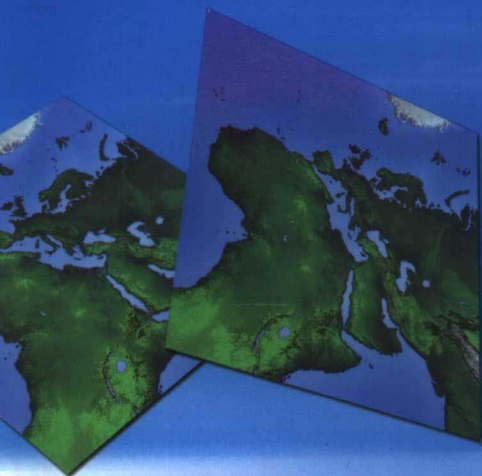


由2维影像 建立3维模型

徐刚 著

郑顺义 译 苏国中 审



WUHAN UNIVERSITY PRESS

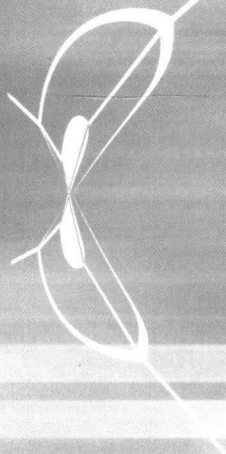
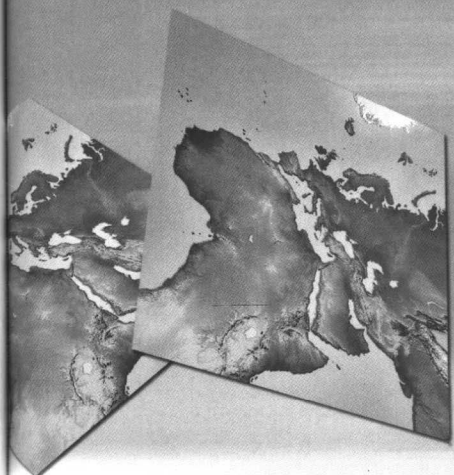
武汉大学出版社

由2维影像 建立3维模型

徐刚 著

郑顺义 译 苏国中 审

Image Based Modeling and Rendering



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

由 2 维影像建立 3 维模型/徐刚著;郑顺义译;苏国中审. —武汉:
武汉大学出版社, 2006. 9

ISBN 7-307-05221-0

I. 由… II. ①徐… ②郑… ③苏… III. 计算机图形学
IV. TP391. 41

中国版本图书馆 CIP 数据核字(2006)第 109522 号

责任编辑:杨 华 责任校对:王 建 版式设计:杜 枚

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:wdp4@whu.edu.cn 网址:www.wdp.com.cn)

印刷:湖北省通山县九宫印务有限公司

开本:880×1230 1/32 印张:4.875 字数:131 千字 插页:1

版次:2006 年 9 月第 1 版 2006 年 9 月第 1 次印刷

ISBN 7-307-05221-0/TP·212 定价:10.00 元

版权所有,不得翻印;凡购买我社的图书,如有缺页、倒页、脱页等质量问题,请与当地图书销售部门联系调换。

内 容 提 要

个人电脑的普及, CPU 运算速度的加快, 图形加速卡性能的提高, 数码相机的普及, Internet 的兴起等要素结合起来, 使得个人手拍影像, 并能建立 3 维模型, 进行自娱自乐并发布的时代来到了, 这样的预测或期待越来越接近现实, 越来越被更多的人所关注.

从输入影像开始, 到恢复 3 维形状, 最后将得到的 3 维结果利用计算机图形学的技术 (CG) 进行表达、显示, 本书对该过程中涉及的原理、概念、表述、算法等进行了详细的叙述.

因为 3 维形状和运动的表示以及计算需要一定的数学知识, 为了便于对一些必需的线性代数和微分知识的学习和理解, 本书中进行了必要的讲解. 这些数学知识作为理解本书的基础, 在附录中有详细的介绍.



前 言

个人电脑的普及, CPU 运算速度的加快, 图形加速卡性能的提高, 数码相机的普及, Internet 的兴起等要素结合起来, 使得个人手拍影像, 并能建立 3 维模型, 进行自娱自乐并发布的时代来到了, 这样的预测或期待越来越接近现实, 越来越被更多的人所关注。

当然, 这是基于 3 维技术基本成熟这样一个前提。

本书的主要内容包括, 通过输入的影像恢复 3 维形状, 得到的结果用计算机图形学的技术 (Computer Graphics, CG) 表示, 本书对整个过程中涉及的原理、概念、描述、算法进行了详细叙述。身处时代前列, 假使能加速时代潮流的发展, 即使甚少, 我也感到很荣幸。

对于 3 维形状和运动的表示和计算, 数学知识不能缺少, 为了更好地理解本书, 需要学习一些线性代数和微分的知识。作为基础的数学知识, 在附录中有相应的介绍, 必要时可以参考。另外, 在第二章以后的各章有一些练习题, 在本书的最后附有相应的参考答案。本书中有多处难以写入的详细说明, 都用作了练习题。

本书收录的大部分算法是作者工作单位立命馆大学理工学部情报学科计算机视觉研究室的试验结果。另外, 本书所写的内容, 可以用于 Kuraves-K 公司所销售的 3 维量测系统的开发。

这里, 对参与本书试验的杉本典子、松井裕司、藤井友和、堀河武志、中山贵央、寺井準一、松冈美希、吉田雅博、森好司、寺本博久等表示感谢。没有他们参与试验, 本书中所写的算法无法完成。另外, 借此场合, 感谢常常与我一起讨论相关问题的朋友——Microsoft 的 Zhengyou Zhang 博士, 经常打扰的同事田中弘美教授、八村広三教授、小川均教授, 秘书室圆女士, 以及为本书的编辑作出努力的近代



科学社的福泽富仁编辑部长。另外,本书的一部分是在 Microsoft 中国研究院以及 Motolola 澳大利亚研究中心执笔撰写的,对那里的 Harry Shum 博士和 Wanging Li 博士表示感谢。

最后,感谢我的家人(明代,星来和悠诗)。

徐刚

2000 年 12 月

目 录

第 1 章 从 2 维影像到 3 维模型	1
1.1 进入由个人拍摄的影像建立 3 维模型的时代	1
1.2 本书中使用的专门术语及数学符号	3
第 2 章 影像、相机及投影	6
2.1 数字影像与数字影像坐标系	6
2.2 针孔相机与中心投影	7
2.3 摄影矩阵及外部参数	9
2.4 规一化相机及内部参数	11
2.5 投影近似: 平行投影, 弱中心投影, 模拟中心投影, 仿射投影	14
2.6 练习题	18
第 3 章 3 维空间中旋转的表示和计算	19
3.1 欧拉角	19
3.2 roll, pitch, yaw	20
3.3 旋转轴及旋转速度	20
3.4 4 元数	22
3.5 正交矩阵、旋转矩阵及反转	23
3.6 利用旋转前后的 3 维向量进行旋转的最优化计算	24
3.7 练习题	27
第 4 章 核线几何	28



4.1	中心投影中核线几何的概念及核线方程式	28
4.2	中心投影中的基本矩阵的性质	30
4.3	弱中心投影中的核线方程式	31
4.4	基于对应点的中心投影核线方程式的线性解法	35
4.5	利用对应点确定仿射投影中的核线方程式	39
4.6	练习题	40
第 5 章 基于弱中心投影影像的 3 维重建		42
5.1	基于 3 张弱中心投影影像的运动与形状恢复	42
5.2	基于奇异值分解利用影像序列进行运动与形状的复原	49
5.3	利用规一化相关实现密集的形状重建	51
5.4	练习题	55
第 6 章 相机标定		56
6.1	基于已知 3 维形状的标定	56
6.2	基于平面图案的相机标定	59
6.3	基于 Kruppa 方程的相机自标定	61
6.4	练习题	64
第 7 章 基于中心投影影像的 3 维重建		66
7.1	基于本质矩阵的运动与形状恢复的线性算法	66
7.2	运动与形状的最优化计算	68
7.3	练习题	74
第 8 章 基于多视数据的物体整体建模		75
8.1	2 视点 3 维数据的综合	75
8.2	多视点 3 维数据的综合	78
8.3	基于多视影像的 3 维整体模型的直接复原	79
8.4	练习题	82

第 9 章 3 维形状的三角网表示	83
9.1 2 维点集的 Delaunay 分割	83
9.2 3 维点集的 Delaunay 分割	86
9.3 基于影像中特征点的可见性信息的 Delaunay 分割	88
9.4 练习题	89
第 10 章 渲染	90
10.1 漫反射与镜面反射	90
10.2 纹理映射	91
10.3 练习题	95
第 11 章 基于影像的渲染	96
11.1 QuickTime VR	96
11.2 Lumigraph, Light Field 和同心拼接	98
11.3 练习题	101
附录 A 向量和矩阵的微分	102
附录 B 逆矩阵及伪逆矩阵	104
附录 C 特征值分解	107
附录 D 奇异值分解	109
附录 E 线性函数的拟合	111
附录 F 非线性函数的拟合	114
附录 G 3 维重建中 Marquart 法的快速算法	119
附录 H 利用 VRML 实现 3 维模型的表示及纹理映射	121
附录 I 习题解说及答案	122
参考文献	140

第1章 从2维影像到3维模型

1.1 进入由个人拍摄的影像建立3维模型的时代

人类依靠2维影像(视网膜中呈现的影像)感知3维世界,利用计算机完成同样的事情是计算机视觉(Computer Vision, CV)领域的主要研究内容。

另一方面,根据已知3维世界(至少假设已知)的投影合成影像的研究,是计算机图形学(Computer Graphics, CG)领域的主要工作。

计算机视觉(CV)与计算机图形学(CG)的研究内容正好相反。从2维影像中提取3维世界的信息的问题在数学上是一个病态问题(ill-posed problem),计算机图形学正好相反,是根据已知3维世界的信息合成影像,这是一个良性问题。前者难度大,应用范围小,后者占有娱乐、广告、虚拟现实等商业领域,是一个 Big Business。看过电影《泰坦尼克号》的读者都会这样想,这是多么酷的CG作品啊。与CV相关的权威性的国际会议ICCV的参加者有数百人,与此相比,与CG相关的权威性的国际会议SIGGRAPH的参加者有数万人,研究人员之外的人也很多。这些数字可以说明二者的差别。

但是最近,可以看到CV在CG中的应用是一个新的发展方向,并充满活力。为了实现具体物体和环境的3维CG表示,对象的3维几何信息是必需的,在CG中以人工输入为主,效率很低,存在缺乏真实感等问题。比如,要用计算机图形学技术表示金阁寺,并在Internet上发布,则必须输入形状信息,没有CAD数据,对于比较大的物体尺寸的量测就难以进行。这时,利用CV技术,从影像中可以得到

到这些相关信息,从 CG 的输入考虑,Image-Based Modeling and Rendering (IBMR) 是 CV 与 CG 结合的领域^[8,19]. 这样,就有基于对象 3 维重建基础上的渲染方法,和基于记录全部影像光线的方法之间的差别. 本书重点处理前者,对于后者,在《第 11 章 基于影像的渲染》这一章进行了简单介绍. 3 维重建的多种方法中,利用多视影像的方法最为成熟. 这时,多视影像间的对应点信息是必需的. 从原理上讲,如果给定对应点,就可以解算出具体的 3 维表示结果,作为对应点坐标的函数,可以计算出新视点下的影像坐标^[37,52],这时,要表示这个物体,如果没有现成的 CG 工具(比如,VRML)的支持,就需要自己完成. 但是,现实中,在 3 维模型已经复原的基础上,能用 CG 表示的地方,如果利用 CG 的工具就方便得多^[67]. 比如,用 VRML 的形式表示,就可以在 Internet 浏览器 (NetScape, Internet Explore) 上显示. 渲染主要通过纹理映射得到实现.

从 2 维影像恢复 3 维运动和形状,是计算机视觉的主要目标,出现了多种方法^[10,42,60]. 这个问题可以分解为两部分:一个是对应问题,另一个是给定对应时运动和形状的计算算法的问题. 前者还有很多问题没有解决,后者已经解决得比较好.

本书中,影像间特征点的对应由人工给定,主要对随后的 3 维计算、三角网的生成、纹理影像的生成等给予详细的介绍. 这个过程基本上可以自动地完成. 影像间特征点的对应,没有确定的方法,还在摸索中.

随着技术的进步,个人电脑、数码相机也开始普及. 另外,个人电脑图形加速卡的性能和规格也在提高. 今后,对我们房间中的大部分物体,通过数码相机照相,把照片输入到计算机中,解算出 3 维形状,再贴上纹理,这样 3 维 CG 都可以自己制作,自娱自乐,甚至在 Internet 上发布,也可以出现在游戏中等. 这样的时代应该不会太远了.

本书中收录的大部分算法是作者工作所在地立命馆大学工学部情报学科计算机视觉研究室的实验结果(一部分程序(可执行文件)在 www.cv.cs.ritsumei.ac.jp 上发布). 这里使用的演示系统 (3d-mode Ver. 0.9, 试用版) 预定在 www.3d-mode.net 上发布更新版



本.

1.2 本书中使用的专门术语及数学符号

3 维形状一般通过 3 维空间中的点、线、面等来表现. 对应具体的 3 维空间, 它们的属性通过数值描述. 通过影像“计算”形状的“数值表示”, 数学的使用不可避免. 本书尽量使用初等数学的知识. 比如, 很多人不习惯投影几何学的表达, 因此本书中可以使用大学中都学习过的线性代数的描述方法来回避. 另外, 必要数学知识的描述包含在附录中, 如果必要可以先行阅读, 然后理解本书就不成问题了.

市面有关于线性代数计算之类的软件,^[1] 另外还有公开出版的程序集.^[2] 用户没有必要亲自去实现线性代数计算的程序, 只需要将问题用线性代数的形式表示, 结果就可以很快计算出来. 因此, 本书中, 没有用几何学概念的语言描述, 而尽量使用线性代数的表现手法, 如果能够完全理解本书, 可以学会一些实用的线性代数“技术”.

本书中使用到的专业术语的中英文对照如下:

pixel coordinate	数字影像坐标
normalized coordinate	规一化影像坐标
normalized camera	规一化相机
intrinsic matrix	相机内部矩阵
intrinsic parameters	相机内部参数
principle point	像主点
extrinsic matrix	外部矩阵
extrinsic parameters	外部参数
epipolar line (plane)	核线(面)
fundamental matrix	基础矩阵
essential matrix	本质矩阵

续表

homography matrix	摄影变换矩阵
perspective projection	透视(中心)投影
orthographic projection	正射投影
weak perspective projection	弱透视(中心)投影
affine projection	仿射投影
camera calibration	相机标定
self-calibration	自标定

本书中,一般情况下,2 维影像坐标用小写,3 维空间坐标用大写,向量用小写的黑体字表示,矩阵用大写的黑体字表示.例外的是,3 维空间坐标向量用大写字母表示.另外,冠以“~”的量表示齐次坐标(扩展向量).例如 $\tilde{x} = [x^T, 1]$.

需要表示多张影像的时候,经常用'来表示第二张影像,用"表示第三张影像.

表 1.1 列出了本书中的通用符号.

表 1.1 符号列表

$\cong$	不考虑缩放比例的情况下相等
$\times$	两个向量的叉乘,比如 $x \times y$
T	向量或矩阵的转置
$^{-1}$	矩阵的逆矩阵
$^{-T}$	矩阵的逆矩阵的转置
$^{+}$	矩阵的伪逆
$\ \cdot \ $	向量的模或矩阵的 Frobenius 模
$\mathbf{0}$	所有元素为 0 的向量

续表

A	3×3 的相机内部矩阵
D	4×4 的3维 Euclidean 变换矩阵
e	影像上的极点
E	3×3 的本质矩阵
f	相机的焦距
F	3×3 的基本矩阵
F_A	3×3 的仿射基本矩阵
H	Hessian 矩阵或投影变换矩阵
I	单位矩阵
m	影像点坐标向量
M	3维空间点坐标向量
P	3×4 的相机投影矩阵
P_A	3×4 的仿射相机投影矩阵
R	3×3 的旋转矩阵
Σ	由奇异值构成的对角线矩阵
t	3×1 的平移向量
U	奇异值分解得到的左正交矩阵
V	奇异值分解得到的右正交矩阵
O	所有元素为0的矩阵



第 2 章 影像、相机及投影

本章叙述影像、相机及投影. 作为一般的输入数据, 这里所说的影像, 一般都是指数字影像. 影像是 3 维空间在相机中的投影. 相机成像模型, 一般采用针孔模型 (pin hole). 采用针孔成像模型的相机投影, 称为中心投影. 相机具有像主点以及焦距等内部参数. 相机坐标系与世界坐标系间的旋转和平移, 称为外部参数. 另外, 平行投影、弱中心投影、模拟中心投影、仿射投影等作为中心投影的近似, 也经常被用到. 以上所述的投影、内部参数、外部参数等在本书中都以线性代数的形式给予简洁的表达.

2.1 数字影像与数字影像坐标系

作为一般的输入数据, 这里所说的影像都是指数字影像. 数字影像由 2 维的像素 (pixel) 阵列构成. 现在, 典型的影像尺寸是 640×480 像素. 一般, 每个像素包含 8bit 的信息量, 可以表示 $0 \sim 255$, 共 256 个不同的数值. 另外彩色影像, 每个像素有红 (R) 绿 (G) 蓝 (B) 三种颜色. 三种颜色分别形成一幅影像. 这样, 它就具有单色影像三倍的数据量.

数字影像的坐标系经常以图 2.1 所示的形式表示. 坐标原点在左上角, 横轴 u 的正方向向右, 纵轴 v 的正方向向下. 各像素由整数坐标值 (u, v) 表示. 这样的坐标系称为数字影像坐标系, (u, v) 称为数字影像坐标 (pixel coordinate).

利用数字影像坐标系表示实际的影像, 使用很方便, 对于投影关系的表达, 利用 2.4 节叙述的规一化影像坐标系比较方便.

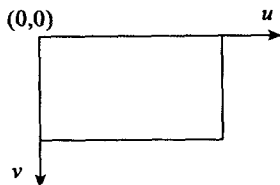


图 2.1 数字影像坐标系的坐标原点在左上角, 横轴 u 的正方向向右, 纵轴 v 的正方向向下

2.2 针孔相机与中心投影

图 2.2 表示针孔 (pinhole) 相机的成像过程. 在到影像平面 I 的距离为 f 的地方放置一个与其平行的平面 F , 它上面有一个小孔, 即点 C . 从物体上来的光线经过小孔, 即点 C , 在影像平面上成像. 物体上的点、小孔、影像平面上的像点在一条直线上. 这种投影称为中心投影 (perspective projection). 点 C 是镜头中心 (focal point), 也称为焦点 (focus), 平面 F 是焦平面 (focal plane), 从镜头中心到影像平面的距离 f 称为焦距 (focal length). 通过点 C 与影像平面垂直的线叫光轴 (optical axis), 光轴与影像平面的交点 c 称为光轴点 (principal point). 毫无疑问, 光轴也同样垂直于焦平面. 这种模型可以正确地描述一般 CCD 相机的成像过程.

下面首先定义描述中心投影方程式的坐标系. 定义影像坐标系的坐标原点为光轴点 c , x 轴和 y 轴分别与相机像素的配置轴重合, 如果将光轴作为第三轴, 其方向可以由右手法则确定. 接下来定义 3 维空间中的坐标系: 将焦点 C 作为坐标原点, 光轴为 Z 轴, X 轴和 Y 轴分别与影像坐标系的 x 轴和 y 轴平行, 方向相反. 这样的话, 符合右手法则, 右手系的旋转矩阵使用起来比较方便. 这种 (C, X, Y, Z) 坐标系称为相机坐标系.

在上述坐标系下, 空间点的 3 维坐标与 2 维影像点坐标之间存在下述的关系:

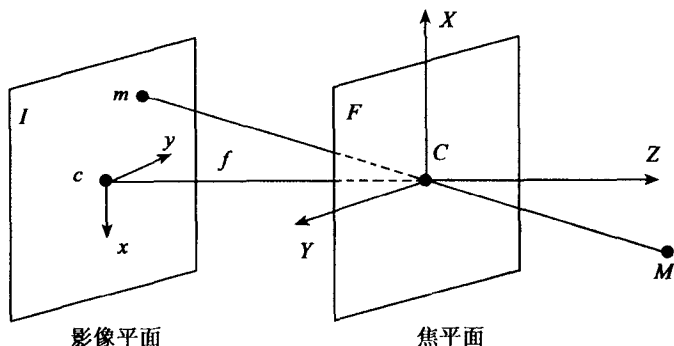


图 2.2 针孔相机模型. 针孔相机的投影为中心投影. 3 维空间坐标系的原点在相机的镜头中心, 像平面在镜头中心的后方

$$\left. \begin{aligned} x &= f \frac{X}{Z}, \\ y &= f \frac{Y}{Z}. \end{aligned} \right\} \quad (2.1)$$

如图 2.3 所示, 影像平面从焦平面的后方移到焦平面的前方, 并反转, 上述的关系仍成立. 计算机视觉中, 更多地使用这种表示方法, 本书以后都使用图 2.3 中所示的坐标系. 注意, 这里影像上的点 (x, y) 在相机坐标系中表示为 (x, y, f) .

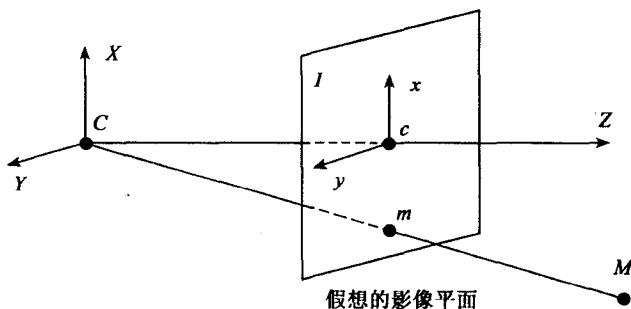


图 2.3 将假想的影像平面放在镜头中心之前的针孔相机模型. 一般情况下都使用该模型