

# 基于图像序列的虚拟场景重建和漫游

吴琼玉 周东翔 蔡宣平

(国防科技大学电子工程与科学学院联合研究中心,长沙 410073)

**摘 要** 基于图像的建模和绘制方法与基于传统的几何绘制方法相比有很多优点,但是在场景几何参数未知的情况下,要实现单幅图片和图像序列的漫游,难度很大。TIP(tour into a picture)技术使得在一幅图像中漫游成为可能。为了实现场景几何参数未知情况下的场景漫游,在对 TIP 技术进行扩展的基础上,提出了一种能够在未知视点路径的图像序列中实现场景漫游的方法。这种算法在相机未定标的情况下,不仅解决了图像序列建模和场景漫游过程中前后图像场景不能平滑过渡的问题,并且扩大了 TIP 中视线方向变化的动态范围,从而在保证图像质量的情况下,实现了由图像序列到场景的无限制漫游。实际图像序列的实验结果表明,该算法是有效的,具有实际应用价值。

**关键词** 场景漫游 图像序列 图像拼接 漫游

**中图法分类号:** TP391.9 **文献标识码:** A **文章编号:** 1006-8961(2006)01-0113-06

## Construction of Virtual Environment from Image Sequences

WU Qiong-yu, ZHOU Dong-xiang, CAI Xuan-ping

(Joint Center for Intelligent Sensing and System, School of Electronic Science and Engineering, NUDT, Changsha 410073)

**Abstract** Comparing with the traditional approach of 3D model-based rendering with explicit 3D structure, the approach of image based modeling and rendering has many advantages, but both fall futile in modeling from one single picture. The method of touring into the picture, proposed by Horry in Siggraph 1997, gives us a surprising way to solve this problem. Our methodology in modeling from one picture can be recognized as an extension to TIP. Based on these extensions on TIP, we consider an image sequence taken with the unknown path of the viewpoint. The path can be a line, curve or a combination of both. Each frame of the sequence is decomposed into sub-images, which are mosaicked into a new texture to be used in the following walking through. Wherever the viewpoint is, it can get high-resolution virtual images. The experimental results show that this method is effective and applicable.

**Keywords** tour into a picture(TIP), image sequence, image mosaic, walk through

## 1 引言

虚拟场景重建的主要目的就是为了实现在一个场景中的穿梭和漫游,这种场景漫游包括改变视点和视线方向,以便获取与 3 维场景所对应的 2 维平面图像。近年来,虚拟场景重建技术受到越来越多的关注,它已成为计算机图形学、计算机视觉和图像处理等领域的热门研究课题。随着经济和科学技术的发展,这种技术必将得到更加广泛的应用。

对于一个场景的重建,所利用的图像序列有多种情况,本文所利用是相机沿任意路径运动而获取的图像序列。目前针对这种图像序列重建场景的研究还不多,它的研究主要存在以下两个难点:一是路径未知,即前后视点之间的空间关系是未知的,这将导致从一幅图像到下一幅图像场景的过渡很难完成;二是前后两幅图像是一种近似于放大缩小的关系,因而无法实现图像间场景的稠密匹配,即使相机定标,也难以获得场景的 3 维信息。

资料证明<sup>[1]</sup>,很多情况下,用图像来恢复场景是

**基金项目:**国家“863”计划项目(2002AA135220)

**收稿日期:**2004-03-30;**改回日期:**2005-03-28

**第一作者简介:**吴琼玉(1978~),女。2002 年获国防科技大学信息与通信工程硕士学位,现为国防科学技术大学电子工程与科学学院博士研究生。主要从事虚拟现实和基于图像的绘制技术方面的研究。E-mail:qy\_wumiao@sina.com

非常困难的。在现有的关于中间视图生成和场景重建的方法<sup>[2-7]</sup>中,只有文献[6]采用了 TIP(tour into a picture)技术,才实现了从一幅图片或一幅画到场景的重建和漫游,但是当虚拟视点偏离原始视点较远时,这种方法生成的图像非常模糊,这是由于单幅图片中的信息不能准确地描述离视点较远的场景,从而导致图片分辨率下降的缘故,而图像序列正好可以弥补这一缺点。鉴于此,本文避开“从运动中恢复结构”等复杂的方法,借鉴 TIP 技术,根据场景的透视投影特点,提出了基于图像序列的 TIP 建模法。

目前,关于 TIP 技术的研究已有不少报导,如 LI 等考虑用多分辨率的方法来提高生成图像的质量<sup>[8]</sup>,但是该方法局限于一幅图片;Sanghyuk Yoon 等通过图像的渐入和渐出方式来切换图像<sup>[9]</sup>,但中间图像非常模糊;而 ZHANG Xiao-hua 等提出的基于蛛网模型的拼接法<sup>[10]</sup>也只能适应场景非常简单的情况,由于其假设场景中只包含 5 个平面,所以该方法的适应范围非常窄。由于本文改用基线建模法,而不是用蛛网模型来建模,从而使方法能够适应任意图像序列,同时结合基于图像的绘制方法,先利用相邻图像来表达场景的连续性,然后将图像序列 3 维模型拼接起来,这样就可以实现前后场景的自然过渡。

## 2 TIP 技术简介

TIP 技术<sup>[5]</sup>最早由 Horry 在 1997 年的 SIGGRAPH 会议上提出来的。它的主要思想是先从图像中恢复场景的伪 3 维模型,然后在其中漫游。TIP 方法的整个流程如图 1 所示。如果要实现在一个场景中的漫游,则首先要恢复场景的 3 维模型,并获得模型的纹理图像。TIP 将整个场景模型分为前景物体模型和背景模型两个部分。为了从图像中交互获得场景的模型,Horry 提出了一种类似于蛛网的交互建模方法。交互界面如图 2 所示,图中 5 个白色的

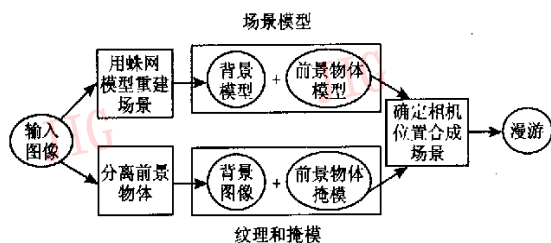


图 1 原 TIP 技术的流程图

Fig. 1 Flow diagram for the original TIP

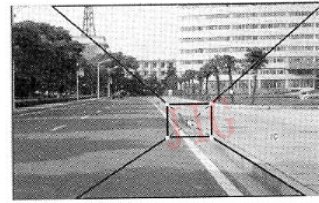


图 2 原 TIP 技术的交互界面图

Fig. 2 Interface of original TIP

点可以交互拖动,而黑色粗线(即蛛网)则将图像分为 5 个部分。用户需要交互地指定模型的内矩形和消失点,然后就可以计算出模型的其他部分。这样背景模型就被分为 5 个部分,分别对应 3 维模型中的 5 个平面,它们是上平面(又称天花板)、下平面(又称地面)、左平面和右平面以及背面。当背景模型确定后,前景物体的模型也可以交互获得,并且每一个前景物体可以看作是背景中的一个多边形,而该多边形相对各个背景平面的位置则可以通过计算得出。

## 3 建模

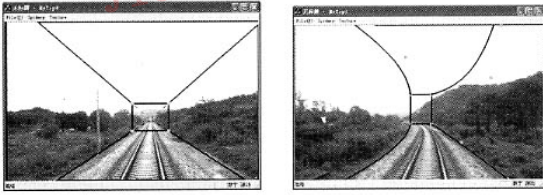
### 3.1 确定图像中心

Horry 的模型是以消失点来近似图像的中心,但当消失点偏离图像中心太远时(如图 1),则这种近似会加大模型和实际场景的偏差,进而会导致漫游时,物体在图像中变形<sup>[11]</sup>。尤其对于一些自然景观,图像中的消失线往往是很明显的,所以本文采用消失线来确定图像的中心(focus of expansion, FOE)。在根据图像的特点指定消失线后,就可以以消失线的中点为图像的中心。当视线方向和地面不平行的时候,即相机除了沿水平方向平移,还有一定的转动时,则指定的图像中心和实际的图像中心存在一定的偏差。在这种情况下,图像可以预先做一个变换,以消除相机旋转引起的偏差。消失线和图像中心在恢复场景 3 维模型的时候会用到。

### 3.2 交互建模

在本文提出的基线建模法中,假设所拍摄的场景地面是水平的。所谓基线<sup>[12]</sup>就是图像中物体和地平面的交线。如图 3 所示,图上地面上的 3 条黑色线段就称为基线。

画完基线后,再依据基线中的每一条线段(包括直线和曲线),图像就可以分为多个垂直于地面的“板”。板以上的部分成为天花板或者天空。场



(a) 基线为直线交互界面 (b) 基线为曲线交互界面

图 3 建模交互界面

Fig. 3 Interface for modeling

中所有的“板”可以认为是垂直于地面的多边形或者曲面。当基线确定后,基线的各个线段的顶点也就确定了,然后只要沿着顶点的垂直线交互拖动这些顶点,就可以确定各个模型的高度和天花板的位置。图 3 是交互界面图,图 3(a) 给出了基线为直线时的交互界面。根据场景的特点,基线可分为 3 个线段,分别对应 3 个垂直于地面的板,这 3 个板本文称它们为左平面、右平面和背面。剩下的两个平面称为地面和天空。本文的图片采用的是自然场景,又可称为室外场景。由于室内场景往往有规则的几何形状,所以基线也很容易制定,这里就不举例说明了。

由于场景并不是所有情况都可以用平面来近似,因此本文还提出了曲面的建模方法,这就使 TIP 的使用范围大大扩展。图 3(b) 是基线为曲线的交互界面。当基线是曲线时,如果采用平面建模,则可能会使地面的纹理映射到左右平面上(如图 4(a) 所示),而且无论直线怎样放置,这一情况都无法避免,所以本文考虑采用曲面的方法来弥补平面的不足,而平面也可以看作是曲面的一个特例。如图 4(b) 所示,白色的曲线准确地分割出了左平面和地面。



(a) 直线建模 (b) 曲线建模

图 4 直线模型和曲线模型比较

Fig. 4 Line model versus curve model

### 3.3 模型计算

在像平面的消失线、消失点和平面或曲面的顶点确定后,接着还要恢复各个面的 3 维空间位置。由于平面的 3 维位置完全是由控制点确定的,所以

下面简单介绍一下本文所采用的相机模型以及控制点的计算流程。

由于物体空间为 3 维的欧氏空间(用  $X, Y, Z$  表示),像平面为 2 维的欧氏空间(用  $x, y$  表示),因此,在实验中不失一般性,将坐标原点定在视点,即视点坐标为  $(0, 0, 0)$ ,像平面和笛卡尔坐标  $xoy$  平面平行, $z$  坐标的方向由右手法则确定(如图 5 所示)。

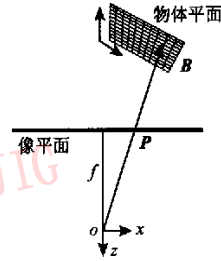


图 5 纹理提取

Fig. 5 Extracting textures

设  $B(X, Y, Z)$  为物体空间的任意一点,  $P(x, y)$  为与这点对应的像空间的像点。像平面位于坐标空间  $z = -f$  处。很明显,光轴和  $z$  轴平行,并垂直于像平面。正如图 5 中所示,  $B(X, Y, Z)$  在像平面的透视投影可以用如下的公式来表达:

$$x = f/Z \times X \tag{1}$$

$$y = f/Z \times Y \tag{2}$$

如果已知点所在的平面,那么它的 3 维坐标计算就变得非常容易。由于地面可以预先确定,因此无论是曲面模型,还是平面模型,都可以先计算在地面上的点。由式(2)算出  $P$  点的  $z$  坐标后,就可以由式(1)确定它的  $x$  坐标。这样所有控制点的 3 维坐标计算完毕后,场景的 3 维结构也随之确定。

## 4 纹理提取

在模型的 3 维结构确定后,还要给模型中的各个面加入纹理。尽管光线映射和透视变换这两种方法都可以获得相同的纹理,但由于透视变换的方法计算起来相对复杂,因此为了计算简便起见,本文采用了光线映射的方法。

光线映射的主要思想就是通过从 3 维物体空间点投影到 2 维空间的像平面来获取该点的像素值,而纹理的分辨率则由背景平面的大小决定。

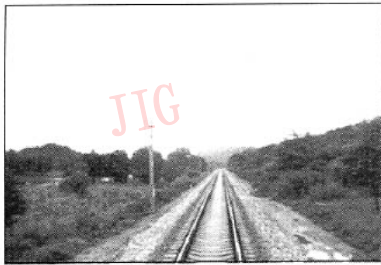
如图 5 所示,曲面或平面上一点的 3 维坐标可以表示为  $B = O + t \times v, B = (X_B, Y_B, Z_B)^T, O$  是视

点,  $\mathbf{v} = (v_x, v_y, v_z)^T$  是直线  $\overrightarrow{OP}$  的方向矢量,  $t$  是参数。假设焦距为  $f$ ,  $P$  点的坐标为  $\mathbf{P} = (x_p, y_p, z_p)^T$ , 当  $Z_B = -f$  时, 若可计算出  $t$ , 就可以得出  $X_B, Y_B$ 。即  $P$  的坐标为  $(X_B, Y_B, -f)$ , 那么  $B$  点的像素值就由  $P$  点的像素值确定。当平面上或曲面上所有的点都遍历后, 该面的纹理图像就提取到了。曲面的表达和遍历方法可以参考文献[13]。

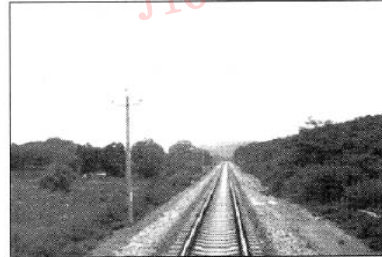
## 5 模型的整合和纹理的拼接

图 6(c)、图 6(d) 是从图 6(a)、图 6(b) 中提取

到的右平面纹理图像。从这幅纹理图像不难看出, 右边纹理表示的近景非常清晰, 而纹理越靠近左边, 就越模糊。这种模糊的纹理是导致漫游图片模糊的主要原因, 而且当视点靠近场景背面的时候, 图像会模糊到无法接受的程度<sup>[5]</sup>。这也是 Horry 算法的主要局限。比较两幅纹理图像不难发现, 图 6(d) 纹理是图 6(c) 纹理的一个补充, 也就是说, 右边纹理中最左边的部分, 在图 6(c) 纹理中是中间某个清晰度不高的地方。如果将两幅纹理拼接, 则可以获得更加清晰的纹理。依此类推, 就可以用拼接的方法来获得从左到右都清晰的纹理。



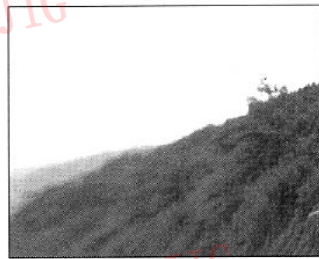
(a) 原图像 1



(b) 原图像 2



(c) 纹理 1



(d) 纹理 2



(e) 拼接纹理

图 6 场景重建

Fig. 6 Construct virtual environment

纹理拼接后, 由于整个模型也发生了变化, 因此还需进行模型整合。模型整合的思想就是将图像序列生成的单个模型依据位置关系串联起来, 以便获得一个整体模型(如图 7 所示)。由于模型中的左右平面位置具有灵活性, 因此要使得模型可以拼接必须满足一定的条件。这个条件就是前后图像的左右平面和地面的交线是连续的, 即左右平面的基线是连续的。事实上, 模型只是场景的一种近似表达, 而真实的场景往往非常复杂, 不是用几个平面就能描述的。这种不准确性和建模的随机性, 可使场景中连续的一条线变得不再连续。这样, 一条连续的线在两个模型中有可能不能很好地连接, 虽然它们

之间可能有一个夹角。但是当这个夹角很小时, 不会影响实验结果。

当图像序列所包含的场景没有很大的跳变时, 如果前后图像中的基线是连续的, 那么模型就可以拼接。如果图像序列中出现场景内容不连续的情况, 则可以把它切割成分段连续的若干段。总之, 基线建模法能保证整个模型的连贯性。当模型的左、右、上、下 4 个平面依次拼接后, 新模型的背面就可由前一模型的背面直接代替。

下面以模型的左平面为例, 简要描述计算新模型的流程。设前后相邻模型 1、2 的左平面纹理图像宽度分别为  $w_1$  和  $w_2$ , 拼接后的纹理宽度为  $w_3$ 。由

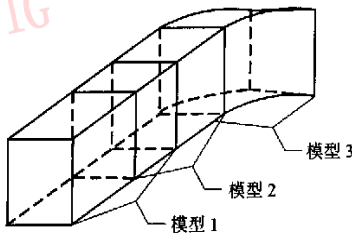


图 7 模型的拼接

(第 1、2 个模型为平面模型,第 3 个模型为曲面模型)

Fig. 7 Model concatenation

(model 1 and 2 are plane model, model 3 is curve model)

于纹理的采样比较密集,它的实际宽度和模型中四边形的宽度不相等,而是成一定的比例关系,所以不能直接用纹理图像的宽度代替模型中四边形的宽度。设模型 2 左平面的宽度为  $w$ ,那么拼接后 3 维模型的宽度  $\hat{w}$  为

$$\hat{w} = w \times \frac{w_3}{w_2} \quad (3)$$

右平面与地面的纹理拼接和模型计算方法相同。对于曲面模型,模型的拼接稍微复杂一些,不过只要在拼接纹理的时候,保留接缝的位置,其计算过程和平面模型则大同小异。

模型拼接完后,将拼接后的纹理映射到新模型,就可以获得一个类似于管道的 3 维虚拟场景(如图 7 所示),确定虚拟相机的参数后,就可以在这个 3 维场景中漫游了。虚拟相机包括以下参数:相机的 3 维坐

标、视线方向和焦距。当这些参数都确定后,则由式(1)、(2)计算就可以得到新图像。为了避免空洞的产生,在实验过程中,还采用了逆向映射的方法,即对像平面中的任意一点  $(x,y)$ ,都可以用式(1)和式(2)来反向寻找虚拟场景中的对应场景坐标  $(X,Y,Z)$ 。

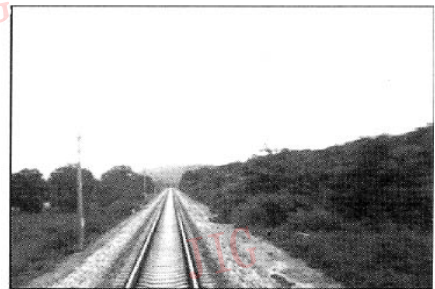
## 6 实验结果

图 6 说明了整个场景的重建过程。图 6(a)和图 6(b)是用来重建场景的两幅原图像,从图像的近景可以看出,图 6(a)和图 6(b)是相机做前向运动拍摄的,图 6(c)、图 6(d)分别是图 6(a)、图 6(b)的右平面纹理,从两幅图像可以看出,纹理图像右端的差别,同时它们也是部分重合的,这就使得两幅纹理能够拼接在一起。图 6(e)是纹理拼接的结果。由于拼接后的纹理充分利用了前后两幅图像清晰的近景信息,因而使拼接纹理比原纹理更加清晰。

图 8 展示了在重建的 3 维场景中漫游获得的图片。本文试验中,拼接了图 6(a)和图 6(b)的左、右、上和下各个平面的纹理,并计算了拼接后的新模型。图 8(a)是图 6(a)和图 6(b)的中间视点图像,图 8(b)是在重建场景中,向前右方漫游获得的图片。图 9 给出了曲面情况下的原图像和场景重建后,向前漫游获得的图片。另外,在重建场景中漫游时,可以一直看到清晰的图片,几乎感觉不到模型的跳变。



(a) 中间视点图像



(b) 向右漫游

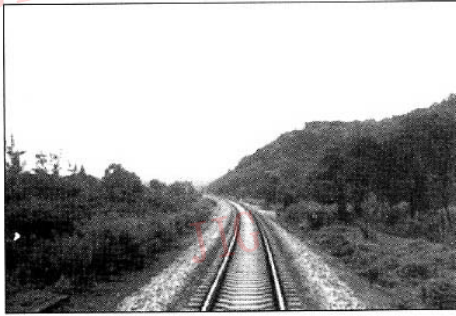
图 8 在重建场景中的漫游

Fig. 8 Touring into the constructed virtual environment

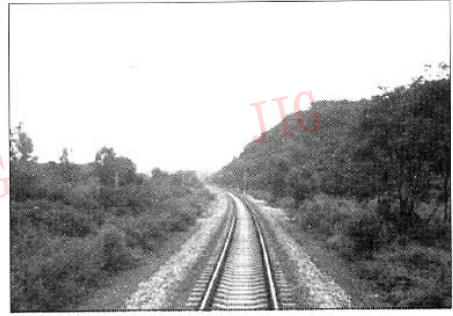
## 7 结 论

本文在 TIP 技术的基础上,提出了一种未知运

动路径情况下的基于图像序列的场景重建和漫游算法。由于受图像序列视角范围的限制,不可能从图片中获得场景的 3 维信息,为此本文根据透视投影关系和运动路径,为场景加入了一个近似的 3 维信



(a) 原图像



(b) 向前漫游

图 9 曲面场景重建和漫游

Fig.9 Constructing curve model and tour into it

息,并将前后图像的模型根据它们的位置关系串连起来,就实现了在图像序列中的漫游。本文还提出了一种曲面建模法,并用其解决了路径为曲线情况下,传统 TIP 技术无法建模的问题。但是在纹理拼接的过程中,由于分辨率的变化,致使拼接纹理的过渡还不太自然。再者,场景的建模还需要人工交互,尚不能全自动完成。这些都有待于进一步的研究。

#### 参考文献 (References)

- 1 Eric W, Grimson L. Why stereo vision is not always about 3D reconstruction [R]. Technical Report 1435, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, Boston, MA, USA, July 1993.
- 2 CHEN Shen-chang Eric, Williams L. View interpolation for image synthesis [A]. In: SIGGRAPH '93 [C], Anaheim, California, USA, 1993: 279 ~ 288.
- 3 Seitz S M, Dyer C R. View morphing [A]. In: SIGGRAPH '96 [C], New Orleans, Louisiana, USA, 1996: 21 ~ 30.
- 4 McMillan L, Bishop G. Plenoptic modeling: An image-based rendering system [A]. In: SIGGRAPH '95 [C], Los Angeles, California, USA, 1995: 39 ~ 46.
- 5 CHEN Shen-chang Eric, Williams L. QuickTime VR—An image-based approach to virtual environment navigation [A]. In: SIGGRAPH '95 [C], Los Angeles, California, USA, 1995: 29 ~ 38.
- 6 Horry Y, Anjyo K, Arai K. Tour into the picture: using a spidery mesh interface to make animation from a single image [A]. In: SIGGRAPH '97 [C], Los Angeles, California, USA, 1997: 225 ~ 232.
- 7 Szeliski R. Video mosaics for virtual environment [J]. IEEE Computer Graphics and Applications, 1996, 16(2): 22 ~ 30.
- 8 Li Nan, HUANG Zhi-yong. Tour into the picture revisited [A]. In: International Conference in Central Europe on Computer Graphics (WSCG) [C], Plzen, Czech Republic, 2001: 41 ~ 48.
- 9 Yoon Sanghyuk, CHEN Hai-jung, Hsu Tom, et al. Web-based virtual tour using the tour into the picture technique [A]. In: International Conference on Distributed Multimedia Systems [C], Miami, Florida, USA, 2003: 81 ~ 88.
- 10 ZHANG Xiao-hua, Hiroki Takahashi, Masayuki Nakajima. Construction of virtual environment from video data with forward motion [A]. In: Proceedings of Advanced Multimedia Content Processing (AMCP) [C], Osaka, Japan, 1998: 301 ~ 312.
- 11 KANG Hyung-woo, Pyo S H, Anjyo K, et al. Tour into the picture using a vanishing line [J]. Computer Graphics Forum, 2001, 20(3): 132 ~ 141.
- 12 CHU Siu-hang. Animating chinese landscape and panoramas [D]. Master's Thesis, Computer Science Department, Hong Kong University of Science and Technology, Hong Kong, China, 2001.
- 13 Hearn Donald, Pauline Baker M. Computer vision [M]. Translated by CAI Shi-jie et al. Beijing: Publishing House of Electronics Industry, 2000. [Hearn Donald, Pauline Baker M 著. 蔡士杰等译. 计算机图形学 [M], 北京: 电子工业出版社, 2000.]