

# 视觉 SLAM 中的视觉里程计

陈浩强<sup>1</sup>

(1. 国防科技大学 智能科学学院, 长沙 410073; )

**摘 要:** 在视觉 SLAM 中, 有一个极其重要的组成部分: 视觉里程计。其作为视觉 SLAM 的前端, 是机器人建图与定位的基础所在。本文主要介绍视觉 SLAM 中视觉里程计的原理和建模过程, 通过对视觉里程计的建模, 理解视觉 SLAM 中后端优化部分的工作目标之所在。进而通过整个视觉 SLAM 原理的阐述, 对 SLAM 这个机器人领域中的分支有进一步的理解, 对 SLAM 的基本问题及关键技术环节有一个清醒的认识。===== 在视觉 SLAM 中, 有一个极其重要的组成部分: 视觉里程计。其作为视觉 SLAM 的前端, 是机器人建图与定位的基础所在。本文主要介绍视觉 SLAM 中视觉里程计的原理和建模过程, 通过对视觉里程计的建模, 理解视觉 SLAM 中后端优化部分的工作目标之所在。进而通过整个视觉 SLAM 原理的阐述, 对 SLAM 这个机器人领域中的分支有进一步的理解, 对 SLAM 的基本问题及关键技术环节有一个清楚的认识。

**关 键 词:** 视觉 SLAM; 视觉里程计; 最小二乘法

**中图分类号:** V221+.3; TB553

**文献标识码:** A

**文章编号:** 1001-5965 (XXXX) XX-XXXX-XX

## 1 引言

Simultaneous Localization And Mapping, 简称 SLAM, 译作“同时定位与建图”, 是近年来机器人领域中发展起来的技术分支, 它指的是通过搭载特定的传感器, 在没有先验环境信息的情况下, 于运动中建立环境的模型并同时估计自己的运动状态。本文所讲的视觉 SLAM 就是将相机作为传感器进行建图与定位的技术。视觉 SLAM 这项技术可以四部分组成视觉<sup>[1]</sup>, 如图 1 为经典视觉 SLAM 工作流程框架, 整个视觉 SLAM 具体工作过程如下 [1]:

1. **传感器数据的读取。**这是将外界环境抽象为数据的过程, 也是计算机理解世界的第一步, 后续过程其实就是对数据的处理过程, 所以在这个阶段数据采集的好坏直接影响到后续工作数量与质量。也因此, 选择质量较好的传感器通常是最为明智的选择, 因为这能大大减少数据处理的工作量和难度。

2. **视觉里程计。**也称为视觉 SLAM 的前端。这是本文主要介绍的部分。视觉里程计的任务是估计相邻两幅图像之间的运动以及局部地图的样子。通过一系列的估计, 可形成对机器人在一段时

间的连续运动的轨迹。

3. **非线性优化。**也称为后端, 是对前端处理结果的优化。传感器产生的数据不可避免地含有噪声, 因此, 前端在不同时刻获得的位姿估计, 受到不同程度的大小的噪声影响, 这不可避免地产生不一致问题, 即如果将不同时刻的位姿拼接起来, 可能出现突变的现象。后端要处理的问题就是对接接收的多个位姿信息进行优化, 从而得到全局一致的轨迹和地图。除此之外, 后端还会对会还检测的信息进行处理。

4. **回环检测。**简单地说, 回环检测就是让机器人检测出自己是否曾经到达这个地方, 如果曾经到达, 则会将信息传给后端进行处理。

5. **建图。**根据估计得轨迹, 建立与任务对应的环境地图。

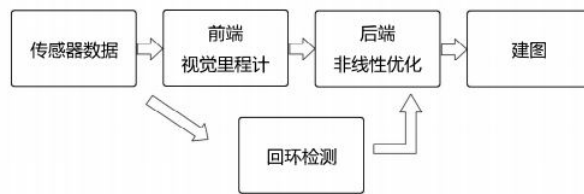


图 1 视觉 SLAM 工作流程图

Fig. 1 Workflow of Visual SLAM

以上是经典的视觉 SLAM 工作流程框架，其在过去的几十年间已有较为丰硕的研究成果，这个框架所包含的研究成果基本定型。在现有的开源机器人程序库中，完全可以依靠这些现有算法搭建一个 SLAM 系统。由于本文主要研究视觉里程计，所以，以下对视觉里程计进行详细的介绍。

视觉里程计关心的是相邻图像间的相机的运动，也即是如何通过相邻的两幅图像估计相机的位姿变换情况。在现实世界中，人眼能够很轻易地通过视觉中镜像的变换感受出自己身体的运动在状态 [2]，在图 2 中，我们人眼能够很轻易的通过图像从左到右的变化猜测出人眼向左旋转或者移动了，因为左边的景物明显变多了。但是如果进一步深究，人眼移动了多少，这样一个量化的问题是难以回答的，因为我们的直觉对数字并不敏感。所以，计算机相比我们，其优势就是可量化地表示出这种量化。在计算机视觉中，图像的表达方式是一个数值矩阵，如何通过一系列的数值矩阵估计相机的运动，就是视觉里程计的本质工作。但是，如果仅仅通过视觉里程计来估计相机的运动轨迹，难免会出现漂移，因为视觉里程计是独立地估计两幅间的运动，每次估计只与上一幅图像相关，所以就会出现误差累计的现象。这导致经过一段时间后，估计的轨迹不再准确。解决这个问题就是后端检测的任务，在此不细讲。



图 2 运动产生的图片变化

Fig. 2 Workflow of Visual SLAM

## 2 在 SLAM 问题中为什么需要视觉里程计

理解这个问题需要对 SLAM 过程进行数学语言的描述。根据文献 [3] 介绍，假设环境有许多路标 (landmark) 组成，将机器人连续运动变成离散

的时刻:  $t=1, 2, \dots, K$ 。图 3 为机器人在某一时刻的运动及工作状态，其中五角星表示路标，虚线的三角形表示里程计估计的机器人的位置 (称为位置 1)，淡实线表示机器人根据对 landmark 的观测确定的机器人位置 (称为位置 2)，深实线表示机器人实际所在的位置 (称为位置 3)。假设机器人携带一个测量自身运动信息的传感器，通过这个传感器可以测量自身运动的里程信息，这个传感器可以是码盘，IMU 或者相机 (此时则是视觉里程计)，但无论这个传感器是什么，我们都可以将机器人的运动状态表示如下：

$$sx_k = S(x_{k-1}, u_k, w_k) \quad (1)$$

这里  $u_k$  是  $k$  时刻传感器的读数,  $w_k$  是  $k$  时刻的噪声,  $u_{k-1}$  是上一时刻机器人所在的位置,  $u_k$  是机器人当前所在位置。机器人运动的原点为机器人开机后机器人所在的位置。式 1 也称为机器人的运动方程。与运动方程相对的，称为观测方程：

$$z_{k,j} = O(y_j, x_k, v_{k,j}) \quad (2)$$

式中  $y_j$  为机器人在  $x_k$  位置观测到的路标,  $z_{k,j}$  为  $k$  时刻，根据该路标产生的观测数据,  $v_{k,j}$  代表噪声。因为该方程是根据外界观测信息对机器人自身的信息 (比如位置信息) 进行估计，所以式 2 称为观测方程。

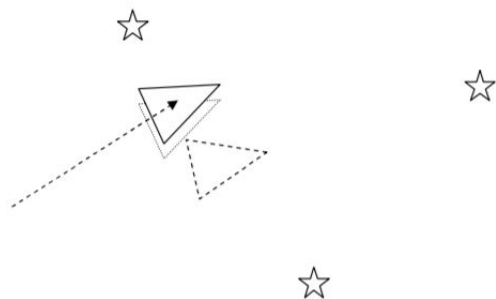


图 3 机器人自我定位原理

Fig. 3 Workflow of Visual SLAM

结合式 1 2 以及图 3, 机器人根据运动方程计算出机器人在  $k$  时刻处于位置 1，然后又根据观测方程观测自己位于位置二。可以看出，运动方程对机器人自身位置的估计准确度比观测方程低。但是这是不是意味这里里程计信息是多余的呢？将上面的分析过程进一步拓展，令  $x_k = [x, y, \theta]_k^T$ ，即机器人的位姿。  $u_k = [\delta x, \delta y, \delta \theta]$ ，代入式 1

$$\begin{bmatrix} x \\ y \\ \theta \end{bmatrix}_k = \begin{bmatrix} x \\ y \\ \theta \end{bmatrix}_{k-1} + \begin{bmatrix} \delta x \\ \delta y \\ \delta \theta \end{bmatrix}_k + w_k \quad (3)$$

对于观测方程, 我们假设路标  $y_j = [p_x, p_y]^T$ , 观测数据  $z_{k,j} = [r, \theta]^T$ , 其中  $r$  为机器人相对路标的距离和夹角。可将观测方程写成:

$$\begin{bmatrix} r \\ \theta \end{bmatrix} = \begin{bmatrix} \sqrt{(p_x - x)^2 + (p_y - y)^2} \\ \arctan \frac{p_y - y}{p_x - x} \end{bmatrix} + v \quad (4)$$

以上两个方程描述了 SLAM 的基本过程, 由于使用的传感器不同, 具体方程的具体形式根据实际情况变化而变化。根据以上两个方程可以解决 SLAM 的基本问题: 定位与建图, 事实上, 定位问题指的就是估计机器人位姿  $x$  的问题, 建图问题则是对路标  $y$  进行估计的问题。这样, SLAM 问题其实就是对  $x, y$  的在噪声中的状态估计问题。而状态估计问题的求解, 根据噪声是否服从高斯分布、运动和观测方程是否线性, 分为高斯/非高斯和线性/非线性系统 [4], 所以有如下四种 SLAM 系统:

表 1 SLAM 系统类型

Table 1 Type of SLAM

编号	类型
1	线性高斯
2	非线性非高斯

线性高斯系统是最简单的系统, 其最优估计可由卡尔曼滤波 (Kalman Filter, KF) 给出, 而非线性非高斯系统则是最复杂的系统, 其代表的求解方法有拓展卡尔曼滤波 (Extend Kalman Filter, EKF) 和非线性优化 [1]。由于现实环境基本多是非线性非高斯, 于机器人鲁棒性而言, 将 SLAM 系统设计成非线性非高斯的, 明显是更优的。事实上, 最早的实时视觉 SLAM 系统则是基于 EKF 开发的, 但是由于 EKF 假设噪声分布是高斯分布, 误差是线性化的, 所以 EKF 还不太符合现实环境的特点。而后人们开始使用粒子滤波器, 随着计算机能力以及非线性优化技术的提高, 当前主流的 SLAM 基本偏向于使用非线性优化技术。

至此, 还未回答本节的目标问题: 为什么 SLAM 系统需要视觉里程计? 似乎根本不需要运

动方程, 我们就可以根据观测方程得到机器人的位置。事实上, 一方面, 由于 SLAM 问题是一个状态估计问题, 所以机器人其实只是“猜测”自己实际所在的位置以及路标的位置, 在观测方程中, 需要根据里程计估计出来的  $x$  对观测数据进行计算, 所以里程计能够辅助机器人更快更准确地计算出自己观测到的数据; 另一方面, SLAM 除了定位之外, 还有建图的任务, 而状态估计出的两个变量  $x$  和  $y$ , 是建图基本的两个参数。基于以上两个原因, 视觉里程计能为视觉 SLAM 这种有一定计算量而又有实时性要求的技术提供很大的帮助。

## 3 正文

### 3.1 量、单位、公式

#### 3.1.1 公式编排

《北京航空航天大学学报》一般不编排单独的符号表, 对于公式中的变量含义需要说明的, 请在公式后的段落中, 采用“式中:  $A$  为某某;  $B$  为某某; ……”的方式加以说明。

$$p_1(h) = \frac{n_{\text{He}}RT}{V} - \rho_{\text{He}}gh \quad (5)$$

式中:  $n_{\text{He}}$  和  $\rho_{\text{He}}$  分别为艇囊内部氦气的物质的量和氦气在温度  $T$  时的平均密度;  $V=36\,893.426\,\text{m}^3$  为艇囊体积;  $T=216.65\,\text{K}$  为艇囊内稳态温度;  $h$  为距离艇囊中心轴线 ( $x$  轴) 的垂直高度. 请使用 Mathtype 编辑。公式中字体的定义尺寸为 10 磅, 上标/下标 68%, 次下标/上标 42%, 符号 150%, 次符号 100%(设置方法: Mathtype-尺寸-定义)。长公式如需转行, 应在记号  $=$ ,  $+$ ,  $-$  等之后断开, 而在下一行开头不再重复这一记号。

#### 3.1.2 量和单位

有关记号的使用应符合国家标准, 例如:  $\sin^{-1}$  应为  $\arcsin$ ,  $\text{ctg}$  应为  $\cot$ ,  $\text{tg}$  应为  $\tan$ , 不要使用非国家法定单位, 如  $\text{ppm}$  等表示法已要求停止使用 ( $\text{rpm}$  应写为  $\text{r/min}$ ); 除  $Re$ ,  $Ma$  (其中  $e, a$  不是下标) 等几个特征数外, 变量应使用单个字母表示, 可以带上标和下标 (否则由多个字母表示单个变量, 易被误解为多个变量相乘)。

#### 3.1.3 字体

矩阵、向量请用粗斜体表示, 变量用一般斜体表示; 下标字母若为说明性的 (如英文缩写) 则用正体表示, 若为量和变动性数字及坐标轴的符号则用一般斜体表示 (设置方法: Mathtype-样式-定义-高级)。

所有文中出现的符号请另附文档说明其是变量、向量等,并说明各变量上下标的含义,以便编辑确定它们应采用的排版字体(变量符号说明表)。

请作者对易于混淆的字母和数字,如数字 0 和字母 o,英文 a 和希腊字母  $\alpha$ , O,P,S,C 等的大小写,批注“英大”(代表英文大写)、“数字 0”、“希小”(代表希腊字母小写)等。

### 3.2 图、表

图、表需给出中英文图题、表题(子图也需给出图题),但图表中图例、线型说明等一律用中文。图表一般不超过 7.7 cm 宽。金相图和计算机云图,其中的比例尺等字编辑过程中都不再重贴,按照照片处理,如有这两类图请保证美观清晰,字体用 times new roman。

#### 3.2.1 图片

对于函数曲线图,采用全框图,并注意检查以下各项:

1) 横纵坐标的标目(即变量名),尽量使用国标变量符号,变量名要在正文中交待,且与正文中符号一致;若正文中无,也可使用中文名称。

2) 坐标轴标目的量纲,对于无量纲化或无单位的,请注明“无单位”。

3) 坐标轴上的刻度线朝内,刻度值完整(坐标轴始末点均应有完整刻度值)。

4) 不同线型或图符是否有说明。

5) 是否矢量图格式,从软件中输出或拷贝矢量图格式直接插入文档中,避免用拷屏办法插图图片,否则后期无法编辑。

6) 类似图片尺寸尽量相同。

《北航学报》自 2014 年起可提供彩版印刷,如有彩印需求请作者在“出版工作单”中注明。若不需彩印,请作者作图时注意用可区分的线形或符号区分不同曲线,以保证黑白图清晰可分辨。

图中文字均用中文或变量名称表示!

图片样例见图 4 和图 5(目前是位图格式,不能编辑,作者应提供可编辑的矢量图)。

#### 3.2.2 表格

请使用三线表。选中表格,点右键打开“边框和底纹”,可对表格的边框等格式进行编辑,三线表的一般格式见表 2。

#### 3.2.3 计算、实验

文章以数值计算为主要内容的,应给出所求解的方程、重要的计算参数、初始或边界条件、难点

表 2 传输线积冰条件

Table 2 Icing conditions of transmission line			
编号	直径/m	静温/K	时间/min
4	0.0349	268.15	30
5	0.01905	268.15	30

问题的处理等,应对方法的适用性和计算精度估计有所说明;文章以实验为主要内容的,应说明实验设备、实验条件,对实验误差的估计等。便于同行重复再现所报道的内容,由于保密原因不便公开某些内容的,应向责任编辑说明。

## 4 参考文献

1) 引用文献应遵循“最新、关键、必要和亲自阅读过”的原则;

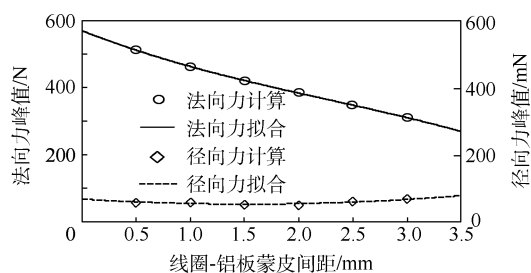
2) 参考文献应是公开出版物;

3) 应在正文中顺次引述(按在正文中被提及的先后来排列各篇参考文献的序号,所有参考文献均应在正文中提及);

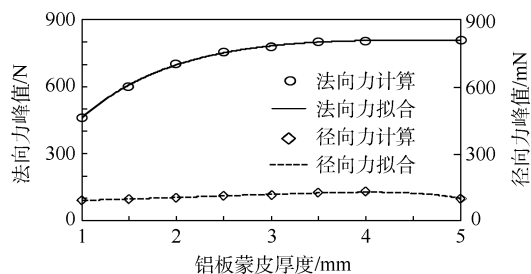
4) 文献条数 15 条以上,且有适量近两年文献;

5) 参考文献中作者为 3 人或少于 3 人应全部列出,3 人以上只列出前 3 人,后加“等”或“et al”;

6) 参考文献中外国人名书写时一律姓前,名后,姓用全称大写,名缩写为首字母(大写),不加



(a) 法向、径向电磁力峰值随线圈-铝板蒙皮间距变化曲线



(b) 法向、径向电磁力峰值随铝板蒙皮厚度变化曲线

图 4 电磁力峰值随线圈-铝板蒙皮间距和铝板蒙皮厚度变化曲线

Fig. 4 Curves of electromagnetic force peak changing with coil-aluminum-plate gap and thickness of aluminum plate

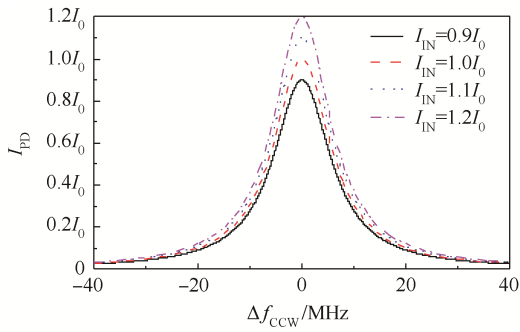


图 5 谐振腔输入光强波动对谐振曲线的影响

Fig. 5 Fluctuation influence of the resonator's input intensity on resonance curve

缩写点:

7) 为便于国际交流, 对外文文献按外文著录; 对于中文文献首先按中文著录, 同时提供英文对照, 并在其后注“(in Chinese)” 注意对中文期刊刊名应使用其标准译法 (通常在文章首页页眉可以找到)。

具体样例详见文后参考文献部分。

表 3 文献类型和标志代码

Table 3 Reference type and identification code

参考文献 类型	文献类型 标识	参考文献 类型	文献类型 标识
专著	M	学位论文	D
会议录	C	报告	R
期刊	J	标准	S
报纸	N	专利	P
汇编	G	档案	A
计算机程序	CP	电子公告	EB
数据库	DB	美图	CM
数据集	DS	其他	Z

表 4 电子文献载体和标志代码

Table 4 Electronic literature and identification code

载体 类型	磁带 (magnetic tape)	磁盘 (disk)	光盘 (CD-ROM)	联机网络 (online)
标志 代码	MT	DK	CD	OL

## 5 其他有关事项说明

1) 文章应着重撰写创新性、关键性内容, 并以一般专业人员看得懂为原则

2) 返回时间: 修改稿一般应在 10 天内返回, 或以责任编辑的要求为准。如作者不能按时返回, 请向责任编辑说明情况

3) 返回文件 (请从系统上传):

① 论文电子版 (修改部分用不同颜色标识)

② 论文修改说明 (写明对专家及编辑部所提意见如何修改)

③ 变量符号说明表 (模板见下载园地)

④ 稿件出版工作单 (word 版, 模板见下载园地); “稿件出版工作单”中有关事项请认真填写, 联系电话最好有手机。后期编辑及发行过程中, 会根据作者填写的信息与作者联系解决稿件问题, 联系方式及寄刊地址有变更的, 请及时通知责任编辑

稿件修改期间请对修改稿仔细审读、精加工, 一经排版, 一般不允许做大的改动

4) 出版过程: 责任编辑在编辑修改稿过程中常会有疑问请作者答复补正, 请作者配合及时答复; 稿件修改符合要求后, 责任编辑将根据文章页码经电子信箱发送缴纳版面费通知单, 作者应根据通知单要求及时缴纳版面费; 编辑部有权对文章进行文字性修改, 使之符合出版体例、规范要求和篇幅限制; 责任编辑在编完稿件后, 将其转至总编辑处, 按来稿先后顺次发表; 文章出版后, 免费向作者提供样刊和抽印本, 每篇文章 1 本样刊及 5 本抽印本, 如作者需要可另购样刊, 刊款可随版面费一并缴纳

5) 提前发表: 本刊一般发表周期为 1 年, 作者若有特殊情况确实需要提前发表的, 请提前向学术编辑联系及说明情况, 编辑部可根据实际情况适当安排

## 6 结论

分点总结, 列出具体的结论, 其他背景、方法都不必赘述。不与摘要和前言重复。具体样例如下:

1) 算法可实现较为优异的检索性能, 例如返回 10 张结果条件下算法检索正确率 83.15%, 召回率 8.42%, 在 60 张下正确率 39.33%, 召回率 24.61%。

2) 算法提出单张图片的引入不会造成原图片库的特征向量集和主题概率分配发生严重畸变的两个假设在一定范围 (待检索图片与原图库特征类似) 内是成立的。

3) 算法的预备工作使检索范围由原先整个库缩小至某个子类中, 虽使召回率有所损失, 但检索

时间得到较大的缩短。

4) 可预估对于特征较接近的图片库，比如人脸库，图片预备工作会产生较大的分类误差，且可能进一步影响检索性能。

为使本文提出的算法能处理各种类型的图片，仍需要优化预备工作和检索实现过程的各项参数。

## 7 模板中一些问题

1) 所有间距都是手动设置，可能与 word 模板有些差别。包括正文行间距、各级节标题前后行间距、文本字与字间距、页面设置（页边距）、双栏间距、公式前后间距、图表（标题）前后间距、页眉页脚间距等等

2) 字体设置：正文中文、英文均是五号字（10.5pt），而公式中设置为 10pt，所以公式中数字会小于普通文本数字，如  $x = 5$  和 5；带单位的量采用 siunitx 生成的话也有这个问题，如速度为 5 m/s 和 5 m/s。公式中上下标看起来与 word 版稍有差别；公式中  $g$  与 word 版  $g$  也不同，默认公式字体可能并不是 times new roman，本模板里未设置。

3) 双栏设置，采用的是 article 模板 twocolumn 选项；multicol 对浮动图表支持要差一些；twocolumn 也有些问题，比如首页跨两栏的脚注，没找到更好的办法，这里使用了 `\enlargethispage{}` 预留出脚注位置，然后用 tikz 手动调 node 的位置。还有跨两栏的图表灵活性稍差，`{figure*}`。

4) 图表中英文题注，使用 ccaption 得到。公式中向量矩阵粗斜体可以使用 `\bm` 得到。

5) 参考文献，为了自动排序，引用方便，使用 BibTex，但是参考文献格式不属于标准的，所以所有参考文献只使用 misc 这个 entry，而且只用到 misc 中 note 这一个 field，也就是把整条参考文献都放到 note 里了。工作量与 word 差不多，但是引用、增删排序更方便些。

6) 变量符号说明表，里面加了一列符号所在位置，需要用到本文件生成的辅助文件，里面有可以引用的 label 信息。

致谢 感谢某某……注意：首页注明基金项目后，文末不必再致谢。

### 参考文献 (References)

[1] 高翔. 视觉 SLAM 14 讲.

[2] 马尔. 视觉计算理论.

[3] Slam for dummies.

[4] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J.J. Leonard. Past, present, and future of simultaneous localization and mapping: Towards the robust-perception age. *IEEE Transactions on Robotics*, 32(6):1309–1332, 2016.

作者简介:

姓名 性别, 学历, 职称。主要研究方向: xxxxxxxxxx。

张某某 男, 博士研究生。主要研究方向: xxxxxxxxxx。

李某某 男, 博士, 教授, 博士生导师。主要研究方向: xxxxxxxxxx。

附录 A:

若确有特殊需要设附录的, 附录部分置于作者简介后, 标题为“附录 A:”、“附录 B:”……。公式用大写字母和数字顺序编号, 例如“(A1)”, “(A2)”。



**Title title title title title title** (不超过 10 个实词, 不出现非公知公用的缩写词)

ZHANG Moumou<sup>1,2</sup>, LI Mou<sup>1,2,\*</sup>, SHANGGUAN Moumou<sup>2,3</sup>, LIN Mou<sup>3</sup>, ZHAO Mou<sup>3</sup>, WANG Mou<sup>3</sup>

(1. School of Aeronautic Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China;

2. School of Astronautics, Beijing University of Aeronautics and Astronautics, Beijing 100191, China;

3. College of Automation, Northwestern Polytechnical University, Xi' an 710072, China)

**Abstract:** (与中文摘要内容对应, 英文摘要字数 150~200 个单词) 英文摘要应和中文摘要对应, 并请导师或专业人士把关, 保证摘要质量, 高质量的摘要有利于文摘被国际权威数据库收录, 及引起同行的重视。如果英文摘要比中文摘要更详细, 应另提供一份英文摘要的中文副本, 以便于本刊英文编辑检查英文。首次出现英文缩写时应注意写明全称。

英文摘要的撰写规范请参考本刊网站“下载园地”中的《Ei 文摘要求》。

**Key words:** keyword1; keyword2; keyword3; keyword4; keyword5 (与中文关键词一一对应, 关键词请尽量从 EI Controlled term 中选择, 以提高 EI 检索的命中率及被引频次, 网址: <http://www.engineeringvillage.com/search/quick.url>)。

---

**Received:** 2017-xx-xx; **Accepted:** 2017-xx-xx; **Published online:** 2017-xx-xx xx:xx

**URL:**

**Foundation item:** National Natural Science Foundation of China (12345678); China Postdoctoral Science Foundation(87654321)

(注: 基金项目英文名称查询“基金项目的中英文名称”)

\***Corresponding author.** Tel.: 010-8231xxxx E-mail: bhxb@buaa.edu.cn