

○ 阚江明 李文彬 著



JIYU JISUANJI SHIJUE DE

基于计算机视觉的 活立木三维重建方法

HUOLIMU SANWEI CHONGJIAN FANGFA

北京林业大学优秀博士
论文基金资助丛书

中国环境科学出版社

北京林业大学优秀博士论文基金资助丛书

基于计算机视觉的活立木 三维重建方法

**Computer Vision based Method for 3D Reconstruction of
the Standing Trees**

阚江明 李文彬 著

中国环境科学出版社·北京

图书在版编目 (CIP) 数据

基于计算机视觉的活立木三维重建方法/阚江明, 李文彬
著. —北京: 中国环境科学出版社, 2011.11
(北京林业大学优秀博士论文基金资助丛书)
ISBN 978-7-5111-0736-7

I. ①基… II. ①阚…②李… III. ①计算机视觉—应
用—测树—研究 IV. ①S776.02-39

中国版本图书馆 CIP 数据核字 (2011) 第 210843 号

责任编辑 周 煜
封面设计 玄石至上

出版发行 中国环境科学出版社
(100062 北京东城区广渠门内大街 16 号)
网 址: <http://www.cesp.com.cn>
联系电话: 010-67112765 (编辑管理部)
010-67112738 (图书出版中心)
发行热线: 010-67125803, 010-67113405 (传真)

印 刷 北京市联华印刷厂
经 销 各地新华书店
版 次 2011 年 11 月第 1 版
印 次 2011 年 11 月第 1 次印刷
开 本 850×1168 1/32
印 张 4.5
字 数 120 千字
定 价 20.00 元

【版权所有。未经许可请勿翻印、转载, 侵权必究】
如有缺页、破损、倒装等印装质量问题, 请寄回本社更换

序 言

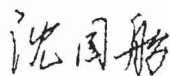
科学技术水平是知识经济时代评价一个国家国力的重要标准。科技水平高则国力强盛，无论在政治、经济、文化、信息、军事诸方面均会占据优势；而科技水平低则国力弱，就赶不上时代的步伐，就会在竞争日趋激烈的国际大舞台上处于劣势。江泽民同志在庆祝北大建校 100 周年大会上也强调指出：“当今世界，科学技术突飞猛进，知识经济已见端倪，国力竞争日益激烈。”因此，提高科学技术水平，提高科技创新能力已为世界各国寻求高速发展时所共识。我国将“科教兴国”作为国策也表明了政府对提高科技水平的决心。博士研究生朝气蓬勃，正处于创新思维能力最为活跃的黄金年龄，同时也是我国许多重要科研项目的中坚力量，他们科研成果水平的高低在一定程度上影响着—所高校、一个科研院所乃至我国科研的整体水平。国务院学位委员会每年一度的“全国百篇优秀博士论文”评选工作是对我国博士研究生科研水平的集体检阅，已被看做是博士研究生的最高荣誉，对激励博士勇攀科技高峰起到了重要的促进作用。北京林业大学不仅积极参加“全国百篇优秀博士论文”的推荐工作，还以此为契机每年评选出三篇校级优秀博士论文并设立专项基金全额资助论文以丛书形式出版，这是一项非常有意义的工作，对推动学校科研水平的提高将发挥重要作用。

从人才培养的角度来看，如何提高博士研究生的创新思维能力和综合素质，高质量地向社会输送人才备受世人关注。提高培养质量的措施很多，但在培养中引入激励机制，评选优秀博士论文并资助出版，不失为一种好方法。博士生和导师可据此证明自

己的学术能力，确立自己的学术地位；也可激励新入学的研究生尽早树立目标，从而在培养的全过程严格要求自己，提高自身的素质。

因学科的特殊性，要想出色完成林业大学的博士论文有许多其他学科所不会遇到的困难，如研究周期长，野外条件难以严格控制，工作条件艰苦等。非常欣慰的是北京林业大学的博士生们不仅克服困难完成了学业，而且已经有人中选“全国百篇优秀博士论文”。而该丛书资助出版的“校级优秀博士论文”所涉及的研究领域、研究成果的水平也属博士论文中的佼佼者，令人欣喜。对这些博士生所取得的成果我表示祝贺，同时也希望他们以及今后的同学们再接再厉，取得更好的成绩报效祖国。

中国工程院副院长、院士



2002年8月10日

北京林业大学优秀博士论文基金资助丛书

编辑委员会

主 任：尹伟伦

副主任：马履一

委 员（按姓氏笔画排序）：

刘俊昌 余新晓 吴 斌 张启翔 李凤兰

李俊清 俞国胜 赵广杰 骆有庆 贾黎明

续九如 蒋湘宁 翟明普

秘 书：王兰珍

目 录

1	绪 论	1
1.1	研究背景与意义	1
1.2	Marr 计算机视觉理论框架	3
1.3	计算机视觉的三维重建研究综述	6
1.4	本文的研究内容	10
2	双目立体视觉基础	15
2.1	射影几何基础	15
2.2	摄像机成像模型	19
2.3	双目立体视觉原理	24
2.4	小结	32
3	活立木枝干提取	33
3.1	图像预处理	33
3.2	图像分割	37
3.3	立木枝干提取	43
3.4	小结	47
4	角点检测与匹配	48
4.1	角点检测	48
4.2	角点匹配	53
4.3	SIFT 角点检测与 NCC 匹配	56

4.4	小结	65
5	基本矩阵估计与射影重建	66
5.1	基本矩阵的估计	66
5.2	基于粒子群优化的基本矩阵估计	72
5.3	射影重建	77
5.4	小结	83
6	相机自标定与欧氏重建	84
6.1	自标定方法概述	84
6.2	基于遗传算法的自标定	93
6.3	欧氏三维重建	98
6.4	小结	105
7	结论与展望	106
7.1	主要研究结论	106
7.2	未来研究展望	108
附 表		109
参考文献		121
后 记		131

1 绪论

1.1 研究背景与意义

我国森林总量不足，人均森林面积仅为 0.132hm^2 ，不到世界平均水平的 $1/4$ 。木材供应量不足，主要靠进口木材满足国内木材消费需求。我国是世界上第二大木材进口国，位列美国之后，日本之前（孙仁山，2006；张俊梅，2005）。大力发展人工林是解决这一问题的最佳途径。我国从 20 世纪 50 年代开始发展人工林，现有人工林面积约 $5\,326\text{万 hm}^2$ ，其中达到速生丰产林标准的林地已有 200多万 hm^2 。由于管理粗放，导致出材率偏低，并且由于没有按用材林标准进行定向培育，致使成林的木材质量较低，不能满足工业用材的质量要求。因此需进行大量人工工业用材林的抚育，这是解决我国木材资源短缺的主要途径之一。同时林业生产作业环境比较恶劣，工人劳动强度大，现有作业机械的使用性能受到很大限制。为此，日本、美国和瑞典等林业发达国家已经或正在开展林业机器人（装备）的研究工作，涉及育苗、造林、森林抚育、病虫害防治、森林采伐与运输、剩余物综合利用、木材精深加工等绝大多数林业生产方面。我国学者在林业机器人方面也开展了相关研究工作，东北林业大学的陆怀民教授、马岩教授及其团队在“863”项目的资助下研制了林木球果采集机器人，开展了轮履复合森林巡防机器人的研究工作；北京林业大学的李文彬教授及其团队在“国家‘十五’科技攻关”项目的资助下研制了遥控型立木整枝机器人，在国家林业公益性行业科研专项项

目的资助下正在研制多功能林果采摘机器人和单轨运输系统；北京林业大学刘晋浩教授及其团队在“863”项目的资助下研制了伐根机器人、防风固沙草方格铺设机器人，在“国家‘十一五’科技支撑”项目的资助下研制了林业多功能采育机器人。在林业机器人自动化与智能化方面，瑞典于默奥大学、瑞典农业科技大学、美国爱达荷大学和爱荷华州立大学等的国外学者开展了一系列的研究。在国内，除了东北林业大学的陆怀民教授、马岩教授、李文彬教授、刘晋浩教授及其团队在研制林业机器人过程中也开展了相关研究外，南京林业大学成立了“智能控制与机器人技术研究所”积极开展林业机器人智能控制方面的研究。哈尔滨林业机械研究所、南京林业大学、中南林业科技大学、湖北林科院、浙江农林大学等教学科研机构的学者也开展了林业机器人相关的研究工作。目前，已经开发成功或正在开发的林业机器人包括林木修剪机器人、人工林间伐机器人、防火带清理机器人、林业除草机器人、林木球果采集机器人、立木整枝机器人、林业收获机器人和森林巡防机器人等。林业机器人的研究，对提高生产效率和减轻工人劳动强度，提高人身安全等方面都有很重要的作用。随着科学技术的发展，特别是计算机技术的迅速发展与进步，林业机器人正向智能化方向发展，逐步形成林业自主作业机器人。这些林业自主作业机器人将会配备计算机视觉系统。

配备机器视觉系统的林业自主作业机器人可以在不需要人参与或者很少参与的情况下自主进行林业生产作业，比如修枝、间伐、清林、采伐等相关作业，将林业生产工人从艰苦的劳动中解放出来，或者减轻劳动者的劳动强度，避免林业生产作业安全事故的发生，提高林业生产作业效率。因此进行配备机器视觉系统的林业自主作业机器人的研究具有很高的实际意义和实用价值（阚江明，2007；Kan J，2008）。

2002年，瑞典开展了林业机器人的导航系统的研究，并开发了无人驾驶集材原形车，解决了自动蔽障和定位等问题；加拿大

从 1993 年就开始自动间伐机器的研究，开发了概念机器，重点是林区环境进行感知自主控制方法的研究；芬兰等国家正在进行无人驾驶联合采伐机的研究，这些林业机器人都配备了计算机视觉系统（Vestlund K, 2006）。

智能型林业装备在行走和作业过程中，其机器视觉系统会从不同角度获得环境中同一活立木的二维图像，为了蔽障和导航，必须进行立木枝干的三维重建，也就是利用二维图像实现对活立木三维信息的估计（Vestlund K, 2006）。这是一个典型的计算机视觉三维重建的问题，也是目前计算机视觉领域的研究热点和难点。比如机器人自主导航中，机器人通过观察周围景物获取环境的图像，再由二维图像计算得到环境的三维信息以确定行走的方向和距离；还有在考古、建筑、精密工业测量、军事等领域研究非常活跃。基于计算机视觉的三维重建也是一个二维到三维的不适定数学问题，需要一定的约束条件才能够有唯一的稳定解。

本研究不仅考虑林业自主作业机器人的实际应用，也着眼于计算机视觉领域的研究热点和难点，其研究可谓具有较强的理论与实际意义。

1.2 Marr 计算机视觉理论框架

计算机视觉的开创性工作是从 20 世纪 50 年代中期开始的，是基于一系列几何光学、古典数学算法，并融合解剖学、神经生理学、计算机技术、统计学、运筹学、图论、信号分析与处理等多学科为一体的综合性科学，它也是人工智能（Artificial Intelligence）的一个重要分支（人工智能的四大分支是：视觉、语音与听觉、学习与推理、触觉与感知）。

1965 年 Roberts 首创了“积木世界”下的计算机视觉试验（Roberts, 1965）。计算机视觉正是在这种最简单的多面体环境下，对最简单的由点、线、面组成的物体识别和定位开始的。

1982 年，麻省理工（MIT）的 Marr 教授第一次从信息处理的角度，运用图像处理技术、心理物理、神经解剖学以及临床精神病理学提出了一个较为完善的计算机视觉系统的框架（Marr，1982）。Marr 教授建立的视觉计算理论使得计算机视觉研究有了一个比较明确的体系，对人类视觉和计算机视觉的研究产生了深远的推动作用。下面简要介绍 Marr 视觉计算理论的基本思想和理论框架。

1.2.1 视觉系统研究的三个层次

Marr 理论认为，视觉是一个信息处理系统，对此系统研究分为三个层次：计算理论层次、表示与算法层次、硬件实现层次，如表 1-1 所示。

表 1-1 Marr 视觉理论三个层次

Tab.1-1 the Hierarchy of Marr's Vision Theory

计算理论	表示与算法	硬件实现
计算的目的是什么？	如何实现这个计算理论？	在物理上如何实现这些表示和算法？
为什么这一计算是合适的？	输入、输出的表示是什么？	
执行计算的策略是什么？	表示与表示之间的变换是什么？	

计算视觉理论要回答视觉系统的计算目的和策略是什么，或视觉系统的输入和输出是什么，如何由系统的输入求出系统的输出。从信息处理的观点来看，至关重要的乃是最高层次，即计算理论层次。这是因为构成视觉的计算本质，取决于解决计算问题本身，而不取决于用来解决计算问题的特殊硬件。换句话说，通过正确理解待解决问题的本质，将有助于理解并创造算法（马颂德，2003）。

1.2.2 视觉信息处理的三个阶段

视觉信息处理过程划分为三个阶段，如表 1-2 所示。第一阶

段也称为早期阶段,是将输入的原始图像进行处理,抽取图像中诸如角点、边缘、纹理、线条、边界等基本特征,这些特征的集合称为基元图。第二阶段中期阶段是指在以观测者为中心的坐标系中,由输入图像和基元图恢复场景可见部分的深度、法线方向、轮廓等,这些信息包含了深度信息,但不是真正的物体三维表示,因此,称为二维半图。在以物体为中心的坐标系中,由输入图像、基元图、二维半图来恢复、表示和识别三维物体的过程称为视觉的第三阶段后期阶段(马颂德, 2003; Milan S, 2003)。

表 1-2 由图像恢复形状信息的表示框架

Tab.1.2 Presentation Frame for Recovering Structure from Images

名称	目的	基元
图像	光强表示	图像中每一个像素点的灰度值
基元图	表示二维图像中的重要信息,主要是图像中的强度变化位置及其几何分布和组织结构	零交叉,斑点,端点和不连续点,边缘片断,有效线段,纹理,曲线,边界
二维半图	在以观测者为中心的坐标系中,表示可见表面的方向、深度值和不连续的轮廓	局部表面朝向“针”基元,离观测者的距离,深度上的不连续点,表面朝向的不连续点
三维模型表示	在以物体为中心的坐标系中,用由体积基元和面积基元构成的模块化多层次表示,描述形状及其空间组织形式	分层次组成若干三维模型,每个三维模型都是在几个轴线空间的基础上构成的,所有体积基元或面积形状基元都附着在轴线上

Marr 理论是计算机视觉研究领域的划时代成就,但该理论不是十分完善的,许多方面还有争议。比如,该理论所建立的视觉处理框架基本上是自下而上,没有反馈;还有,该理论没有充分地重视知识的应用。尽管如此,该理论还是给了我们研究计算机视觉许多珍贵的哲学思想和研究方法,同时也给计算机视觉研究领域创造了许多研究起点。本文在 Marr 的视觉理论框架基础之上,针对其第二阶段中的立体视觉三维重建进行了研究,面向智能型林业装备的实际应用,研究活立木的三维重建方法。

1.2.3 计算机视觉研究的复杂性

对于人类视觉来说，识别和理解周围场景是一件非常容易的事，但对计算机或其他仪器设备来说，识别和理解周围场景却是一件很困难的事，主要原因有如下几方面：

(1) 图像的多义性：由于二维图像中深度和不可见部分的信息丢失，会出现不同形状的三维物体投影成相同的图像，另一方面，一个物体从不同角度可以投影成不同的二维图像。

(2) 环境因素影响：场景中的诸多因素，包括照明、物体形状、表面颜色、相机位置以及空间关系等，任一因素的变化都将引起二维图像的变化。

(3) 知识导引：同样的图像在不同的知识导引下，将会产生不同的效果。

(4) 大量数据：图像的位图性质决定了每幅图像都有较大的数据量，随着图像数字化的分辨率提高、图像色彩位数加大、序列图像的应用以及实时性的要求，对图像处理的理论和设备都有了更高的要求。

1.3 计算机视觉的三维重建研究综述

三维重建是计算机视觉研究的主要内容之一，属于 Marr 视觉计算理论框架中的中级视觉部分。三维重建是通过二维图像中的基元图来恢复三维空间信息，也就是要研究三维空间点、线、面的三维坐标与二维图像中对应点、线、面的二维坐标间的关系，实现定量分析物体的大小和空间物体的相互位置关系。

双目立体视觉是计算机立体视觉的重要分支，有主动双目立体视觉和被动立体视觉之分。主动立体视觉需要使用一个专门的光源装置来提供目标物体周围的照明；而被动立体视觉则是由周围物体的光线来提供照明，比如自然光。针对应用实际情况，本

论文的研究针对被动双目立体视觉。

早期双目立体视觉的三维重建都必须对摄像机进行预标定,因此只能应付静止或者已知环境的三维重建工作(Longuet-Higgins H C, 1981)。未标定图像的三维重建始于 1992 年 Faugeras 和 Hartley 提出的利用未标定图像序列计算射影结构的理论(Hartley R, 2000)。基于未标定图像的三维重建主要从图像匹配、摄像机标定和三维重建三个方面进行研究。

三维重建的首要任务是要解决两幅图像间的对应关系即立体匹配。立体匹配包括两个子问题:特征检测和特征匹配。常用的匹配特征有特征点、特征线、特征区域,其中以点特征研究较多。Dreschler 和 Nagel 等提出了基于 Gaussian 曲率原则的检测方法(Dreschler, 1982)。Kitchen 和 Rosenfeld 提出的角点检测器利用了灰度沿边界轮廓梯度方向变化最大的性质(Kitchen, 1982)。Nobel 试图用微分几何给出角点检测的理论公式(Nobel, 1988)。Moravec 于 1977 年首次提出了用“兴趣算子”提取角点的方法(Moravec, 1977)。1988 年 Harris 和 Stephens 对算子进行了改进,提出了 Harris 算子(Harris, 1988)。Smith 提出了著名的 SUSAN 角点检测算子,该方法对像素周围区域最小化,用统计特性来决定该像素的属性,即角点、边上的点还是圆上的点(Smith, 1997)。Lowe D G 1999 年提出 SIFT 算法,其是一种提取局部特征的算法,在尺度空间寻找极值点,提取位置、尺度、旋转不变量(Lowe, 1999; Chang M Y, 2005)。

特征匹配往往以灰度相似性为基础,辅以对极几何约束或其他约束进行搜索。Beardsley 等将提取的角点作为特征点,运用相关性进行匹配,将匹配的结果用奇异值分解求取了基本矩阵(Beardsley P, 1996)。Pritchett 和 Zisserman 等提出了用单应矩阵(Homography)取代传统的灰度相似性和极线约束作为匹配的准则,他们近似认为特征点及其周围小块区域是空间中平面的成像,因此匹配点对之间应近似满足单应矩阵的关系。另外,他们试图

寻找一种整体相似变换,以使两幅图像在相差一个常数因子的情况下具有最大相关性,根据整体变换,估计图像间局部区域的仿射变换,用局部变换来寻找匹配点(Pritchett P, 1998)。Pritchett 等的方法对于某些含有丰富平面信息的图像特别有效。Lhuillier 和 Luong 等提出了一种稠密匹配的策略,该策略综合了以对极几何约束为代表的全局约束和以灰度相似、单应矩阵为代表的局部约束,在每次匹配过程中,选取当前灰度相似性最大的匹配对,在它的周围小区域内寻找更多的匹配对,重复这种过程直至匹配对充满整幅图像,最后再应用对极几何约束去除错误匹配(Lhuillier M, 2000)。该方法对于纹理稠密的图像特别有效,缺陷是精度不高(Lhuillier M, 2000)。Pilu M 提出了一种基于奇异值分解的匹配方法,首先构造一个包含特征点的 Gaussian-weighted 距离信息的强度矩阵 G 并对其进行奇异值分解,接着将得到的对角矩阵的对角元素全部替换为 1 而得到一个同维矩阵 P ,通过矩阵 P 中各个元素的值就能判断出对应的特征点是否匹配(Pilu M, 1997)。由于整个计算过程中全部是代数意义下的运算,所以匹配的精度有待进一步提高。

要实现欧氏空间下的三维重建,必须确定摄像机的参数,即对摄像机进行标定(刘金颂, 2008)。微软研究院的张正友提出了一种利用平面模板的标定方法,该方法只需从不同角度对模板拍摄几幅图像,通过每幅图像的单应矩阵即可计算出摄像机内参数,缺点是每次都需要标定,且都需要标定物(Zhang Z, 1999)。近年来,无需标定物、基于图像序列的自标定方法已成为标定研究的一个重要方向。Faugeras, Luong 和 Maybank 等首先提出了自标定的概念,从射影几何的角度出发证明了每两幅图像间存在着两个形如 Kruppa 方程的二次约束,通过直接求解 Kruppa 方程组可以解出摄像机的内参数(Faugeras O, 1992; Maybank S J, 1992),但是 Kruppa 方程只能表示图像两两之间的关系,不能将整个图像序列相联系,且求解困难。鉴于直接求解 Kruppa 方程的困难,研

究者们又提出了分层逐步标定的思想,即首先对图像序列做射影重建,在此基础上再进行仿射标定和欧氏标定,该类方法以 Hartley 的 QR 分解法(Hartley R, 1996)、Triggs 的绝对二次曲面法(Triggs B, 1997)和 Pollefeys 的模约束法(Pollefeys M, 1996)等为代表,这些方法的本质都是求解无穷远平面的绝对二次曲线或曲面,缺点是对输入数据有很高的精度要求。自标定方法不能使用解析式表达,需要使用非线性优化,非线性优化带来的计算难度和误差是不可避免的。

在三维重建方面,Tomasi 和 Kanade 等在假定摄像机为正交投影模型的前提下,利用仿射分解的方法解出了三维结构和摄像机运动。该系统使用基于光流的跟踪器技术来解决特征点的匹配问题(Tomasi C, 1991)。但由于该系统采用的摄像机模型是正交投影模型,这种模型只有在当物体的深度远大于物体的尺寸时才合理,因此有一定的局限性。Pritchett 和 Zisserman 完成的视觉导航系统利用了分层重构的思想,即首先对图像序列做射影重构,再将射影重构逐步提升到仿射重构和欧氏重构(Pritchett P, 1992; Moons T, 1994)。Debevec、Taylor 等完成了著名的建筑物重构系统 Façade (Debevec P E, 1996)。该系统要求首先得到建筑物的粗略几何模型和摄像机运动模数,然后将该模型反投影到图像上与实际图像作比较,通过减小反投影误差最终计算出建筑物的精确三维结构。为增强视觉效果,该系统还使用了基于视点的贴纹理技术。该系统不足之处在于需要预先得到建筑物的几何模型,而且建筑物的几何结构不能太复杂。Shum H Y 等提出了一种人机交互式重构系统,可以从一组全景图中恢复出三维结构或者将场景表示成一系列按深度划分的分层的集合(Shum H Y, 1998)。该系统不需要作图像间的匹配,但要求用户在重构过程中交互式地确定场景中的一些几何约束,如共面、平行、垂直等。Faugeras 等的系统利用分层重构、自标定等方法从图像序列中重构出了建筑物(Faugeras O, 1998)。在他们的系统中,考虑了建筑物的特殊