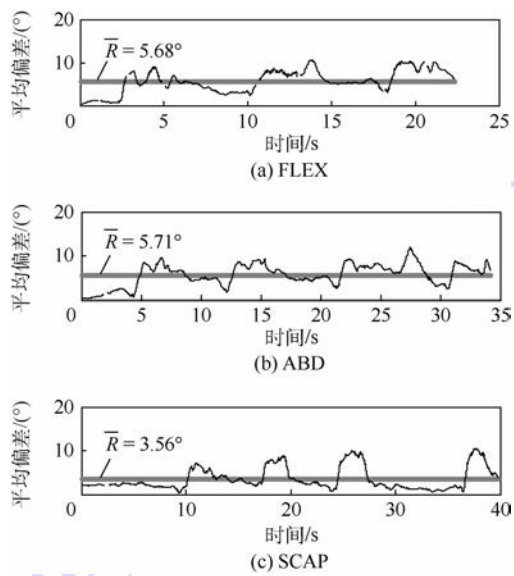


图 13 肩胛骨姿态的模型预测值与实验测量值平均偏差

Fig. 13 Average deviations between scapula postures obtained from model prediction and experiment measurement

6 结 论

1）本文通过将肩胛骨与胸廓的相对运动关系定义为类似于圆柱-平面副的线面接触运动约束，从而将包含胸骨、锁骨、肩胛骨和肱骨在内的肩部骨骼系统抽象成一个 4 杆 6 自由度的混联空间机构模型，并完成了肩带机构的位置建模和封闭解求解。

2）通过骨骼尺寸的缩放，建立了基于个体特征的肩部骨骼机构学模型，为个性化肩部模型的建立和运动分析提供了基础。

3）通过获得自肩部运动实验的数据，验证了

所提出的混联机构模型对于人体肩部各骨骼之间运动约束关系描述的正确性。

参考文献（References）

[1] YANG J, FENG X, KIM J H, et al. Review of biomechanical models for human shoulder complex[J]. International Journal of Human Factors Modelling & Simulation, 2010, 1(3): 271-293.

[2] YANG J J, FENG X, XIANG Y, et al. Determining the three-dimensional relation between the skeletal elements of the human shoulder complex[J]. Journal of Biomechanics, 2009, 42(11): 1762-1767.

[3] DEMPSTER W T. Mechanisms of shoulder movement[J]. Archives of Physical Medicine & Rehabilitation, 1965, 46: 49-70.

[4] ENGIN A E, CHEN S M. Statistical data base for the biomechanical properties of the human shoulder complex—Ⅰ: Kinematics of the shoulder complex[J]. Journal of Biomechanical Engineering, 1986, 108(3): 215-221.

[5] ENGIN A E, TUMER S T. Three-dimensional kinematic modelling of the human shoulder complex—Part Ⅰ: Physical model and determination of joint sinus cones[J]. Journal of Biomechanical Engineering, 1989, 111(2): 107-112.

[6] XU X, DICKERSON C R, LIN J H, et al. Evaluation of regression-based 3-D shoulder rhythms[J]. Journal of Electromyography and Kinesiology, 2015, 17: 28-33.

[7] MASO F D, BLACHE Y, RAISON M, et al. Glenohumeral joint kinematics measured by intracortical pins, reflective markers, and computed tomography: A novel technique to assess acromiohumeral distance[J]. Journal of Electromyography and Kinesiology, 2015, 29: 4-11.

[8] HARISH V, BAKSH A, UNGI T, et al. Measurement of electromagnetic tracking error in a navigated breast surgery setup [C]// Proceedings of the SPIE Medical Imaging. Bellingham: SPIE Press, 2016, 9786: 1-8.

[9] ZHANG C, SKALLI W, LAGAC P Y, et al. Investigation of 3D glenohumeral displacements from 3D reconstruction using biplane X-ray images: Accuracy and reproducibility of the technique and preliminary analysis in rotator cuff tear patients[J]. Landscape Design Journal of the Landscape Institute, 2015, 29(10): 12-20.

[10] DAL M F, RAISON M, LUNDBERG A, et al. Glenohumeral translations during range-of-motion movements, activities of daily living, and sports activities in healthy participants[J]. Clinical Biomechanics, 2015, 30(9): 1002-1007.

[11] GARNER B A, PANDY M G. Musculoskeletal model of the upper limb based on the visible human male dataset[J]. Computer Methods in Biomechanics & Biomedical Engineering 2001, 4(2): 93-126.

[12] MAUREL W, THALMANN D. Human sholder modeling including scapulo-thoracic constraint and joint sinus cones[J]. Computers & Graphics, 2000, 24(2): 203-218.

[13] MAUREL W, THALMANN D, HOFFMEYER P, et al. A biomechanical musculoskeletal model of human upper limb for dynamic simulation[C]// Proceedings of the Eurographics Workshop on Computer Animation and Simulation. Vienna: Springer, 1996: 121-136.

[14] TONDU B. Modelling of the shoulder complex and application the design of upper extremities for humanoid robots[C]// Proceedings of the IEEE-Ras International Conference on Humanoid Robots. Piscataway, NJ: IEEE Press, 2005: 313-320.

[15] BOLSTERLEE B, VEEGER H E J. Modelling clavicular and scapular kinematics: From measurement to simulation[J]. Medical & Biological Engineering & Computing, 2014, 52(3): 283-291.

[16] SHOLUKHA V, JAN S V S. Combined motions of the shoulder joint complex for model-based simulation: Modeling of the shoulder rhythm (ShRm) [M]. London: Springer, 2014: 205-232.

[17] WARNER M B, CHAPPELL P H, STOKES M J. Measurement of dynamic scapular kinematics using an acromion marker cluster to minimize skin movement artifact[J]. Journal of Visualized Experiments Jove, 2015, 96: e51717.

[18] The National Center for Simulation in Rehabilitation Research (NCSRR). Shoulder movements database: The supplement of Delft Shoulder and Elbow Model Project[DS/OL]. (2015-03-20)[2016-05-30]. https://simtk.org/projects/dsem.

[19] WU G, HELM F C T V D, VEEGER H E J, et al. ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—Part Ⅱ: Shoulder, elbow, wrist and hand[J]. Journal of Biomechanics, 2005, 38(5): 981-992.

[20] GARNER B A, PANDY M G. A kinematic model of the upper limb based on the visible human project (VHP) image dataset [J]. Computer Methods in Biomechanics & Biomedical Engineering, 1999, 2(2): 107-124.

作者简介：

聂超　男，硕士研究生。主要研究方向：肩部骨骼肌肉模型、康复训练策略。

宋智斌　男，博士，讲师。主要研究方向：上肢康复机器人、人机交互、生体电信号处理等。

戴建生　男，博士，教授，博士生导师。主要研究方向：变胞机构、可重构机构与可重构机器人、旋量代数与旋量系理论、Origami 机构、机器人与跨学科研究等。

北航学报 赠阅

A shoulder skeletal kinematic model based on spatial hybrid mechanism

NIE Chao[1], SONG Zhibin[1,*], DAI Jiansheng[1,2]

(1. School of Mechanical Engineering, Tianjin University, Tianjin 300350, China;
2. King's College London, University of London, London WC2R 2LS, UK)

Abstract: In order to represent the movement of the shoulder skeletal system, a spatial hybrid mechanism model is proposed, which describes the articulation between the scapula and thorax as a kinematic constraint similar to a cylinder-plane pair. Firstly, types of joints in shoulder are determined, and thus the degrees of freedom of the shoulder girdle and shoulder mechanism can be analyzed. After the definition of local coordinate systems attached to each skeleton, the vector theory method and homogeneous coordinate transformation are used to establish the position analysis equation of the shoulder mechanism, and the closed-form solutions of joint positions are obtained subsequently. Finally, to verify the validity of the mechanism model, the skeletal posture dataset obtained from a shoulder movement experiment is used to inversely drive the model, the calculation results of the scapular posture are compared with the experimental data. The results indicate that the mechanism model has achieved a good consistency in predicting the skeletal movement of the shoulder complex. Furthermore, this model can be adapted for different individuals' geometric skeletal characteristics by scaling.

Keywords: shoulder complex; shoulder girdle; skeletal movement; homogeneous coordinate transformation; spatial hybrid mechanism

北京航空航天大学学报

Received: 2016-12-15; **Accepted**: 2017-01-13; **Published online**: 2017-03-08 09:29
URL: kns.cnki.net/kcms/detail/11.2625.V.20170308.0929.002.html
Foundation items: National Natural Science Foundation of China (51475322, 51535008)
* **Corresponding author**. E-mail: songzhibin@tju.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0017

随机优化的改进交叉熵方法

任超，张航，李洪双*
（南京航空航天大学 航空宇航学院，南京 210016）

摘 要：随机优化的交叉熵方法具有高效性和自适应性的特点，在高维和非线性等复杂优化问题中具有巨大的开发潜力。针对传统交叉熵优化方法精度不足的缺点，提出使用“当前精英样本”和“全局精英样本”构建新的参数更新策略，以充分提取迭代历史中的有用信息。采用自适应的平滑策略和变异操作进一步提升计算性能。通过3个计算实例证明，改进后的方法比传统交叉熵方法具有更高的计算精度和更强的全局搜索能力。

关 键 词：随机优化；交叉熵；精英样本；参数更新策略；自适应平滑策略；变异操作
中图分类号：O224
文献标识码：A **文章编号**：1001-5965(2018)01-0205-10

工程优化设计是以产品结构为基础，根据设计要求、产品功能等，结合现代优化技术，在可行域中搜索最优设计方案，以达到缩短设计周期、提高产品质量和降低生产成本的目的。目前的结构优化设计方法大致可以分为两大类：基于梯度的优化设计方法和随机优化设计方法。前者利用梯度信息从初值点逐渐向最优值点搜索[1]；后者则是研究人员受自然现象、动物行为和人类社会等启发而建立的。但是实际工程中的结构优化设计问题，往往拥有较多的设计变量和约束条件，目标函数和约束条件也比较复杂，此时第一类方法的可行性与准确性受到了极大地考验，而随机优化设计方法则可以很大程度地弥补这一不足。随着计算机性能的不断提高，随机优化设计方法更加受科研人员青睐。

交叉熵(Cross Entropy, CE)方法是一种可靠性分析与随机优化设计的统一方法，其基本概念由 Rubinstein[2] 于 1997 年最先提出并用于模拟稀疏事件。随后 Rubinstein 又将其拓展为随机优化设计方法[3]。之后，国内外众多科研人员对交叉熵方法进行了详细地研究与改进。2001 年，Helvik 和 Wittner[4] 在原有交叉熵方法的基础之上做出改进，使其运用于含有多个相互独立的智能主体网络的路径寻找问题，结果证明交叉熵方法拥有很好的收敛性，并且能够得出理想的优化结果。2005 年，Alon 等[5] 采用交叉熵方法优化缓冲区分配问题，基于交叉熵的优化方法计算效率高，优化结果与理论值的误差不超过 1% 的概率大约为 99%。同年，Kuo 等[6] 采用交叉熵方法处理车辆路径优化问题并得到了令人满意的结果，证明了交叉熵方法在解决这类问题时，对不同的目标函数和约束条件都有很好的效果，并且针对各类具体问题(尤其是车辆路径问题)都能得出很好的优化结果。2006 年，Kroese 等[7] 针对多极值和非线性约束的连续优化问题，提出了改进的交叉熵方法。2010 年，Ho 和 Yang[8] 使用交叉熵优化方法处理电磁场中的逆问题，引入变异操作增强了算法的全局搜索能力。2014 年，丁卫平等[9] 将交叉熵方法与混合蛙跳算法相结合，处理 MRI (Magnetic Resonance Images)图像分割问题，改善

收稿日期：2017-01-12；**录用日期**：2017-05-05；**网络出版时间**：2017-06-19 15:48
网络出版地址：kns.cnki.net/kcms/detail/11.2625.V.20170619.1548.003.html
基金项目：南京航空航天大学研究生创新基地（实验室）开放基金（kfjj20160113）；国家自然科学基金（U1533109）
* **通信作者**. E-mail: hongshuangli@nuaa.edu.cn

引用格式：任超，张航，李洪双. 随机优化的改进交叉熵方法[J]. 北京航空航天大学学报，2018，44(1)：205-214.
REN C，ZHANG H，LI H S. Stochastic optimization method based on improved cross entropy[J]. Journal of Beijing University of Aeronautics and Astronautics，2018，44(1)：205-214 (in Chinese).

北航学报 赠阅

了原始算法的计算精度。同年,李国成和肖庆宪[10]利用交叉熵方法改进蝙蝠算法,降低了原始算法陷入局部最优的概率,并验证了该方法处理高维优化问题的可行性。2015 年,Ghidey[11]对交叉熵方法在可靠性优化设计问题中的精度和效率进行了详细的研究,结果表明相比于解析方法和现有的可靠性优化方法,交叉熵方法在精度和效率上都具有一定的优势。这些研究成果表明将交叉熵方法和其他优化方法相结合,具有很大的发展潜力和应用前景。但总体而言,目前国内外对基于交叉熵的可靠性分析与随机优化设计方法的研究还少见诸于公开发表文献。所以,作为一种新兴的、高效的可靠性分析与随机优化的统一方法,交叉熵方法值得深入探索和研究。本文着眼于交叉熵方法在随机优化问题中的应用,详细分析传统交叉熵方法在处理优化问题时的优缺点,尝试从参数更新策略、平滑参数调整和变异操作 3 个方面作出改进,并利用 2 个数值算例和 1 个工程算例验证改进交叉熵方法的效率、精度和鲁棒性。

1 传统交叉熵方法

交叉熵定义为 2 个概率密度函数比值对数的期望:

$$D(f,h)=\int f_X(\boldsymbol{x})\ln\left(\frac{f_X(\boldsymbol{x})}{h_X(\boldsymbol{x})}\right)\mathrm{d}\boldsymbol{x}=E\left(\ln\left(\frac{f_X(\boldsymbol{x})}{h_X(\boldsymbol{x})}\right)\right) \tag{1}$$

式中:$E(\cdot)$为期望算子;$\boldsymbol{x}$ 为 n 维随机输入变量;$f_X(\boldsymbol{x})$和 $h_X(\boldsymbol{x})$为 2 个概率密度函数。从式(1)可以看出,交叉熵可以反应 2 个概率密度函数之间的差异,当 $f_X(\boldsymbol{x})=h_X(\boldsymbol{x})$时,$D(f,h)=0$。因此交叉熵也被称为 KL 距离(Kullback-Leibler distance)。对于式(2)所示的求最大值问题,即

$$\gamma^*=\max S(\boldsymbol{x})\qquad \boldsymbol{x}^{\mathrm{L}}\leqslant\boldsymbol{x}\leqslant\boldsymbol{x}^{\mathrm{U}} \tag{2}$$

式中:γ^*为优化问题的最优解;$S(\boldsymbol{x})$为目标函数;$\boldsymbol{x}^{\mathrm{U}}$ 和 $\boldsymbol{x}^{\mathrm{L}}$ 分别为设计变量 $\boldsymbol{x}$ 的上下限。当使用交叉熵方法时,需要先构建一个人造可靠性问题:

$$P_{\mathrm{f}}=P(S(\boldsymbol{x})\geqslant\gamma)=\int_{S(\boldsymbol{x})\geqslant\gamma}f_X(\boldsymbol{x};\boldsymbol{v})\mathrm{d}\boldsymbol{x}=E(I_{S(\boldsymbol{x})\geqslant\gamma}) \tag{3}$$

式中:P_{f} 为可靠性问题的失效概率;γ 为一个阈值;$\boldsymbol{v}$ 为未知的分布参数;$I_{S(\boldsymbol{x})\geqslant\gamma}$为示性函数。若令 $\ell(\gamma)=P_{\mathrm{f}}$,则存在 2 类估计问题:①给定一个 γ 求相应的 ℓ;②给定一个 ℓ 求相应的 γ。前者是传统的可靠性分析问题,而后者则是一个优化问题。但是在使用交叉熵方法时,并不是事先指定失效事件及相应的阈值,而是通过 KL 距离去更新概率密度函数 $f_X(\boldsymbol{x};\boldsymbol{v})$,得到一系列的概率密度函数 $f_X(\boldsymbol{x};\boldsymbol{v}_1)$,$f_X(\boldsymbol{x};\boldsymbol{v}_2)$,$\cdots$,$f_X(\boldsymbol{x};\boldsymbol{v}^*)$。在分布参数 $\boldsymbol{v}$ 更新的同时,概率密度函数 f 和变量 γ 也相应地更新,当 $\boldsymbol{v}$ 取得最优值 $\boldsymbol{v}^*$ 时,也得到了概率密度函数 f 的最优值 $f^*=f_X(\boldsymbol{x};\boldsymbol{v}^*)$和参数 γ 的最优值 γ^*。由此可见,最优概率密度函数 f^* 会使得随机变量 $\boldsymbol{x}$ 有极大概率落在最优解 $\boldsymbol{x}^*$ 上或其附近,则由 f^* 生成的任意随机样本都可以看作是最优解的逼近值,而 $\gamma^*=S(\boldsymbol{x}^*)$则为式(2)的最大值。交叉熵方法的核心在于参数 $\boldsymbol{v}$ 的更新,传统的更新策略如下[12]:

1)设第 k 次迭代中,分布参数为 $\boldsymbol{v}_{k-1}$,从 $f_X(\boldsymbol{x};\boldsymbol{v}_{k-1})$中生成 N 个随机样本$\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N\}$,计算目标函数的值$\{S(\boldsymbol{x}_i)\}=\{S(\boldsymbol{x}_1),S(\boldsymbol{x}_2),\cdots,S(\boldsymbol{x}_N)\}$,并将$\{S(\boldsymbol{x}_i)\}$按照从大到小的次序排列,即 $S(\boldsymbol{x}_1)>S(\boldsymbol{x}_2)>\cdots>S(\boldsymbol{x}_N)$。

2)计算目标函数的 ρ 分位数,使得

$$P(S(\boldsymbol{x})<b_k)=\rho \tag{4}$$

式中:b_k 为功能函数值次序统计量的 ρ 分位数,即 $b_k=S(\boldsymbol{x}_{\rho N})$,而且目标函数最优值 γ^* 在第 k 次迭代中的估计值 $\hat{\gamma}_k$ 即为 $S(\boldsymbol{x}_{\rho N})$。

3)第 $k+1$ 次迭代中的分布参数 $\boldsymbol{v}_k$,可以通过式(5)的最大化问题求得

$$\max_{\boldsymbol{v}}D(\boldsymbol{v})=\max_{\boldsymbol{v}}E_{\boldsymbol{v}_{k-1}}(I_{\{S(\boldsymbol{x})>\hat{\gamma}_k\}}\ln f(\boldsymbol{x};\boldsymbol{v}_{k-1})) \tag{5}$$

式(5)对应的随机问题为:对于给定的 $\hat{\gamma}_k$ 和 $\hat{\boldsymbol{v}}_{k-1}$,从式(6)求得第 k 次迭代的分布参数的估计值:

$$\max_{\boldsymbol{v}}\hat{D}(\boldsymbol{v})=\max_{\boldsymbol{v}}\frac{1}{N}\sum_{i=1}^{N}I_{\{S(\boldsymbol{x})>\hat{\gamma}_k\}}\ln f(\boldsymbol{x}_i;\boldsymbol{v}_{k-1}) \tag{6}$$

设从式(6)求得的分布参数为 $\tilde{\boldsymbol{v}}_k$,则第 $k+1$ 次迭代中的分布参数估计值并非直接取 $\tilde{\boldsymbol{v}}_k$,而是引入一个平滑参数 α 构造如式(7)所示的更新策略:

$$\hat{\boldsymbol{v}}_k=\alpha\tilde{\boldsymbol{v}}_k+(1-\alpha)\hat{\boldsymbol{v}}_{k-1} \tag{7}$$

传统交叉熵方法可以高效地处理大部分的简单优化问题,并且具有极快的收敛速度。但是当目标函数较复杂或约束函数数目较多时,传统方法很可能出现计算精度不足或陷入局部最优的情况,因此,本文针对性地提出 3 个改进策略,以提高交叉熵优化方法的计算性能。

2 改进的交叉熵方法

1)改进的更新策略

一般地,大部分的优化问题中,交叉熵方法的

北航学报 赠阅

收敛速度明显高于常用的其他随机优化算法,如遗传算法(Genetic Algorithm, GA)、蚁群优化(Ant Colony Optimization, ACO)算法和粒子群优化(Particle Swarm Optimization,PSO)算法等。交叉熵方法之所以计算速度快,主要是由于更新策略主要是基于 ρN 个"精英样本"的,其中 ρ 为稀薄参数,其取值范围通常为 0.01 ~ 0.1。所以,即使在约束条件较多或者很少有样本点满足约束的情况下,交叉熵方法也可以优先选择最有意义的"精英样本"计算该次迭代中所更新的参数。

同理,由于"精英样本"的数目偏少,并不能完全提取总样本中的有用信息,所以其计算精度是偏低的。另外,由于交叉熵方法的收敛速度过快,所以比较容易陷入局部最优;且如果初期的优化方向错误,则方法可能无法收敛。

本文通过提出"当前精英样本"(current elite samples)和"全局精英样本"(global elite samples)的概念,将参数更新策略改进为

$$\hat{\boldsymbol{v}}_{k+1}=\alpha_1\hat{\boldsymbol{v}}_{c,k}+\alpha_2\hat{\boldsymbol{v}}_{g,k}+\alpha_3\hat{\boldsymbol{v}}_{k-1} \tag{8}$$

式中:$\hat{\boldsymbol{v}}_{c,k}=\{\hat{\boldsymbol{\mu}}_{c,k},\hat{\boldsymbol{\sigma}}^2_{c,k}\}$ 为"当前精英样本"的分布参数,$\hat{\boldsymbol{\mu}}_{c,k}$ 为当前精英样本均值,$\hat{\boldsymbol{\sigma}}^2_{c,k}$ 为当前精英样本方差;$\hat{\boldsymbol{v}}_{g,k}=\{\hat{\boldsymbol{\mu}}_{g,k},\hat{\boldsymbol{\sigma}}^2_{g,k}\}$ 为"全局精英样本"的分布参数,$\hat{\boldsymbol{\mu}}_{g,k}$ 为全局精英样本均值,$\hat{\boldsymbol{\sigma}}^2_{g,k}$ 为全局精英样本方差;α_1、α_2 和 α_3 为新的平滑参数,且 $\alpha_1+\alpha_2+\alpha_3=1$。对于"当前精英样本"和"全局精英样本",可以作如下解释。

"当前精英样本"为第 k 次迭代中使得目标函数值最小的前 ρN 个样本;将第 k 次的"精英样本"与使得目标函数值最小的历史最优值对比,若"当前精英样本"更好,则将其作为第 k 次迭代的"全局精英样本"。"全局精英样本"的初值为第一次迭代的"当前精英样本"。

新更新策略的意义在于:迭代历史中的信息并非都是无用的,尽管新的样本点可能更接近最优,但是历次迭代中的历史样本中也可能存在有用信息。因此,充分利用历次迭代中的所有有用样本将有助于提升算法的计算精度。

2)自适应平滑参数

借鉴只有一个平滑参数的经验,α_3 的取值通常在 0.1 ~ 0.3 之间,α_1 的值要大于 α_2 的值。然而该方法在某些条件下依然可能出现计算结果不理想的情况[13],因此有必要增加一种自适应平滑参数。

仔细分析式(8)可以发现,α_3 的取值越大,则样本保留上一次迭代的趋势越大,在全局搜索的空间更广;相反,α_3 越小,则新生产样本更趋向于最新的"精英样本"靠拢,即更新的速度更快。因此,为了综合全局搜索能力和迭代后期的收敛速度,理想方案是在搜索的初期 α_3 取较大值,随着迭代的进行,α_3 逐渐减小并趋近于预设的最小值。对于 α_1 的取值,本文推荐在 0.5 ~ 0.7 之间。据此,本文提出如下平滑参数自适应调整公式:

$$\begin{cases}\alpha_3=\alpha_{3\max}-\dfrac{\alpha_{3\max}-\alpha_{3\min}}{k_N}k\\ \alpha_2=1-\alpha_1-\alpha_3 \qquad 0.5\leqslant\alpha_1\leqslant 0.7\end{cases} \tag{9}$$

式中:k 为迭代次数;k_N 为预设的总迭代次数;$\alpha_{3\max}$ 和 $\alpha_{3\min}$ 分别为 α_3 的最大值和最小值。

3)变异操作

"变异"是进化计算的基本思想之一,广泛应用于遗传算法、蚁群优化算法和粒子群优化算法等。变异操作的目的在于增加样本的随机性,以降低优化算法陷入局部最优的概率。通常变异操作的方法来对每次迭代过程中更新的参数施加扰动,例如在原有方差的基础上增加一个合适的数。该操作可能使得交叉熵方法的收敛出现明显的波动,但只要扰动量设置合理,就可以在保证计算效率的同时降低算法陷入局部最优的可能性。关于扰动,本文选用了如式(10)所示的扰动函数[14]:

$$\hat{\boldsymbol{\sigma}}^2_{z,k+1}=\hat{\boldsymbol{\sigma}}^2_{k+1}+Z_{k+1} \tag{10}$$

式中:$\hat{\boldsymbol{\sigma}}^2_{k+1}$ 为未施加扰动的第 $k+1$ 次迭代中的样本方差;$\hat{\boldsymbol{\sigma}}^2_{z,k+1}$ 为施加扰动后的样本方差;Z_{k+1} 为扰动函数,可以取一个常值或者是一个关于迭代次数 k 的函数。由于样本方差随迭代次数增加呈递减趋势,因此,扰动函数取一个随迭代次数增加而减小的函数比较合理,本文所采用的扰动函数为

$$Z_k=\max\left(Z-\frac{k}{10},0\right) \tag{11}$$

式中:Z 为一个常数,通常建议取 5 ~ 10。

4)约束处理方法

对于一个有约束的优化问题,通常不能直接对目标函数值进行处理。本文采用"罚函数"法处理有约束问题,当样本点落入不可行域时,对目标函数值进行惩罚。具体方法为利用一个"惩罚函数"对原始目标函数进行增广,增广后的目标函数为

$$S'(\boldsymbol{x})=S(\boldsymbol{x})+\sum_{i=1}^{n_g}h_iD_i(\boldsymbol{x})+\sum_{j=1}^{n_h}h_jD_j(\boldsymbol{x}) \tag{12}$$

式中:n_g 和 n_h 分别为不等式约束和等式约束的数目;h_i 和 h_j 为罚因子,通常罚因子可以取一个很大的常数以有效区分可行域内外的样本点;

北航学报 赠阅

$D_i(\boldsymbol{x})$和$D_j(\boldsymbol{x})$为约束函数值乘子,且有:

$$D_i(\boldsymbol{x})=\begin{cases}0 & g_i(\boldsymbol{x})\leqslant 0\\ |g_i(\boldsymbol{x})| & g_i(\boldsymbol{x})>0\end{cases}\quad 1\leqslant i\leqslant n_g \tag{13}$$

$$D_j(\boldsymbol{x})=\begin{cases}0 & |h_j(\boldsymbol{x})|\leqslant \varepsilon\\ |h_j(\boldsymbol{x})| & |h_j(\boldsymbol{x})|>\varepsilon\end{cases}\quad 1\leqslant j\leqslant n_h \tag{14}$$

其中:ε为一个很小的阈值;$g_i(\boldsymbol{x})$为不等式约束函数;$h_j(\boldsymbol{x})$为等式约束函数。

综上所述,改进交叉熵(Improved Cross Entropy, ICE)优化方法步骤如下:

步骤 1 初始化待定分布参数$\boldsymbol{v}_0=(\boldsymbol{\mu}_0,\boldsymbol{\sigma}_0)$,其中$\boldsymbol{\mu}_0$为初始均值,$\boldsymbol{\sigma}_0$为初始标准差,迭代数记为$k=0$。

步骤 2 从截断正态分布$N(\hat{\boldsymbol{\mu}}_k,\hat{\boldsymbol{\sigma}}_k)$中生成$N$个随机样本$\{\boldsymbol{x}_{1,k},\boldsymbol{x}_{2,k},\cdots,\boldsymbol{x}_{N,k}\}$,计算目标函数的值并按照式(2)进行增广,得到$\{S'(\boldsymbol{x}_{i,k})\}=\{S'(\boldsymbol{x}_{1,k}),S'(\boldsymbol{x}_{2,k}),\cdots,S'(\boldsymbol{x}_{N,k})\}$,并将$\{S'(\boldsymbol{x}_{i,k})\}$按照从小到大的次序排列,即$S'(\boldsymbol{x}_{1,k})<S'(\boldsymbol{x}_{2,k})<\cdots<S'(\boldsymbol{x}_{N,k})$。

步骤 3 根据$b_k=S'(\boldsymbol{x}_{\rho N,k})$确定$b_k$的值,并记录前$\rho N$个随机样本$\{\boldsymbol{x}_{1,k},\boldsymbol{x}_{2,k},\cdots,\boldsymbol{x}_{\rho N,k}\}$。

步骤 4 将$\{S'(\boldsymbol{x}_{1,k}),S'(\boldsymbol{x}_{2,k}),\cdots,S'(\boldsymbol{x}_{\rho N,k})\}$分别和$\{S'(\boldsymbol{x}_{1,k-1}),S'(\boldsymbol{x}_{2,k-1}),\cdots,S'(\boldsymbol{x}_{\rho N,k-1})\}$进行比较,若较小,则用$\boldsymbol{x}_{i,k}$的值替代$\boldsymbol{x}_{i,k-1}$,$i=1,2,\cdots,\rho N$,并将这一组值作为全局精英样本。

步骤 5 根据式(15)和式(16)分别对步骤 3 中“当前精英样本”和步骤 4 中“全局精英样本”的均值和方差,分别记作$\hat{\boldsymbol{v}}_{c,k}=\{\hat{\boldsymbol{\mu}}_{c,k},\hat{\boldsymbol{\sigma}}^2_{c,k}\}$和$\hat{\boldsymbol{v}}_{g,k}=\{\hat{\boldsymbol{\mu}}_{g,k},\hat{\boldsymbol{\sigma}}^2_{g,k}\}$。

$$\hat{\boldsymbol{\mu}}_{k+1}=\sum_{i=1}^{\rho N}\boldsymbol{x}_i/[\rho N] \tag{15}$$

$$\hat{\boldsymbol{\sigma}}^2_{k+1}=\sum_{i=1}^{\rho N}[\boldsymbol{x}_i-\hat{\boldsymbol{\mu}}_{k+1}]^2/[\rho N] \tag{16}$$

步骤 6 根据式(8)和式(9)进行自适应平滑处理。

步骤 7 根据式(10)进行变异操作。

步骤 8 若$\max(\hat{\boldsymbol{\sigma}}_{k+1})\leqslant\varepsilon'$($\varepsilon'$为预设的精度值)或者$k=k_N$,算法停止,$\hat{\boldsymbol{\mu}}_{k+1}$为优化问题的最优解,$S'(\hat{\boldsymbol{\mu}}_{k+1})$为优化问题的最优值;否则,令$k=k+1$,转步骤 2。

3 算 例

1) 算例 1

该算例来自文献[15]的问题 06,求式(17)和式(18)所示的最小化问题:

$$\min S(\boldsymbol{x})=(x_1-10)^3+(x_2-20)^3$$

$$\text{s.t.}\quad\begin{cases}g_1(\boldsymbol{x})=-(x_1-5)^2-(x_2-5)^2+100\leqslant 0\\ g_2(\boldsymbol{x})=(x_1-6)^2+(x_2-5)^2-82.81\leqslant 0\end{cases} \tag{17}$$

式中:$13\leqslant x_1\leqslant 100$,$0\leqslant x_2\leqslant 100$。文献[15]中给出的最优解参考值为$\boldsymbol{x}^*=[14.0950,0.8430]^{\mathrm{T}}$,最优值参考值为$S(\boldsymbol{x}^*)=-6961.8139$。

在本文的交叉熵优化方法的计算过程中,并不是单纯地以精度或迭代次数为终止条件,而是参考文献[15]中的方法,以迭代计算所需的总样本量为终止条件。因实际计算过程中目标函数和约束条件的复杂程度而异,单层样本量的合理值不尽相同,因而迭代次数也有所差异。本文中均以总样本量5×10^5为终止条件。

首先尝试采用传统交叉熵优化方法进行计算,单层样本量为 2 000,迭代次数为 250 次,稀薄参数$\rho=0.01$,平滑参数$\alpha=0.9$,分布参数初值为$\boldsymbol{\mu}_0=[56.5,50]^{\mathrm{T}}$和$\boldsymbol{\sigma}_0=[20,20]^{\mathrm{T}}$。计算结果不甚理想:最优解$\boldsymbol{x}^*=[14.6044,2.1558]^{\mathrm{T}}$,最优值$S(\boldsymbol{x}^*)=-5582.6584$。尽管该最优解满足 2 个约束条件:$g_1(\boldsymbol{x})=-0.3334$,$g_2(\boldsymbol{x})=-0.6853$,但是最优值明显不如文献[15]中的结果。为验证是迭代次数不足还是方法自身缺陷导致的结果不理想,绘制目标函数值变化曲线和最优解变化曲线如图 1 所示。

从图 1 中可以看出,传统交叉熵优化方法的收敛速度很快,这一点与可靠性分析中的情况相吻合。目标函数值在大约 5 次左右已经达到收敛,最优解也在 10 次迭代后收敛。因此实际上有大约超过 200 次的迭代过程是没有意义的。为不失一般性,采用完全随机的抽样方法,重复多次试验,并记录优化计算的结果,分析该方法计算不精确是否为个别现象。图 2 为重复 100 次试验的最优值统计结果。

100 次的重复试验中,最优值的最小值为$\min S(\boldsymbol{x}^*)=-6036.7026$,最大值为$\max S(\boldsymbol{x}^*)=-4646.1624$,中位数为$\text{median}\ S(\boldsymbol{x}^*)=-5331.3439$,方差为$\sigma^2=1.1241\times10^5$。从这一结果可以看出,尽管单次试验中交叉熵优化方法的收敛速度很快,但是其优化结果明显陷入了局部最优,且计算结果不够稳定,因此必须对传统交叉熵方法进行有效的改进。

使用改进的交叉熵方法计算该算例。为了具

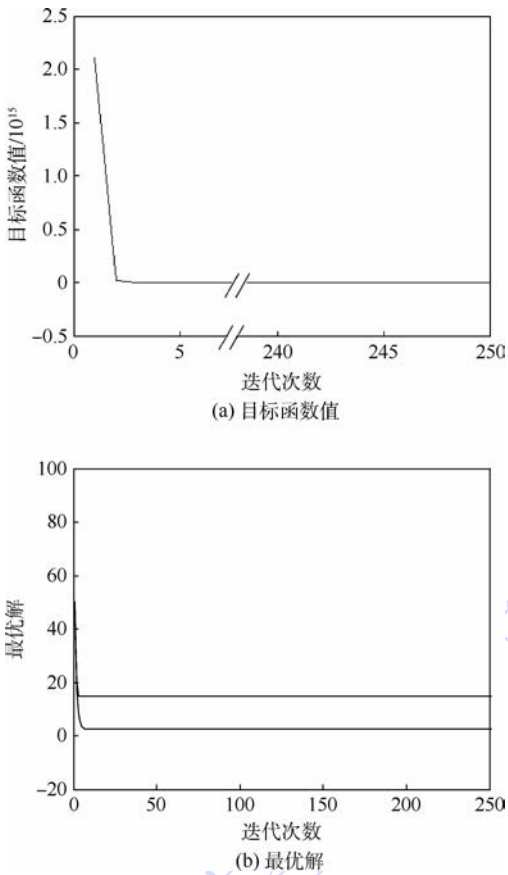


图 1　算例 1 目标函数值和最优解迭代曲线(传统交叉熵方法)

Fig. 1　Iterative curves of objective function value and optimal solution for Example 1(traditional CE method)

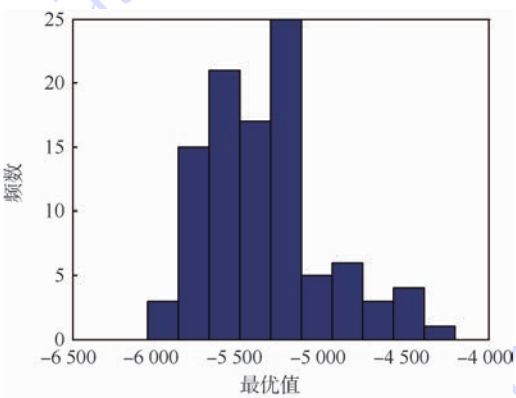


图 2　算例 1 统计结果(传统交叉熵方法)

Fig. 2　Statistical results of Example 1 (traditional CE method)

体分析通过改进更新策略、改进平滑参数、增加变异操作对算法造成的影响,采用依次增加改进策略的方法进行优化计算。第 1 步,增加变异操作,防止陷入局部最优;第 2 步,在第 1 步的基础上引入改进的更新策略和自适应平滑参数,提高计算的精度。

第 1 步:扰动函数为 $Z_k = \max(10 - k/10, 0)$,图 3 为增加变异操作后的目标函数值变化曲线图和最优解变化曲线图。

最优解 $\boldsymbol{x}^* = [14.1558, 0.9785]^{\mathrm{T}}$,最优值 $S(\boldsymbol{x}^*) = -6\,810.4916$。类似的,采用该方法进行 100 次重复试验,最优值统计结果如图 4 所示。

100 次的重复试验中,最优值的最小值为 $\min S(\boldsymbol{x}^*) = -6\,860.4566$,最大值为 $\max S(\boldsymbol{x}^*) = -6\,686.3876$,中位数为 $\text{median } S(\boldsymbol{x}^*) = -6\,775.7076$,方差为 $\sigma^2 = 1.8001 \times 10^3$。从这一

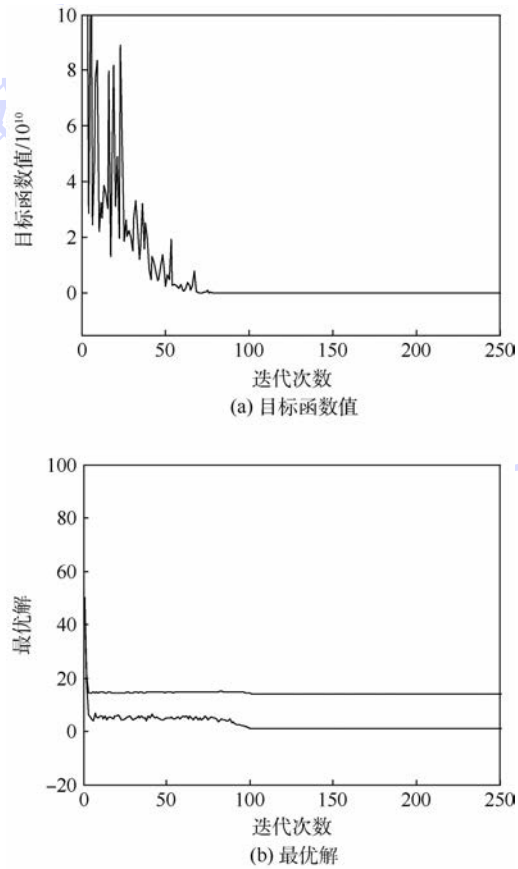


图 3　算例 1 目标函数值和最优解迭代曲线(变异操作)

Fig. 3　Iterative curves of objective function value and optimum solution for Example 1 (with mutation operation)

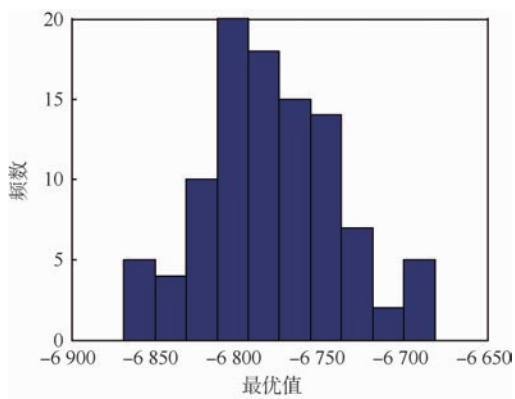


图 4　算例 1 统计结果(变异操作)

Fig. 4　Statistical results of Example 1 (with mutation operation)

结果可以看出,增加了变异操作之后,计算结果跳出了局部最优,与理论值已经十分接近;但是从100 次的统计结果看,增加了变异操作后,交叉熵方法精度依然不是很高,交叉熵方法的稳健性也有待提高,因此有必要进行第 2 步改进。

第 2 步:使用改进的更新策略和自适应的平滑参数,即为完整的改进交叉熵方法。为了和传统交叉熵方对比,设均值和方差的初值点与改进前相同,即:$\boldsymbol{\mu}_0 = [56.5, 50]^{\mathrm{T}}$,标准差初值$\boldsymbol{\sigma}_0 = [20, 20]^{\mathrm{T}}$,稀薄参数和单层样本量保持不变,光滑参数与算例 1 保持一致。使用改进交叉熵方法得到的目标函数值变化曲线图和最优解变化曲线如图 5 所示。

从图 5 可以看出,虽然目标函数值在迭代初期剧烈振荡,但是在 75 次迭代后趋于收敛,最优解则在 100 迭代左右趋于收敛。优化计算的结果为:最优解 $\boldsymbol{x}^* = [14.089\,2, 0.830\,9]^{\mathrm{T}}$,最优值 $S(\boldsymbol{x}^*) = -6\,968.586\,7$,约束函数 $g_1(\boldsymbol{x}) = 0.005\,5$,$g_2(\boldsymbol{x}) = 0.006\,1$。虽然该最优解略微没有满足约束条件,但是无论是最优解还是最优值的计算结果,均与理论值十分接近,因此可以认为,相较

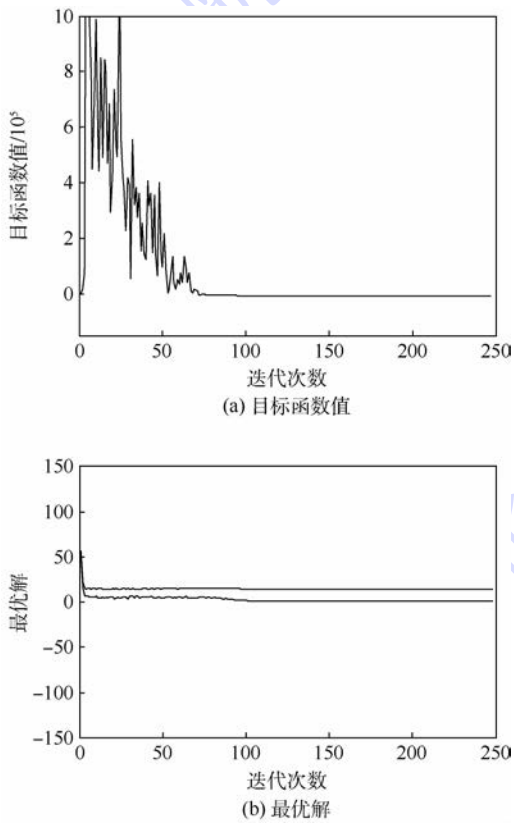


图 5 算例 1 目标函数值和最优解迭代曲线(改进交叉熵方法)

Fig. 5 Iterative curves of objective function value and optimal solution for Example 1 (ICE method)

传统交叉熵方法,改进交叉熵方法不仅可以有效防止算法陷入局部最优,还可以明显提高计算精度。另外,在实际计算的过程中,如果想要加快收敛的速度,只需将式(11)中的参数 Z 适当调小即可,但是注意 Z 取值过小可能造成变异不足。最后,采用改进交叉熵方法进行 100 次重复试验,结果如图 6 所示。

从图 6 可以看出改进后的交叉熵方法鲁棒性明显提高,绝大部分的最优值计算结果都十分接近理论值。

为了进一步显示改进方法的优越性,尝试从重要抽样密度均值的初值 $\boldsymbol{\mu}_0$ 入手,将 $\boldsymbol{\mu}_0$ 设置在 $\boldsymbol{x}$ 的约束范围之外,以检测该方法的通用性。表 1 为 $\boldsymbol{\mu}_0$ 取不同值时改进前后的优化计算结果,$S_{\mathrm{CE}}(\boldsymbol{x}^*)$ 和 $S_{\mathrm{ICE}}(\boldsymbol{x}^*)$ 分别为采用传统交叉熵方法和改进交叉熵方法求得的目标函数值。

从表 1 可以看出,当均值的初值点处于感兴趣的区域外时,传统交叉熵方法出现了明显的错误,这种错误并不是由迭代次数不足引起的,而是寻优方向出现了错误导致算法没有收敛;相反,改进交叉熵方法不论初值点是否在感兴趣的区域,最终都能够得到十分精确的结果,证明改进交叉熵方法鲁棒性更强。

2) 算例 2

该算例来自文献[15]的问题 16,由于该问题过于冗长,对于该问题的描述详见文献[15]。该

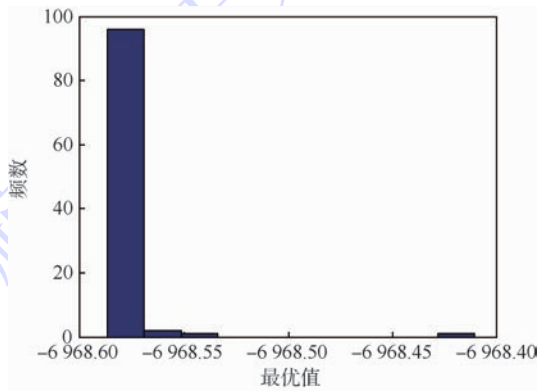


图 6 算例 1 统计结果(改进交叉熵方法)

Fig. 6 Statistical results of Example 1 (ICE method)

表 1 不同初值改进前后方法优化结果

Table 1 Optimal solutions of method before and after improvement with different initial values

$\boldsymbol{\mu}_0$	$S_{\mathrm{CE}}(\boldsymbol{x}^*)$	$S_{\mathrm{ICE}}(\boldsymbol{x}^*)$
$[56.5, 50]^{\mathrm{T}}$	$-5\,582.658\,4$	$-6\,968.586\,7$
$[100, 100]^{\mathrm{T}}$	$1.946\,6 \times 10^{14}$	$-6\,968.586\,8$
$[0, 0]^{\mathrm{T}}$	$-6\,289.121\,0$	$-6\,968.586\,6$
$[150, 150]^{\mathrm{T}}$	$1.128\,9 \times 10^{16}$	$-6\,968.586\,7$
$[-150, 150]^{\mathrm{T}}$	$1.161\,8 \times 10^{14}$	$-6\,968.586\,8$

问题一共有 5 个输入变量,38 个不等式约束,是一个约束情况较为复杂的优化算例。文献[15]中给出的最优值参考值为 $S(\boldsymbol{x}^*) = -1.905\,155\,26$,最优解为 $\boldsymbol{x}^* = [705.175, 68.600, 102.900, 282.325, 37.584]^{\mathrm{T}}$。

分别采用改进前后的交叉熵方法进行 100 次重复试验。为了方便对比,同时防止改进前的传统交叉熵方法出现找不到最优解的情况发生,将重要抽样密度均值设在感兴趣的区域之内,即 $\boldsymbol{\mu}_0 = [800,150,60,250,55]^{\mathrm{T}}$。标准差初值 $\boldsymbol{\sigma}_0 = [30,30,30,30,30]^{\mathrm{T}}$,单层样本量 $N = 2\,500$,迭代次数为200 次,稀薄参数 $\rho = 0.01$。传统交叉熵方法光滑参数 $\alpha = 0.8$,改进交叉熵方法参数设定与算例 1 相同。

取其中一次的试验结果绘制目标函数值和最优解迭代曲线如图 7 所示。

采用传统交叉熵方法,结果依然不是非常理想,目标函数最优值 $S(\boldsymbol{x}^*) = -1.387\,7$,最优解为 $\boldsymbol{x}^* = [854.724, 78.267, 93.233, 266.129, 63.013]^{\mathrm{T}}$。

改进交叉熵方法,结果与理论值十分接近,目标函数最优值 $S(\boldsymbol{x}^*) = -1.905\,2$,最优解为 $\boldsymbol{x}^* = [705.175, 68.600, 102.900, 282.325, 37.584]^{\mathrm{T}}$。值得注意的是,本算例中目标函数的变化曲线并不像算例 1 中波动明显,这是由于本算例中目标函数值的波动幅度过大,因此这里舍去了部分目标函数值极大的点,仅画出波动较明显的区域。其中一次结果的目标函数值和最优解迭代曲线如图 8 所示。表 2 给出了改进前后的交叉熵方法目标函数值收敛情况。

从表 2 可以看出,传统交叉熵方法在第 10 次左右便已经接近收敛,并且陷入了局部最优;而改进交叉熵方法虽然收敛速度稍慢,但是却可以逐渐找到全局最优解。

由于一次试验的结果具有一定的随机性,因此绘制 100 次重复试验的结果统计对比如图 9 所示。

100 次的重复试验中,传统交叉熵方法目标函数最小值为 $\min S_{\mathrm{CE}}(\boldsymbol{x}^*) = -1.617\,5$,最大值为 $\max S_{\mathrm{CE}}(\boldsymbol{x}^*) = -1.080\,6$,中位数为 $\mathrm{median}\, S_{\mathrm{CE}}(\boldsymbol{x}^*) = -1.325\,5$,方差为 $\sigma^2_{\mathrm{CE}} = 0.023\,4$。改进交叉熵方法目标函数最小值为 $\min S_{\mathrm{ICE}}(\boldsymbol{x}^*) =$

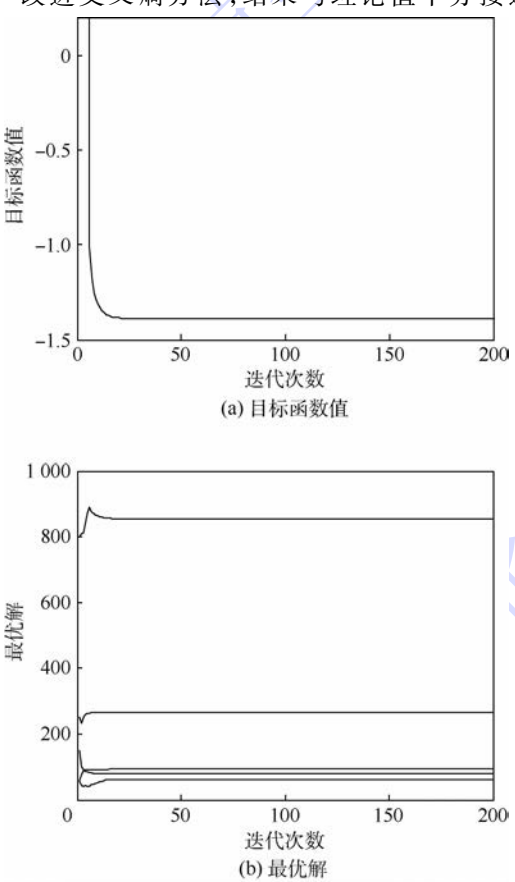


图 7　算例 2 目标函数值和最优解迭代曲线(传统交叉熵方法)

Fig. 7　Iterative curves of objective function value and optimal solution for Example 2(traditional CE method)

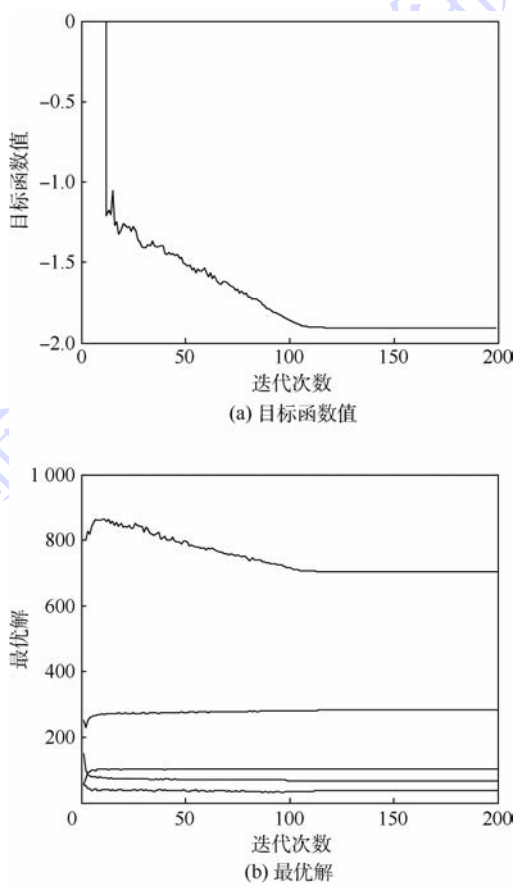


图 8　算例 2 目标函数值和最优解迭代曲线(改进交叉熵方法)

Fig. 8　Iterative curves of objective function value and optimal solution for Example 2 (ICE method)

北航学报 赠阅

$-1.905\,20$，最大值为 $\max S_{\mathrm{ICE}}(\boldsymbol{x}^*) = -1.904\,97$，中位数为 $\mathrm{median}\ S_{\mathrm{ICE}}(\boldsymbol{x}^*) = -1.905\,16$，方差为 $\sigma^2_{\mathrm{ICE}} = 1.8415\times10^{-9}$。对比改进前后的统计结果，可以明显看出改进交叉熵方法有更大的概率求得接近理论值的解。不仅如此，改进交叉熵方法比改进前的结果拥有更好的鲁棒性，因此可以认为在处理含有多个不等式约束的优化问题时，改进的交叉熵方法在计算的精确性，鲁棒性方面都具有明显的优势。

表 2 传统交叉熵和改进交叉熵方法目标函数值变化过程

Table 2 Comparison of objective function value in iteration history between traditional CE and ICE methods

迭代次数	$S_{\mathrm{CE}}(\boldsymbol{x}^*)$	$S_{\mathrm{ICE}}(\boldsymbol{x}^*)$
1	$3.046\,0\times10^{17}$	$1.870\,5\times10^{17}$
2	$1.885\,2\times10^{16}$	$2.409\,6\times10^{16}$
3	$4.116\,8\times10^{15}$	$5.599\,3\times10^{15}$
4	$8.374\,1\times10^{14}$	$3.043\,2\times10^{15}$
5	$1.273\,0\times10^{14}$	$9.092\,1\times10^{14}$
6	$-1.002\,9$	$3.657\,2\times10^{14}$
9	$-1.141\,5$	$1.002\,9\times10^{13}$
10	$-1.312\,0$	$-0.893\,3$
50	$-1.387\,7$	$-1.508\,1$
100	$-1.387\,7$	$-1.854\,4$
150	$-1.387\,7$	$-1.905\,2$
200	$-1.387\,7$	$-1.905\,2$

3）算例 3

该算例来自文献[16]，是一个传动箱优化设计问题。该问题的目标函数和约束函数为

$$\begin{aligned}\min\ & S(\boldsymbol{x}) = (0.020\,4x_1x_4 + 0.060\,7x_1x_4x_5^2)(x_1 + x_2 + x_3) + (0.018\,7x_2x_3 + 0.043\,7x_2x_3x_6^2)(x_1 + 1.57x_2 + x_4)\\ \text{s.t.}\ & \begin{cases} g_1(\boldsymbol{x}) = 2.07 - 0.001x_1x_2x_3x_4x_5x_6 \leqslant 0\\ g_2(\boldsymbol{x}) = -1 + 0.000\,62x_1x_4x_5^2(x_1 + x_2 + x_3) + 0.000\,58x_2x_3x_6^2(x_1 + 1.57x_2 + x_4) \leqslant 0 \end{cases}\end{aligned} \tag{18}$$

式中：$x_i \geqslant 0, i = 1,2,\cdots,6$。文献[16]中给出的目标函数参考值为 $S(\boldsymbol{x}^*) = 135.076\,0$，使用传统交叉熵方法和改进交叉熵方法计算出的目标函数值分别为：$S_{\mathrm{CE}}(\boldsymbol{x}^*) = 136.876\,2$ 和 $S_{\mathrm{ICE}}(\boldsymbol{x}^*) = 135.279\,9$，可以看出改进交叉熵方法结果更接近参考值。为了进一步衡量改进前后方法的鲁棒性，做 100 次重复试验，其结果统计对比如图 10 所示。

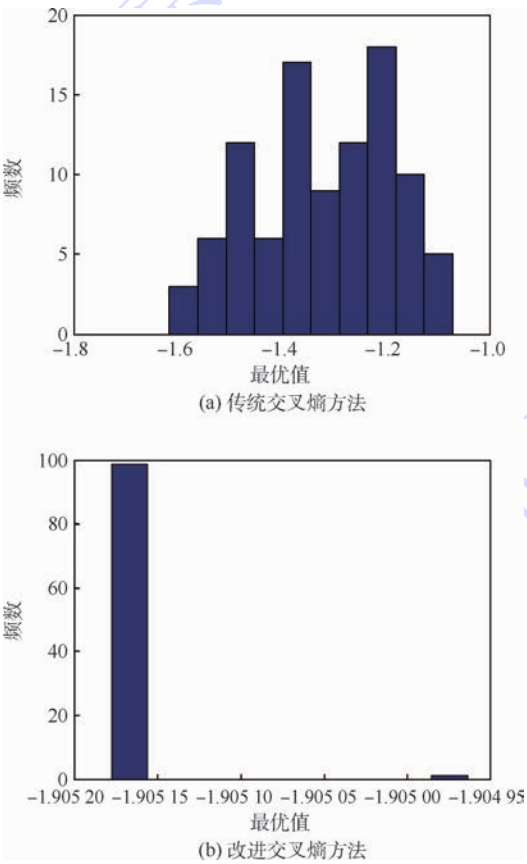


图 9 算例 2 统计结果

Fig. 9 Statistical results of Example 2

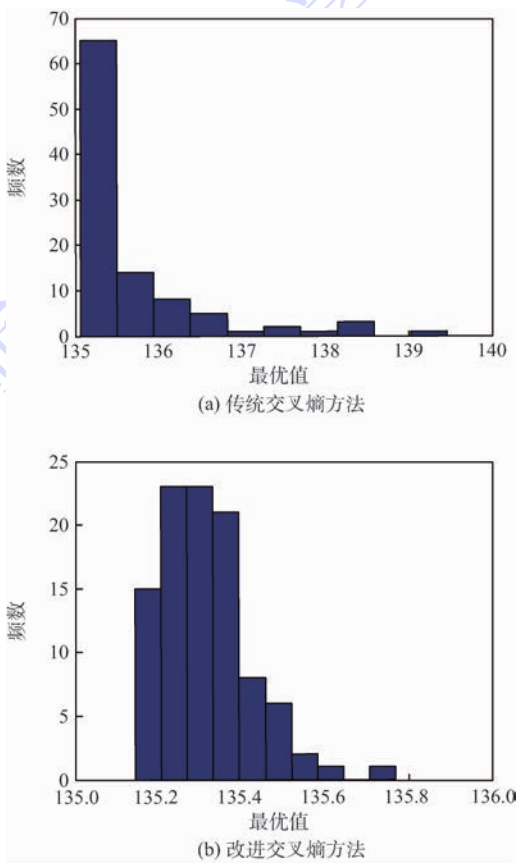


图 10 算例 3 统计结果

Fig. 10 Statistical results of Example 3

从图 10 可以看出，传统交叉熵方法得到的目标函数值最大值超过了 139，而改进交叉熵方法最大值不超过 136，因此可以认为改进交叉熵方法具有更好的鲁棒性。同时注意到即使是改进交叉熵方法，其计算结果和理论值仍存在少许误差，因此将来仍有改进的空间。

4　结　论

本文针对传统交叉熵方法的缺陷，提出了一种高效、高鲁棒性的改进交叉熵方法，可以得出以下结论：

1）提出一种增加“全局精英样本”的方法，适当降低传统交叉熵方法的计算速度，保留迭代历史中的有用信息并应用于参数的更新中，可提出计算的精确度。

2）引入变异操作，使得改进交叉熵方法在其迭代初期尽可能在更大的范围类搜索可行解，在迭代后期逐渐趋于最优解。

3）改进交叉熵方法，保留了传统交叉熵方法一定的高效性，提高了精确性，降低了陷入局部最优的可能性。

4）通过多次的重复试验，证明改进交叉熵方法鲁棒性更高，尤其是对重要抽样密度函数均值的初值点不敏感，因此可以在不同的初值条件、不同的随机种子的前提下都尽可能趋于一个最优解。

参考文献（References）

[1] HAFTKA R T, GÜRDAL Z. Elements of structural optimization [M]. Dordrecht: Kluwer Academic Publishers, 1990.

[2] RUBINSTEIN R Y. Optimization of computer simulation models with rare events[J]. European Journal of Operational Research, 1997, 99(1): 89-112.

[3] RUBINSTEIN R Y. The cross-entropy method for combinatorial and continuous optimization[J]. Methodology & Computing in Applied Probability, 1999, 1(2): 127-190.

[4] HELVIK B E, WITTNER O. Using the cross-entropy method to guide/govern mobile agent's path finding in networks[M]. Berlin: Springer, 2001.

[5] ALON G, KROESE D P, RAVIV T, et al. Application of the cross-entropy method to the buffer allocation problem in a simulation-based environment[J]. Annals of Operations Research, 2005, 134(1): 137-151.

[6] KUO C, YEARGAIN J R, DOWNEY W J, et al. Solving the vehicle routing problem with stochastic demands using the cross-entropy method[J]. Annals of Operations Research, 2005, 134(1): 153-181.

[7] KROESE D P, POROTSKY S, RUBINSTEIN R Y. The cross-entropy method for continuous multi-extremal optimization[J]. Methodology & Computing in Applied Probability, 2006, 8(3): 383-407.

[8] HO S L, YANG S. The cross-entropy method and its application to inverse problems[J]. IEEE Transactions on Magnetics, 2010, 46(8): 3401-3404.

[9] 丁卫平，王建东，陈森博，等. 基于改进混合蛙跳算法的粗糙属性交叉熵优化约简[J]. 南京大学学报(自然科学), 2014, 50(2): 159-166.
DING W P, WANG J D, CHEN S B, et al. Rough attribute reduction with cross-entropy based on improved shuffled frog-leaping algorithm[J]. Journal of Nanjing University (Natural Sciences), 2014, 50(2): 159-166 (in Chinese).

[10] 李国成，肖庆宪. 求解高维函数优化问题的交叉熵蝙蝠算法[J]. 计算机工程, 2014, 40(10): 168-174.
LI G C, XIAO Q X. Cross-entropy bat algorithm for solving high-dimensional function optimization problem[J]. Computer Engineering, 2014, 40(10): 168-174 (in Chinese).

[11] GHIDEY H. Reliability-based design optimization with cross-entropy method[D]. Trondheim: Norwegian University of Science and Technology, 2015.

[12] 李洪双，马远卓. 结构可靠性分析与随机优化设计的统一方法[M]. 北京：国防工业出版社, 2015: 146-147.
LI H S, MA Y Z. Unified methods for structural reliability analysis and stochastic optimization design[M]. Beijing: National Defense Industry Press, 2015: 146-147 (in Chinese).

[13] HUI K P, BEAN N, KRAETZL M, et al. The tree cut and merge algorithm for estimation of network reliability[J]. Probability in the Engineering & Informational Sciences, 2002, 17(1): 23-45.

[14] 吴烈. 电磁场逆问题鲁棒优化设计理论和算法研究[D]. 杭州：浙江大学, 2012.
WU L. Numerical methodologies for robust optimizations of inverse problems[D]. Hangzhou: Zhejiang University, 2012 (in Chinese).

[15] LIANG J J, RUNARSSON T P, MEZURA-MONTES E, et al. Problem definitions and evaluation criteria for the CEC 2006 special session on constrained real-parameter optimization[R]. Singapore: Nanyang Technological University, 2006.

[16] 彭宏，杨立洪，郑咸义，等. 计算工程优化问题的进化策略[J]. 华南理工大学学报(自然科学版), 1997, 25(12) 17-21.
PENG H, YANG L H, ZHENG X Y, et al. A new evolutionary strategy for solving engineering optimization problems[J]. Journal of South China University of Technology (Natural Science Edition), 1997, 25(12): 17-21 (in Chinese).

作者简介：

任超　男，硕士研究生。主要研究方向：结构可靠性设计、灵敏度分析、随机优化方法。

张航　男，硕士研究生。主要研究方向：结构可靠性设计、灵敏度分析、随机优化方法。

李洪双　男，博士，副教授，硕士生导师。主要研究方向：飞行器可靠性设计、飞行器结构优化设计。

北航学报 赠阅

Stochastic optimization method based on improved cross entropy

REN Chao, ZHANG Hang, LI Hongshuang*

(College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Cross entropy method is an efficient and adaptive stochastic optimization method and has immense potential in complex optimization problems with high dimension and nonlinear constraints. However, the traditional cross entropy method is lack of accuracy. In this study, both the concepts of current elite samples and global elite samples are introduced to extract more useful information from the whole iterative history. Then, a new parameter updating strategy is established based on these two concepts. New adaptive smoothing strategy and mutation operation are also applied to improve its computing performance. The proposed algorithm is illustrated by three numerical examples. The computational results indicate that the improved cross entropy method has higher calculation accuracy and better global search capability.

Keywords: stochastic optimization; cross entropy; elite samples; parameter updating strategy; adaptive smoothing strategy; mutation operation

北京航空航天大学学报

Received: 2017-01-12; **Accepted**: 2017-05-05; **Published online**: 2017-06-19 15:48

URL: kns.cnki.net/kcms/detail/11.2625.V.20170619.1548.003.html

Foundation items: Foundation of Graduate Innovation Center in NUAA (kfjj20160113); National Natural Science Foundation of China (U1533109)

* **Corresponding author**. E-mail: hongshuangli@nuaa.edu.cn

《北京航空航天大学学报》征稿简则

北航学报
赠 阅

《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(作者顺序始终不得更改),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,引言,正文,参考文献,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。首、末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:100083 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 44 卷 第 1 期　2018 年 1 月

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　100083
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

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUAA)
(Monthly, Started in 1956)
Vol.44 No.1　January 2018

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 100083, 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　50.00 元／期

ISSN 1001-5965
9 771001 596182 01>