

从二维系列摄影图片提取剪影重构 三维实体的光线跟踪算法

陈越 余峰

(浙江大学信息工程学院计算机系, 杭州 310027)

摘要 给出了利用二维系列摄影图片提取剪影重构三维实体的一种新算法,即围绕一个三维实体从多个角度拍摄照片,然后从照片中提取出实体的边界,通过基于光线跟踪的、对二维影像的合成得到原物体近似表面的三维点坐标,与传统的基于体像素的算法相比,该算法节省存储空间,近似精度与三维分辨率无直接关系,而且速度有所提高,特别当实体体积较大时,效果明显。

关键词 三维重构 光线跟踪 边界提取 体像素 剪影 掩模

中图法分类号: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2002)08-0806-08

Ray-Tracing Algorithm for Reconstruction of 3D Objects Using Silhouettes Subtracted from a Series of 2D Photos

CHEN Yue, YU Feng

(Department of Computer Science & Engineering, Zhejiang University, Hangzhou 310027)

Abstract This paper presents a new algorithm for reconstruction of a 3-dimensional object using silhouettes subtracted from 2-dimensional photos. The basic idea is to shoot a series of photos around a 3-dimensional object from different angles, subtract the boundaries of the object from the photos, and then combine the 2-dimensional information by tracing the rays to obtain the 3-dimensional coordinates of the points on the approximation surface of the object. Compare to some available method based on the concept of "voxel", our algorithm has the following advantages: (1) It significantly saves space because only the 2-dimensional mask boundaries need to be saved in memory; (2) The accuracy of approximation is independent of the 3-dimensional resolution since we are not using voxels; and 3. the speed is also greatly improved especially when the object has large volume since instead of projecting every voxel, only the points on the surface are computed according to our method.

Keywords 3-D reconstruction, Ray-tracing, Boundary subtraction, Voxel, Silhouette, Mask

0 引言

三维重构是计算机图形学中具有广泛应用背景的重要课题之一,其应用包括:机器人加工控制——在加工一个零件的毛坯之前,机器人必须先建立毛坯的三维模型;计算机辅助设计——一个自然三维实体模型的自动生成,可作为CAD系统的输入;动

画制作——为动画制作者提供大量真实世界的模型,等等。目前三维重构的处理方法大致可分为两大类:一类是动态重构,利用三维激光扫描仪等仪器得到的物体表面的三维点坐标,这种方法的局限性在于:扫描对象的尺寸有限,无法应用于大型实体,如建筑物、火箭、航天飞机、山川等等,另外,目前成本还比较高,且要求用户具备一定操作水平;另一类是静态重构,利用实体的二维几何与拓扑信息,重构其

基金项目:教育部优秀青年教师基金(2000)

收稿日期:2001-04-27;改回日期:2001-12-10

三维信息,例如由三视图重构的方法及由透视图重构对称形体的算法等^[1]。与动态方法相比,静态重构对设备要求不高,且从原理上讲,不受空间尺寸的限制,但算法相对复杂一些。

从二维系列摄影图片重构三维自然实体问题的研究在计算机视觉领域已经不是一个崭新的话题,它通常包括相机标定、对应点匹配、以及三维重构3个部分。本文重点研究三维重构的方法。

三维重构算法^[2]是这个领域的经典算法之一,它解决了在已知对应点和相机内参数的情况下,重构物体的问题;另一种经典算法则通过“体像素”(voxel)在二维图象上的投影来判断物体所在空间位置^[3~7]。

对于第1种算法,在空间取任意一点 A ,在相机 C 中观察到其像点 a ,则 A 必定在 Ca 连线上的某一点。如果同时用相机 C 、 C' 观察同一点 A ,则 A 必定在 Ca 和 $C'a'$ 连线的交点上,它的空间位置是唯一的。当提供多幅图的时候,空间点的位置受到了更多的约束,所得的解会更精确。但这种算法需要对应点匹配,即对于用相机 C 观察到的 A 的像点 a ,要在用相机 C' 观察到的图象中找到与其相对应的像点 a' 。而通用的匹配算法至今没有很好的解决方案,在实际的使用当中,通常利用特定的相机位置和图象位置来简化匹配的问题。

对于第2种算法基本可归结为利用“体素”或类似的概念,将实体所在空间,以类似于平面像素的方法划分为立体像素,将掩模与视点确定的投影锥内的体像素记为黑,外部体像素记为白,对所有不同角度获得的投影锥进行AND运算,这样就得到由黑色体像素构成的三维实体,但算法的缺点是时间和空间复杂度很高,体像素空间的分辨率直接影响近似精度和空间复杂度。自20世纪80年代以来,人们对体像素方法进行了不断改进,又提出了基于上述思想的四叉树、八叉树算法等等,但时空复杂度并没有十分显著的改善。

本文给出的算法与经典的体像素思想不同,因为其目的是得到物体表面的点坐标,从而构造三维(封闭)曲面的模型,所以其并不需要物体内部的体像素,而只需考虑如何得到表面点,在计算中,也只需要掩模的边界,即空心的剪影,这样不仅所耗的存储空间大大节省,近似精度与三维空间分辨率也无直接关系,而且时间复杂度一般情况下,优于经典的体像素方法。

1 算法基本原理

算法可分为两个模块,即剪影边界的提取和通过光线跟踪计算近似表面点。

1.1 剪影边界的提取

照片摄取的过程可以看成是物体固定在桌面上,而照相机绕着物体旋转一周进行拍照。

重构计算时,以剪影为底面(剪影可以是多连通域)、以其对应的照相机的焦点为顶点,可确定一个锥体,称为投影锥。将锥体向桌面方向延伸,其表面应与三维物体外切,如图1所示。



图1 相机围绕物体拍照

因只需找到切点的近似位置,所以不需要把实心锥体进行延伸,而只要把锥的母线进行延伸即可。这就需要首先得到投影锥的准线——剪影的边界轮廓线。

轮廓线追踪的算法^[4]可用于从实体的黑白掩模中提取边界像素,并用点列记录。这样得到的点列还是有序的(顺时针或逆时针),这一点对于后面的求交计算很重要。

另外由于算法的复杂度与每张剪影轮廓上的平均取点数的平方成正比,而且照片数一般不是很大(24张、36张、或72张),轮廓点取得太密不利于重构后的三角剖分,所以并不是取的点越密越好,而是用尽可能少的点表示出剪影轮廓的特征,即在曲率变化大的地方多取一些点,反之则少取一些点。

1.2 光线跟踪

一般设桌面所在平面为 $x-z$ 平面,与桌面垂直的旋转轴为 y 轴,方向满足右手系关系,此坐标系称为“参照坐标系”。

为了便于直观地解释算法,从二维问题开始观察。

图2(a)显示从一个焦点延伸出的投影“锥”,4条射线是二维剪影的4个边界点所对应的光线。

图2(b)显示两个投影锥相交的情况,以第1个

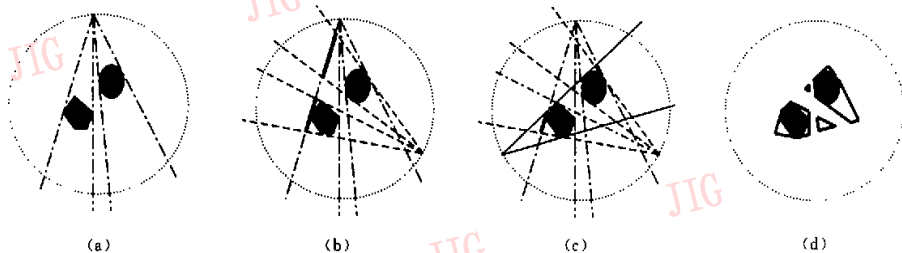


图2 光线跟踪法的2维情形

锥的最左边一条射线为例,它与第2个锥的交集是图中用粗线表示的两条直线段。事实上,图中的8条射线均被另一个锥切割成两条线段,而物体必然在这些线段所包围的4个连通域中。

图2(c)是第3个锥加入后的情形。仍以第1个锥的最左边一条射线为例,它与第2、3个锥的共同交集是图中用粗线表示的一条直线段,也是3个锥的交集区域的一条边,线段的两端点即是结果多边形的两个顶点。同样的,每个锥的边界射线都被其他的两个锥切割,而物体必然在这些交集线段所围成的数个连通区域内,如图2(d)所示,所以每条线段的端点就是当前得到的近似多边形的顶点。

三维情形的道理是类似的:第 k 个焦点在参照坐标系中的位置记为 f_k ,它和掩模的离散化边界上两个相邻的点 $P_{k,j}$ 和 $P_{k,j+1}$ 形成一个楔形平面。所有从 f_k 延伸出的楔形平面共同形成一个(或多个,若剪影非单连通)棱锥面,即是投影锥的一个分段线性近似(如图3所示),其近似精度由离散精度确定。

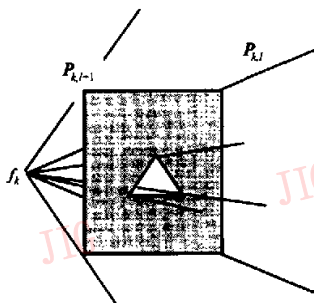


图3 由楔形平面构成的2个棱锥面

每一个焦点对应一个锥,三维物体即在这些锥相交的区域中,即如图1所示。事实上,这些棱锥面相交的结果,就是由小多边形组成的物体表面的一个近似,每个小多边形都是某个楔形平面的一部分,

与物体至少在一个点处外切。

与体像素计算的观点不同,其不考虑物体内部的点,而直接计算构成物体近似外表面的多边形的顶点坐标,所以计算速度得以提高。对于从第 i 个焦点 f 出发,经过其剪影上第 j 个边界点 $P_{i,j}$ 的射线,求它与所有焦点 $f_k(k=1,2,\dots, \text{且 } k \neq i)$ 延伸出的投影锥的交集,得到该射线上的一系列直线段。当只有一个交点时,可以看成是两个相同的交点,即一段退化的线段(实际上,由于计算误差的存在,相同的点几乎是不可能的)。很明显,不可能没有交点,而这些直线段的交集一定与三维物体相切,也必定是构成物体近似外表面的多边形的一条边。这样就得到了多边形的两个顶点坐标,或者也可输出线段中点并作为近似点。

2 算法描述

2.1 算法

输入:

- (1) n 张系列照片;
- (2) 照相机的焦距 f 和焦点相对于参照坐标系的坐标 $f_i = (f_x[i], f_y[i], f_z[i])$ ($i = 0, 1, \dots, n-1$);
- (3) 照相机的3个旋转角:仰视角 α_i (以 x 轴为轴旋转的角度)、旋转角 β_i (以新的 y 轴为轴旋转的角度)、偏转角 γ_i (以新的 z 轴为轴旋转的角度), ($i = 0, 1, \dots, n-1$)。

输出:

物体近似表面点的三维坐标。

算法:

- (1) 读入输入数据,对于每一张照片,执行:
 - ① 取一阈值将照片二值化,以区分物体剪影与背景;

其中, M^T 是旋转矩阵的转置, $P_{c,i,j} = (x_{c,i,j}, y_{c,i,j}, -f)^T$.

3个坐标系之间的关系如图5所示.

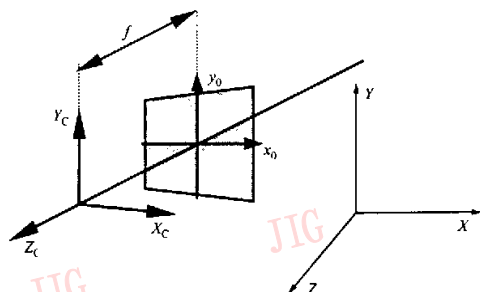


图5 参照坐标系(X, Y, Z), 相机坐标系(X_c, Y_c, Z_c), 照片坐标系(x_0, y_0)

2.5 求射线与楔形平面的交点

首先考察从 f_i 出发经过 $P_{i,j}$ 的射线, 用参数式表达可写为

$$\mathbf{x} = \mathbf{f}_i + t_{i,j} \mathbf{u}_{i,j} \quad (3)$$

其中, $\mathbf{u}_{i,j} = \mathbf{P}_{i,j} - \mathbf{f}_i$ 是此射线的方向向量, $t_{i,j} \in [1, +\infty)$ 是参数. $(t_{i,j}^{(a)}, t_{i,j}^{(b)})$ 对应射线上一段直线段, 也即物体表面点所在的区间. 通过射线与多幅照片求交, 可得若干直线段, 而求所有这些直线段的交集线段, 便可逐步精确表面点所在的范围. 在初始化时, 直线段链表置为 $[1, +\infty)$.

其次, 分别考察从 f_k 延伸出的各个锥, 计算射线与这些锥的交点. 从 f_k 延伸出的锥的第 l 和 $l+1$ 个点对应的楔形平面方程为

$$\mathbf{x} = \mathbf{f}_k + r_{k,l} [\mathbf{u}_{k,l} + s_{k,l} (\mathbf{u}_{k,l+1} - \mathbf{u}_{k,l})] \quad (4)$$

其中, $r_{k,l} \in [1, +\infty)$, $s_{k,l} \in [0, 1]$ 为参数.

求射线式(3)与平面式(4)的交点, 即令

$$\mathbf{x} = \mathbf{f}_i + t_{i,j} \mathbf{u}_{i,j} = \mathbf{f}_k + r_{k,l} \mathbf{u}_{k,l} + r_{k,l} s_{k,l} (\mathbf{u}_{k,l+1} - \mathbf{u}_{k,l}) \quad (5)$$

从中解出3个参数的值. 问题化为求解下列线性方程组:

$$\begin{bmatrix} u_x[k, l+1] - u_x[k, l] & u_x[k, l] & -u_x[i, j] \\ u_y[k, l+1] - u_y[k, l] & u_y[k, l] & -u_y[i, j] \\ u_z[k, l+1] - u_z[k, l] & u_z[k, l] & -u_z[i, j] \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} f_x[i] - f_x[k] \\ f_y[i] - f_y[k] \\ f_z[i] - f_z[k] \end{bmatrix} \quad (6)$$

其中, $\mathbf{v} = (v_1, v_2, v_3)^T = (r_{k,l} s_{k,l}, r_{k,l}, t_{i,j})^T$. 用 Gauss 消去法首先解得 $v_3 = t_{i,j}$, 判断其是否大于1, 若是, 则继续解 $v_2 = r_{k,l}$, 再判断是否大于1, 若是, 则继续解 $v_1 = r_{k,l} s_{k,l}$, 判断 $v_1/v_2 = s_{k,l}$ 是否在 $[0, 1]$ 内, 若是,

则得到一合法交点, 记录 $t_{i,j}$. 若其中任何一个判断不满足, 即中断求解方程组, 表示射线与平面的交点不在楔形区域内.

在整个轮廓线上的楔形平面求交之后, 可得若干区间. 对于表面处处为凸的物体, 射线与每个投影锥至多在两个楔形面上相交, 因此对于从 f_k 延伸出的锥只得到一个区间, 记为 $(t_{i,j}^{(1)}, t_{i,j}^{(2)})$. 而对于一般的物体, 可能会得到两个或两个以上的交点, 排序后得到区间 $(t_{i,j}^{(1)}, t_{i,j}^{(2)}), (t_{i,j}^{(3)}, t_{i,j}^{(4)}), \dots$.

3 性能分析与实验

3.1 复杂度分析

由上述算法描述可以看出, 当选择合适的拍摄角度, 且当 m_i 和 n 无限大时, 计算结果应能无限接近精确表面的点. 但如果拍摄角度不对, 使表面某些区域的点不能从掩模边界看到, 则无论是经典的基于体像素的算法, 还是光线跟踪法, 仅从二维系列照片出发的重构方法都不能奏效. 两种方法的计算量主要取决于物体表面积和体积的大小.

从算法的描述可以看出, 光线跟踪的核心是对 $\sum_{i=0}^{n-1} m_i$ 条射线, 求每条射线与 $\sum_{k \neq i} m_k$ 个楔形平面的交点, 算法最坏情况下的时间复杂度约为 $O(m^2 n^2)$, 空间复杂度为 $O(mn)$, 其中, n 是照片个数, 实际应用中, 一般为24张或36张、72张; m 是每张照片上取的剪影离散轮廓点个数的平均值. 注意, m 并非越大越好, 如果 n 的大小有限, m 太大会给随后进行的三角剖分带来负面影响.

经典的 voxel 算法需要存储所有照片的每个像素值, 而不只是剪影轮廓的像素值, 其空间复杂度比文中算法高一个阶. 另外, 体像素空间的分辨率直接影响结果的精度, 其时间复杂度约为 $O(nV)$, 这里 V 是给定分辨率下, 体像素空间的体积.

在相同的精度要求下, 文中算法在空间方面的优势是明显的. 对于体积较大的物体, 即 $V \gg m^2 n$ 时, 速度上也有比较明显的优势, 而且多边形顶点的产生有一定的规律, 使得很容易通过三角剖分, 一次性生成物体的全封闭表面模型.

3.2 实验结果

实验对象包括球体(因其为典型的表面处处为凸的单连通体, 且其剪影可以人为生成, 是验证算法正确性最经济的对象)、圆柱体(因其上下表面与侧

面衔接不光滑,且并非严格凸,可以检验该算法的可推广性)、戒指(因其为带洞的物体,可以检验算法处理非单连通情况的效果)、以及埃菲尔铁塔(因其结构复杂,可以检验算法的健壮性)。

实验硬件环境:CPU:PⅢ 733,内存:128MB

3.2.1 球体重构

取每张照片上的掩模为 1 个单位半径的圆,照相机的焦距为 10 个单位,照片与 y 轴平行,距 y 轴为 20 个单位,则可精确计算出对应的三维球体的半径为 1.990 74 个单位。以三维球体的球心为三维坐标系的原点,照相机绕球心在 x-z 平面旋转一圈,间隔相同的角度拍摄 1 张照片。

计算结果取每条线段的中点,并判断该点到球心的距离是否在 $(1.990\,74-0.01, 1.990\,74+0.01)$ 范围内,即将绝对误差在 0.01 范围内的点视为合格点,并计算平均相对误差,记录其计算时间。

表 1 球体重构的精度与速度

照片数	每张照片取点数	总点数	合格率(%)	平均相对误差	重构时间(ms)
24	100	2 400	99.708	0.000 719	4 056
36	100	3 600	100	0.001 109	9 464

从实验数据可见,该算法生成 2 400 个点用时约 4.056s,3 600 个点用时约 9.464s,运算时间对于计算点数基本上呈平方增长关系,即 $2\,400^2:3\,600^2\approx 4.056:9.464$ 。以 2 400 点的用时可推算生成 5 000 个点的用时约为 $(5\,000/2\,400)^2\times 4.056\approx 17.6\text{s}$;以 3 600 点的用时可推算生成 5 000 个点的用时约为 $(5\,000/3\,600)^2\times 9.464\approx 18.3\text{s}$ (重构结果见图 6)。

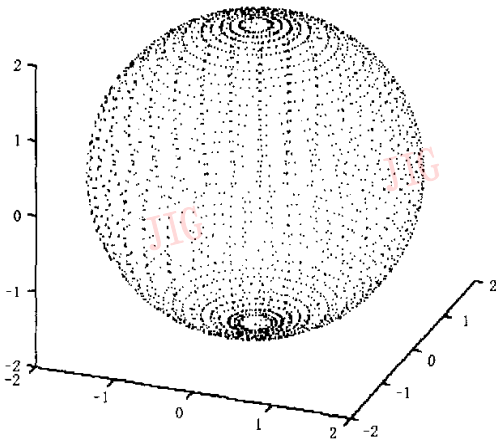


图 6 用 36 张剪影重构的球体

而 voxel 算法^[2]中给出的实验数据表明,对一

只玩具熊的 36 张照片进行处理,二维分辨率为 $640\times 480\text{ pixel}$,体像素分辨率为 $320\times 320\times 240\text{ pixels}$,产生 5 000 个点最少用时 102s(CPU: Pentium Ⅱ 400MHz, Memory: 128MB)。而本实验用机 CPU 的速度约是它的 1.8 倍,用时约是它的 1/5.6,可见对于球体这种体积明显大于表面积的物体,本算法的速度优于 voxel 算法。重构结果见图 6。

3.2.2 圆柱体重构

取照相机焦距为 20 个单位,照片距旋转轴为 20 个单位,假设三维实体圆柱的底半径为 1.0 个单位,高 4.0 个单位。以圆柱体的中心为三维坐标系的原点,照相机绕中心在 x-z 平面旋转一圈,间隔相同的角度拍摄照片。计算结果取每条线段的中点,并判断该点到 z 轴的距离是否在 $(1.00-0.01, 1.00+0.01)$ 范围内,即将绝对误差在 0.01 范围内的点视为合格点计算对于圆柱体底面半径的平均相对误差,并记录计算时间。

表 2 圆柱体重构的精度与速度

照片数	平均取点数	总点数	合格率(%)	平均相对误差	重构时间(ms)
24	80	1 920	89.74	-0.001 824	3 134
36	80	2 880	99.97	-0.001 009	7 110

实验运算时间与计算点数之间的关系仍然近似满足平方关系,即 $2\,880^2:1\,920^2\approx 7.110:3.134$ 。以 1 920 点的用时来推算生成 5 000 个点的用时约为 $(5\,000/1\,920)^2\times 3.134\approx 21.3\text{s}$;以 2 880 点的用时来推算生成 5 000 个点的用时约为 $(5\,000/2\,880)^2\times 7.11\approx 21.4\text{s}$,用时约是文献[2]中实验用时的 1/4.8,仍然优于 voxel 算法。重构结果见图 7。

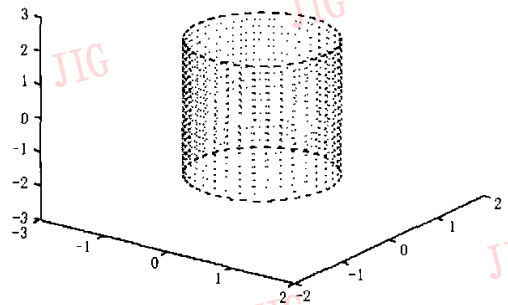


图 7 用 36 张剪影重构的圆柱体

3.2.3 戒指重构

从图 8 可以看到,由于照片中并没有去除阴影,使掩模中多出了阴影部分,这必然影响结果。修正的方法是选取少量照片作为“过滤器”,将计算所得的表

面点向过滤器中的照片投影,若投影点不在照片掩模内,则删除该点.实验数据如表 3.

表 3 戒指重构的精度与速度

照片数	分辨率 (pixel)	过滤用 照片数	处理照片 用时(s)	重构时间 (ms)	得到表面 点数
36	320×240	12	1 166	12 628	1 820

3.2.4 埃菲尔铁塔重构

埃菲尔铁塔重构实验结果如图 9 及表 4.

表 4 埃菲尔铁塔重构的精度与速度

照片数	分辨率 (pixel)	过滤用 照片数	处理照片 用时(s)	重构时间 (ms)	得到表面 点数
36	320×240	12	1 312	33 338	2 251

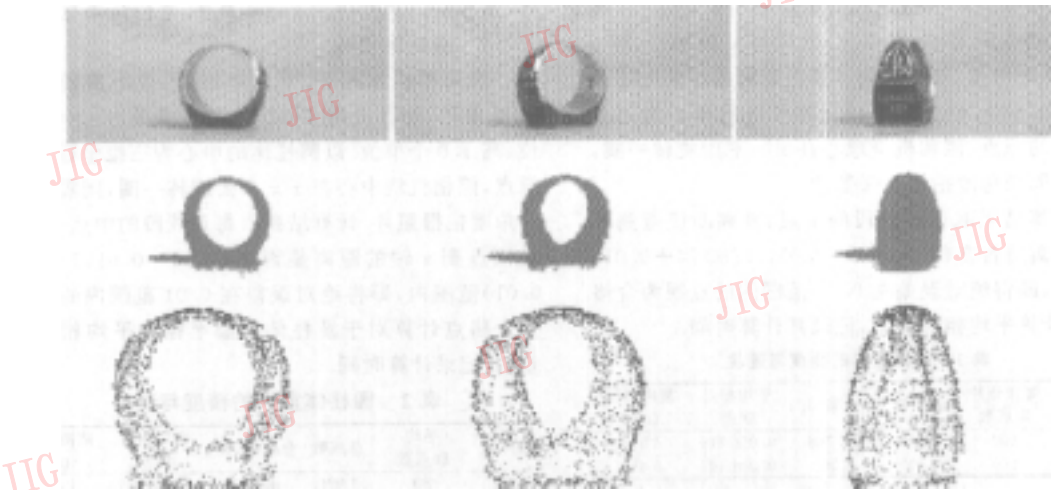


图 8 用 36 张剪影重构的戒指

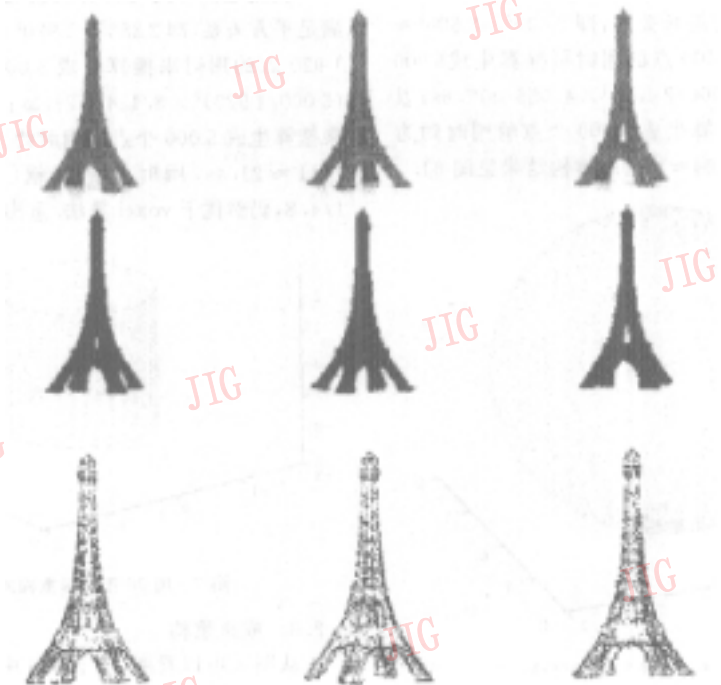


图 9 用 36 张剪影重构的埃菲尔铁塔

4 结 论

光线跟踪算法能快速地从小系列二维剪影中计算出物体表面点的信息。通过实验可以看到,对于凸的单连通以及非单连通物体,算法都能得到较好的效果。

该算法对输入数据的精度要求较高,且扰动会产生个别偏差较大的点。修正的方法是,牺牲一定时间空间,取 p 张不同角度的照片,将所得射线区间上的线段端点投影到这些照片上(使用 DLT 模型^[8])。若投影点在掩模内,则保留该点;若为背景,则舍去。

照相机的拍摄位置很重要,其视点应选取使得表面上的每一部分都可以从剪影轮廓上被看到。而对于表面有不能被剪影轮廓线体现的凹陷物体,此算法就不能给出凹表面的点,必须与三维扫描仪结合做适当的局部处理。

参 考 文 献

- 1 孙家广等编著. 计算机图形学(第3版)[M]. 北京:清华大学出版社,1998.
- 2 Xu Gang, Terai Junchi, Shum Heung-Yeung. A linear algorithm for camera self-calibration, motion and structure recovery for multi-planar scenes from two perspective images[A]. In: Proc of IEEE Conference on Computer Vision and Pattern Recognition [C], South Carolina, June 2000, II:474~479.
- 3 Yukinori Matsumoto, Tsukuba Ibaraki, Kouta Fujimura. CyberModeler: A compact 3D scanner based on monoscopic camera[J]. SPIE, 1999, 3640-02:2~10.
- 4 Niern Wolfgang. Robust and fast modeling of 3D natural objects from multiple views[J]. SPIE 1994, 2182:388~397.
- 5 Szeliski R. Rapid octree construction from image sequence[J]. Comp. Vision Graphics Image Processing (CVGIP): Image Understanding, 1993, 58(1):23~32.
- 6 Potmesil M. Generating octree models of 3D objects from their silhouettes in a sequence of images [J]. Computer Vision, Graphics, and Image Processing, 1987, 40(1):1~29.
- 7 Chien C H, Aggarwal J K. Identification of 3D objects from multiple silhouettes using quadrees/octrees [J]. Computer Vision, Graphics, and Image Processing, 1986, 36 (2/3): 256~273.
- 8 Sabel J C. Calibration and 3D reconstruction for multicamera marker based motion measurement [M]. Netherlands, Delft University Press, Nov, 1999.



陈 越 1967 年生,副教授,1996 年获美国 University of Connecticut 数学系博士学位。目前主要研究方向为自由曲面造型理论。

余 峰 1979 年生,硕士研究生,2001 年获浙江大学计算机系学士学位。目前主要研究方向为三维成像及数字图象处理。