全视野时空视觉导航

真实景物的成象、建模与表示

朱志刚

清华大学计算机科学与技术系

First Hundred National Excellent Doctorate Dissertations Series

China Higher Education Press

December 2001

Full View Spatio-Temporal Visual Navigation

Imaging, Modeling and Representation of Real Scenes

Zhu Zhigang
Department of Computer Science and Technology
Tsinghua University

First Hundred National Excellent Doctorate Dissertations Series
China Higher Education Press
December 2001

论文简介

自然环境中视觉导航的研究涉及到计算机视觉的许多方面，是一个极有意义但又很有难度的综合性课题。视觉导航的基本要求是机器人能够自主进行全局定位、道路跟踪和障碍物检测，其中每一部分都要以环境建模为基础。论文提出了面向任务的多尺度全视野环境综合建模方法，将路标定位的全景图象建模，道路识别的全方位图象建模和障碍物检测的双目注视图象建模相结合，改进了依赖局部、单一信息的视觉导航方法。论文研究了灵巧传感器设计、鲁棒视觉算法和多层次模型表示等问题，主要贡献有：基于图象稳定和全景外极面图象分析的 3D 环境建模方法，全方位图象特征和神经网络相结合的道路建模方法，和基于重投影变换的双目障碍物检测方法。论文给出了大量室外景物建模的实验结果。

摘　要

自然环境中视觉导航的研究涉及到计算机视觉中的各个主要方面，是一个极有意义但又很有难度的综合性课题。机器人视觉导航的基本要求是机器人能够自主进行全局定位、道路跟踪和障碍物检测，其中每一部分都要以环境建模为基础。而真实景物的环境建模又离不开三个主要的部分：灵巧视觉成象、鲁棒视觉算法和多层次视觉表示。本书是作者 1997 年在清华大学博士论文的修改版。在论文工作中作者系统深入地研究了视觉导航中的自然环境建模问题，包括：

1. 提出了面向任务的多尺度全视野的环境综合建模方法。将路标定位的全景图象建模，道路识别的全方位图象建模和目标(障碍物)检测的双目注视图象建模相结合，改进了依赖局部、单一信息的视觉导航方法。研究了面向任务的环境建模方法中的传感器设计、视觉算法和模型表示问题。

2. 提出了由非精确摄象机运动下的图象序列建立 3D 环境全景模型的两步法，即基于运动滤波的图象稳定和基于时-空-频域遮挡模型的全景外极面图象分析。本文推广了全景图象方法和外极面图象方法，从而使之能适用于具有抖动的图象序列分析，并避免了一般运动视觉方法的不适定问题和特征对应问题、基于空域约束的迭代方法的局部最小化等问题。

3. 提出了全方位图象特征和神经网络相结合的道路建模方法，较好地解决了机器人依赖局部视野信息导致迷路的问题、视觉算法依赖特定环境特征导致推广性差等问题。本文提出了主分量分析和 Fourier 变换相结合的全方位图象特征空间数据压缩和旋转不变特征提取方法，设计了在识别道路类型基础上进行道路方向估计的组合神经元网络，从而提供了一种能够使机器人自动适应不同道路类型的可行方法。

4. 提出了基于重投影变换的障碍物检测方法，跳出了传统立体视觉方法中特征抽取、匹配和三维恢复的模式。设计了无特征提取和对应的立体视觉新算法，并给出了克服摄象机俯仰影响的动态重投影变换算法，增强了系统在颠簸道路上

运行的适应性。利用重投影变换后的双目图象对中路面特征零视差的特性，通过障碍物有无判断、区域确定和三维测量的“分步渐进”过程，实现了高效可靠地对路面障碍物的实时检测。

5. 系统实现：(1) 实现了全景建模中的图象稳定、遮挡恢复和深度分层，从而为机器人全局定位的自然路标提取和真实环境再现的图象合成打下了基础; (2) 设计和实现了适合于室外自然环境的全方位成象系统和单摄象机双目立体成象系统； (3)实验验证了全方位道路图象神经网络学习方法的有效性；(4) 实现了实时障碍物检测系统，经过大量的室外道路环境下障碍物检测的试验，说明系统具有很强的实用性。

关键词：视觉导航，景物建模，图象稳定，重投影变换，全景视觉，全方位视觉，神经元网络，灵巧传感器，路标定位，道路理解，障碍物检测。

Abstract

Visual navigation of a mobile robot in a natural environment has always been a very interesting but challenging problem. It involves almost every aspect of computer vision issues - from visual sensors, robust algorithms to visual representation. The fundamental tasks of visual navigation include global localization, road following and obstacle detection. Visual environment modeling is the foundation of visual navigation - and to an extension, most of the real world problems in computer vision. Three key components – smart visual sensing, robust visual algorithms and hierarchical visual representations are equally important in real scene modeling. This book is based on the author's Ph.D. thesis at Tsinghua University. The book presents a systems approach to visual modeling of a natural scene for robot navigation and in large to scene understanding and representation. There are four main contributions in the thesis:

1. A purposive, multi-scale and full-view visual scene modeling approach is proposed for visual navigation in a natural environment. It combines three novel modules into an integrated system, the **POST**: ***P****anoramic vision for scene modeling,* ***O****mnidirectional vision for road understanding and* ***ST****ereo vision for obstacle detection*. This approach tries to overcome the drawbacks of traditional visual navigation methods that have mostly depended on local and/or single view visual information. However, the proposed approach is not just a simple combination of the three novel sensors and methods, but rather a systematic integration under the strategy of purposive vision ("the right way for the right work"), and under the philosophy of systems approach which emphases that "the whole is more than sum of its components". Thus, right sensor design, adequate level of scene representation and corresponding robust algorithms are specifically explored for each given task while the interconnection among the vision sub-systems are taken into consideration under the overall goal of autonomous navigation.

2. A two-stage method is presented for 3D panoramic scene modeling from image sequences captured by a video camera subject to vibration on a common road surface. First, a 3D image stabilization method is proposed which eliminates vibration from

vehicle's motion so that "seamless" epipolar plane images (EPIs) and panoramic view images (PVIs) can be generated. Second, an efficient panoramic EPI analysis method is proposed to combine the advantages of both PVIs and EPIs efficiently in three important steps: locus orientation detection, motion boundary localization, and occlusion/resolution recovery. The two-stage method not only combines Zheng-Tsuji's PVI method with Baker-Bolle's EPI analysis, resulting in the so-called panoramic EPI method, but also generalizes them to handle image sequence subject to unpredictable camera vibration. Finally, a compact layered representation for a large-scale scene has been proposed, which can be used in both visual navigation and image-based rendering.. Since camera calibration, image segmentation, feature extraction and matching are avoided, all the proposed algorithms are fully automatic and rather general.

3. A new road following approach, the ***R**oad **O**mni-View **I**mage **N**eural **N**etworks* (**ROVINN**), has been proposed. It combines the omni-directional image sensing technique with neural networks in such a manner that the robot is capable to learn recognition and steering knowledge from the omnidirectional road images which in turn guarantee the robot never to get lost. The ROVINN approach brings Yagi's COPIS (conic omnidirectional projection image sensor) method to the outdoor road scene and provides an alternative solution different from the CMU's ALVINN. Compact and rotation-invariant image features are extracted by integrating the principle component analysis (PCA) and the Fourier transform (DFT). The modular neural networks of the ROVINN estimate road orientations more robustly and efficiently by classifying the roads as a first step, which enables the robot to adapt to various road types automatically.

4. A novel method called *Image Reprojection Transformation (***IRT***)* is presented for stereo-vision-based road obstacle detection. Obstacle detection is modeled as a reflexive behavior of detecting anything that is different from the planar road surface. Dynamic reprojection transformation algorithms are developed so that the algorithms can work in rough road surface. The novelty of the (dynamic) reprojection transformation method, which ensembles the gaze control of the human vision, lies in the fact that it brings the road surface to zero disparity so that the feature extraction and

matching procedures of the traditional stereo vision are avoided in the proposed obstacle detection algorithms. The progressive processing strategy from yes/no verification, focus of attention, to 3D measurement based on the reprojection transformation make the obstacle detection efficient, fast and robust.

To validate the proposed strategies and methods, we have implemented the following algorithms and systems.

(1) 3D **Scene modeling system**. In the 3D panoramic scene modeling, the algorithms of motion filtering and image stabilization, kinetic occlusion detection and depth layering, have been developed, and the novel depth layered setting (DLS) models have been constructed for many image sequences. These efforts found a ground base for landmark selection of global localization and image synthesis of photo-realistic image-based rendering.

(2) **Designs of novel sensors**. An omni-view image sensor is designed and realized, and its properties for the outdoor road understanding are thoroughly studied. A patented single camera binocular vision system is also designed and constructed, and has been put into real use of obstacle detection. An integrated full view smart sensor POST (Panoramic, Omnidirectional and STereo vision) is proposed, which integrates 360-degree omnidirectional view, binocular forward view as well as both left and right side views by using a single camera.

(3) **Real scene experiments.** Experimental results of training and testing the ROVINN using real road images have shown that the proposed method for road following is quite promising. A real-time visual obstacle detection system has been set up and tested in the outdoor road scene extensively.

Index Terms: Visual navigation; Scene modeling; Image stabilization; Reprojection transformation; Panoramic vision; Omnidirectional Vision; Neural network; Smart sensors; Landmark localization; Road understanding; Obstacle detection.

致　谢

我的博士论文是在博士生导师石纯一教授和博士生导师徐光佑教授的指导下完成的。石纯一教授严谨的治学态度和富有启发性的指导使我获益匪浅。徐光佑教授引导我进入视觉环境建模的研究领域，本论文的许多思想方法和视觉导航体系的建立和他的指导是分不开的。他对问题的洞察力和远见使我学到了超出专业研究范围的东西，并将使我终身受益。林学言教授在视觉算法的深入研究方面给予我许多启迪和帮助，并共同提出了重投影变换的原理，他那追求创新的精神对我永远是巨大的鞭策。石定机教授启发我设计新型的视觉传感器，并且共同研制了双目立体成象系统和全方位成象系统，他使我时刻牢记：研究必须与工程和应用相结合。在论文工作过程中也得到吴伯新高工、陈永康高工等老师的帮助，在此深表感谢。

在论文工作中还得到清华计算机系视觉实验室研究生、工作人员和我家人的协助和关心。杨波帮助我实现了实时障碍物检测系统；图象运动检测和图象稳定的算法是由杨雨东、张强帮助实现的。尤其要感谢我的妻子李晓彦，她是我十几年如一日坚持研究工作的强大后盾。

本书的出版我应该感谢我的博士论文评审专家及答辩委员会，首届优秀博士论文评审委员会和高教出版社，他们卓越的工作和极具价值的评审意见促使我的论文能够入选中国首届百篇优秀博士论文，并在高教出版社正式出版。在本书成稿的过程中，正值著者在安城麻省大学（UMASS-Amherst）作访问教授、资深研究员，UMASS 计算机系视觉实验室的联合主任 Edward M. Riseman 教授和 Allen R. Hanson 教授不仅给予许多便利的研究条件，而且还提出许多极有价值的意见和建议。高教出版社的康兆华老师和肖娜女士的有效工作使我远在万里之外能够顺利完成书稿。 在此一并表示衷心的感谢。