

# 由抖动图象序列建立自然环境三维真实感全景模型

朱志刚 徐光祐 林学闯

(清华大学计算机系信息教研组, 北京 100084)

**摘要** 由真实环境中的现场图象进行三维环境建模是目前国际上研究的热点问题。本文依据合理的运动模型, 提出和实现了由包含抖动的摄像机运动下的图象序列建立 3D 环境全景模型的两步法。首先通过运动滤波和运动分解获得运动稳定的图象序列, 然后采用无特征提取的时空纹理方向精确估计、深度边界确定和遮挡恢复算法, 建立全局自然景物的真实感三维环境模型。提出了 2 种三维全景图象的表示方法, 即非阵列方式深度分层区域表示和阵列方式的深度分层布景表示, 可用于机器人全局定位的自然路标提取和真实环境虚拟再现的图象合成。该研究推广和结合了外极面图象的方法和全景图象的方法, 放宽了对运动的要求, 从而可使该方法适用于室外颠簸的道路环境。和现有运动分层方法相比, 避免了该类方法迭代过程中的局部最小化问题, 并具有计算和存储效率高, 适应性强, 算法鲁棒性好的优点。

**关键词** 基于图象的 VR, 图象稳定, 外极面图象, 全景图象, 分层表示

## 1 引言

在现有的虚拟现实(VR)系统中, 物体及环境的三维模型主要是通过图形学和 CAD 的方法建立。这种方法可以获得较为精细而准确的三维模型, 但对于自然景物, 由于三维模型极其复杂, 并且需有其各自对应的纹理, 因此数据量和工作量巨大。因此从 90 年代开始, 人们开始将注意力转移到如何自动获取环境或物体三维模型的方法, 即基于现场图象的 VR 建模技术<sup>[1,2]</sup>。该技术尤其适用于那些难以用 CAD 方法建立真实感模型的自然环境, 以及需要真实重现环境原有风貌的场合。环境建模除可用于虚拟现实(VR)外, 还可用于视觉导航<sup>[3]</sup>、图象压缩<sup>[4]</sup>、视频检索和浏览等<sup>[5]</sup>。

运动序列理解是建立环境三维模型的主要方法之一。长时 2 帧图象虽可计算出较好的深度, 但求对应困难, 因为深度不同会带来明显的遮挡和非公共区域。采用短时 2 帧图象易求出图象运动场, 但由于运动视差幅度太小, 深度计算误差大。光流的方法即属于短时 2 帧的方法, 由于时空域上均在很小的区

域内操作, 因而图象运动的计算不精确。提高计算精度和可靠性一般有 2 种方法, 即时域约束的方法和空域约束方法。

增加时域范围(序列图象的帧数), 可提高光流计算的精度, 但若运动复杂, 轨迹线的跟踪变得非常困难。外极面图象的方法<sup>[6]</sup>假设运动恒定, 因此可将运动分析转化为二维外极面图象上轨迹线的提取, 但由于对运动的要求过于苛刻, 限制了其实际应用。全景图象的方法<sup>[3,7]</sup>拓宽了图象在水平方向的视野, 适合于表示大型的室外景物, 但在获取三维信息方面, 却将两空域图象的匹配问题转化为两帧时空全景图象的匹配问题。Factorization 的方法<sup>[8,9]</sup>利用多帧图象上对应特征点的冗余信息, 通过奇异值分解的方法计算运动和结构, 但需要假设正交投影成象模型, 并需找出多幅图象上对应的特征点。为了达到快速 3D 计算, 有的作者通过精确控制摄像机运动以避免运动估计和对应问题。Dilmia 等<sup>[10]</sup>通过控制平移运动的精度, 从时空梯度信息中获得不同质量的深度信息, 但要求达到 0.1mm 的摄像机运动精度; Murry<sup>[11]</sup>则通过精确旋转摄像机, 采用时空图象分析技术获得深度, 但只能用简明的方法获得一条

\* 国防预先研究课题(No. 16. 10. 1)和 863 课题(No. 863-306-2D-10-22)资助项目

收稿日期: 1997-11-14; 收到修改稿日期: 1998-01-09

中心扫描线上的 3D 信息。

增加空域范围会包含多个运动/深度区域,因此需解决多运动分割这一难题,目前常用的方法是直接运动估计方法<sup>[4,5,12]</sup>。直接方法又可分为主运动分析方法和多运动同时估计方法。在一般的运动情况下,完全 3D 运动和深度计算有困难且为病态问题,目前比较成功的方法将多图象运动近似为平面仿射变换<sup>[4]</sup>,即 2 幅图象之间几个深度层的运动符合平面仿射变换,这样实际上对运动和景物的近似,虽可用于图象压缩,但难以生成任意新视角的图象。Sawhney 等<sup>[5]</sup>提出带射影深度的运动估计模型,将各点深度离射影平面的距离约束加到深度分层数和射影变换后 2 幅图象残差的最小描述长度(MDL)约束中。虽然该模型更具一般性,且无需进行特征提取,但其缺点是:计算量大;深度变化最小的假设和分层区域最少的假设不一定合理;由于采用最小化方法,可能会导致局部最小而达不到正确的结果。

在任意(未知)运动下自动恢复 3D 信息尚无成熟的方法。为达到有效 3D 建模,本文提出一个可用于实际环境建模的非精确运动的模型,即已知的平面运动加上未知的任意抖动,而对景物的 3D 结构无任何限制。这种模型适合于由移动车体上拍摄的运动序列建立自然环境模型的场合。在此基础上提出图象稳定加全景外极面图象分析的两步法,有效地将复杂运动和复杂 3D 景物参数分开,从而使全景图象方法和外极面图象方法可应用于实用环境中。深度估计算法为不进行特征提取和对应的直接方法。最后,提出了系统解决遮挡问题和均匀区域深度估计问题的一种方法,并设计了分层区域表示和分层布景表示,在较低的计算代价下构造出景物的 3D 全景表示,表示紧凑、时空效率高。

## 2 图象稳定和全景外极面图象生成

在时间段  $t \in [0, T]$  内,车体运动由 2 部分组成(图 1):(1) 已知轨迹的平面运动。本文假定车体沿道路做速度为  $V$  的匀速直线运动,摄像机光轴垂直于运动方向,其它情况下可进行图象采样和校正<sup>[13]</sup>;(2) 在相邻 2 个时刻( $t, t+1$ )之间,车体有微小的随机抖动,用 3 个自由度的旋转角  $\Omega_x, \Omega_y, \Omega_z$  和 3 个自由度的平移量  $T_x, T_y, T_z$  表示。图象稳定的作用是消除或减弱车体的随机抖动,获得平稳运动下的图象序列。图象稳定算法分为 3 个步骤:图象运动矢量估计,时域运动滤波和图象几何校正,具体算法

已在文献<sup>[13,14]</sup>给出。

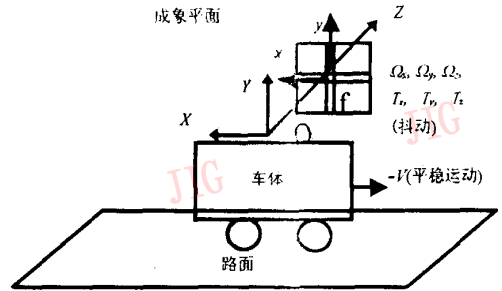
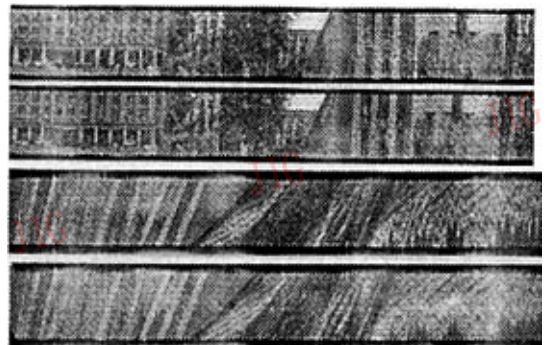


图 1 车体运动模型

在上述运动假设下,图象稳定后的效果相当于车体只作匀速直线运动。在这种设置下,空间点  $P$  的轨迹  $(x, y, t)$  就约束到平行于  $x$  轴的扫描线上,透视成像可表示为:

$$x(t) = f \frac{X + Vt}{Z}, y(t) = f \frac{Y}{Z} \quad (1)$$

其中  $(X, Y, Z)$  表示  $P$  点在时刻  $t = 0$  时的三维坐标。通过图象稳定,2 种有意义的二维时空截面图象——全景图象(PVI)和外极面图象(EPI),就可在颠簸的路面环境下的图象序列中获得。其中,前者为三维时空图象中平行于  $y-t$  的截面(如图 2(b)),后者为平行于  $x-t$  的截面(如图 2(d))。PVI 相当于在  $y$  方向上透视投影而在  $x$  方向平行投影的结果,它基本上映了侧面景物的表面特性;而在 EPI 中,由于运动及景物空域特征所形成的纹理方向则反映了景物中物体的深度。这种方向性纹理称为时空纹理或运动纹理。原图象的一个特征点  $p(x, y)$  在 EPI 上对应一条直线轨迹线,它的斜率反比于景物点  $P(X, Y, Z)$  的深度  $D$ , 即  $D = Z = f \frac{V}{v}$ , 其中  $v = \frac{dx}{dt}$  是直线

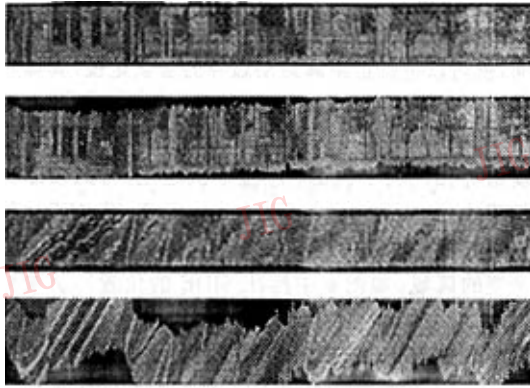


(a) 稳定前的 PVI (b) 运动滤波后的 PVI  
(c) 稳定前的 EPI (d) 运动滤波后的 EPI

图 2 MB 序列图

轨迹线的斜率。图象稳定中的运动滤波(运动分解)正是利用了这一性质。

TREE 序列是将固定在三角架上的摄像机载于手推的小车上,在预制板块路面上基本保持匀速移动拍摄的,图象具有较大幅度的平移和旋转性抖动,在预制板块的交接处尤为剧烈。数字化后的图象序列的数据量(字节数)为 128(行)×128(列)×1 024(帧),图象获取频率为每秒 19 帧。图 3(a)和(c)为稳定前的 PVI 和 EPI(均为 128×1 024),显见图象的抖动情况。在经过稳定的 EPI 上难以得到可靠的深度信息。考虑到该景物中的物体在每幅图象中的深度布局非常相似,故可采用直线拟合的运动滤波方法,图 3(b)中 PVI 的水平线,图 3(d)中 EPI 中的轨迹线尤其是背景物体的轨迹线,均得到明显的改善。图 3(d)的 EPI 上也显示出摄像机摆动的幅度几乎为半幅图象,说明图象稳定方法可以处理较大幅度的图象抖动。



(a) 稳定前的 PVI (b) 线性运动滤波后的 PVI  
(c) 稳定前的 EPI (d) 线性运动滤波后的 EPI。  
PVI 为每幅图象中央一列拼接而成( $x=0$ ),  
EPI 对应于图象正中央( $y=0$ )的水平扫描线。

图 3 TREE 序列

将摄像机用三脚架固定在轿车上,在较平整的广场上运动获得 MB(Main Building)序列,数字化后图象的数据量为 128(行)×128(列)×1 024(帧),图象获取频率也为每秒 19 帧。MB 序列大约有 1 个象素左右的抖动(见图 2)。采用平滑运动轨迹的运动滤波方法,可以获得较好的全景图象和外极面图象。在 PVI 上,从中央主楼①和过街天桥②的水平边可以看出垂直方向抖动的校正情况;从东主楼门口的垂直窗户棂③可见对水平方向摆动的校正情况。在 EPI 上,注意图中②、③处树木和楼门的轨

迹线的校正情况。在图象右侧(处出现的大片流水状纹理,是原图象中水平边的孔径效应以及图象的轻微抖动造成的非对应点在 EPI 中的表现,在后面的深度估计中可以看到此处得不到正确的深度。实验说明图象稳定虽无法完全消除抖动,但在很大程度上减轻之。

### 3 全景外极面图象分析方法

在传统的外极面图象(EPI)处理中,需要对 EPI(即  $xt$  图象)上每一点进行处理,计算量很大。在本文中,只以所关心的二维 PVI 上各点为中心,有选择地分析 EPI 区域即可得出较完整的深度信息,我们称之为全景外极面图象分析方法。基于时空纹理方向检测的能量模型,本文采用了一个大窗口 Gauss-Fouier 方向检测算子 GFOD 来检测时空纹理方向。大窗口可以满足方向检测的精度,高斯窗口可以照顾到边界的精确定位<sup>[13]</sup>。

在实际应用中,我们提出扩展的全景图象 XPI(Extended Panoramic Image)的表示,以更多利用原序列的信息。设图象序列数据量为  $W$ (列)× $H$ (行)× $F$ (帧),方向检测算子 GFOD 的尺寸为  $m$ × $m$ ,则有效的 PVI 大小为  $W$ ×( $F-m$ )。从长运动序列( $F \gg m$ ,如  $F=1 024, m=64$ )生成的 PVI( $y, t$ )包含了所获取的大部分视野的信息,只在对应于 PVI 的  $x$  坐标(如  $x_0=0$ )处沿  $t$  轴的各点进行 EPI( $y, t$ )的分析即可。但当运动序列较短且单幅图象分辨率较高时,有效的 PVI 只有很小一段,没有体现出“全景”的优点。如对于常用于运动分析的花园(FG, Flower Garden)序列<sup>[4,5]</sup>,图象数据量为  $W$ × $H$ × $F=352$ × $240$ × $115$ ,PVI 的大小仅为  $H$ × $F=240$ × $115$ ,若  $m=64$ ,则有效的 PVI 大小为  $H$ ×( $F-m$ )= $240$ × $51$ 。FG 序列为稳定的平移序列,图 4 显示了的第 32 帧和最后 83 帧图象,图 5(a)、(b)为生成的全景图 PVI 和草地处的一幅外极面图 EPI。为此,本文生成了图 5(c)所示扩展的全景图象 XPI,其构成为:第  $m/2$  帧的前半帧,第  $m/2 \sim F-m/2$  帧每帧的中间列( $x=0$ )和第  $F-m/2$  帧的后半帧。对各 EPI 的分析基本只需考虑对应于 XPI 上各点(图 5(b)中的折线上)。

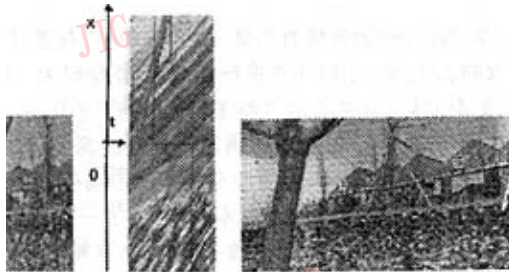
全景外极面图象分析算法分 7 个步骤:

(1) 方向可度量性计算 从全景图象  $I_{pvi}(y, t)$  上,可计算出点( $y, t$ )的运动纹理方向估计的可度量性



(1) 第 32 幅图象 (2) 第 83 幅图象

图 4 FG 序列



(a) PVI (b) EPI (c) XPI

图 5 FG 序列全景图,外极面图和扩展全景图

$$B(y,t) = G_x(y,t) - G_y(y,t) \quad (2)$$

其中  $G_x(y,t), G_y(y,t)$  分别为 PVI 上点  $(y,t)$  处水平方向和垂直方向的梯度。水平梯度超过垂直梯度的值越大,可度量性越好。

(2) **多方向检测** 在对应于 PVI 每一垂直坐标  $y$  的外极面图象  $I_{EPI}(x,t)$  上,只在 PVI 对应的  $x$  位置(一般  $x=0$ )上,顺序找出位置  $t_i$ ,使  $B(y,t_i) > T_b$ (一般取  $T_b \geq 2$ )。以  $(x,t_i)$  为中心,采用  $m \times m$  的 GFOD 算子,检测出表示不同深度的时空纹理方向角  $\theta_k(k=1, \dots, K)$ (一般情况下  $K=1 \sim 2$ )。从而可以得到  $K$  个可能图象速度  $v_k = \tan \theta_k$ 。

(3) **运动边界定位** 在产生多方向的位置需通过进一步的分析,定位运动边界。由于摄像机运动抖动和自然物体(如树叶、草地)的非刚体“抖动”难以完全消除,故不能严格保证灰度的时域同一性(如图 5(b)中草地的黑白交替运动),采用轨迹线跟踪的方法易失败。因此我们在用 GFOD 算子检测出多个方向后沿各方向采用了多尺度窗口的灰度/梯度

一致性的比较判断,选择一致性最好的为此处的方向,从而可准确定位运动边界(图 6)。

(4) **方向值内插** 为了得到致密的运动纹理方向(深度),对无纹理的区域采用合理的内插方法。内插方法基于这样一个基本事实:在深度变化处几乎一定有明显的遮挡边界或影调边界,因此该处几乎总能检测出深度。对方向角设定一个阈值  $T_{dis}$ ,在 2 个可度量出方向角  $\theta_1, \theta_2$  的时刻  $t_1, t_2$  之间的均匀区域,内插值定义如下:  $|\theta_1 - \theta_2| < T_{dis}$  时(即深度“连续”),将中间线性内插;  $|\theta_1 - \theta_2| \geq T_{dis}$ (深度不连续),将中间内插为远处的深度(因为近处物体的遮挡边界发生在远处物体的无纹理区域)。

(5) **运动边界二次定位和纹理方向精化** 近距离的物体(前景)由于图象速度矢量幅度大,在时域方向上宽度较窄(图 5(b))。用大窗口 GFOD 算子检测方向,当前景的纹理较弱而背景物体纹理较强时,前景的纹理会淹没在强纹理背景中,在运动边界处可能没有检测出深度,故要特殊处理。一般情况下,只要能检测到近距物体一个以上点的运动纹理方向,就可以进行近距离运动边界的重新定位,具体方法类似于运动边界的定位的灰度一致性判断<sup>[13]</sup>。对于 FG 序列图象,由于树干的纹理很弱,图 7(a)中深度边界偏向树干内部。经过 2 次定位,得到较好的深度边界,见图 7(b)。在单方向区域,用 GFOD 算子估计的方向受固定窗口大小  $m$  限制,而且在方向渐变的区域(如图 6 中标注 SIDE 的位置),会受到周围纹理的影响,因此检测出的方向可能会与正确的方向有偏差。以 GFOD 算子估计的初始方向角为基础,可进一步用基于最佳灰度一致性的方向精化方法找出更准确的方向,而且对远距离的点,可利用更大的时域范围(更多的帧)。

(6) **被遮挡区域的恢复和近距离分辨率的恢复** 由于 PVI 上只保留了一个视角的信息,因此在有些视点的原始图象上能看到的景物在 PVI 上丢失了。我们通过遮挡关系的分析将其从其他时空位置的图象信息中补出(图 6)。在遮挡边界点  $p_0$  处,确定遮挡区域  $p_0, p_c, p_e$  的类型(OCCLUDED 或 SIDE),并将恢

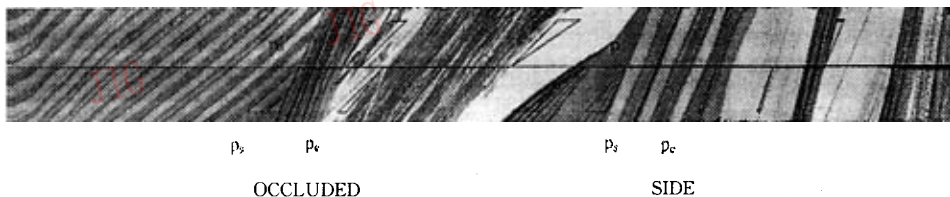
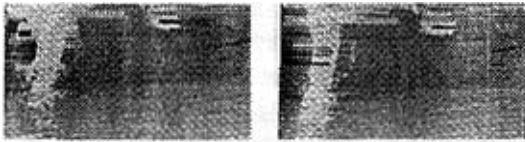


图 6 被遮挡区域的恢复和近距离分辨率的恢复

复出的时空段  $p_s, p_e$  补在  $p_0$  处。对于近距离的物体 ( $v > 1$  时), 在 PVI 上损失了原图象的分辨率, 因此可根据图象速度的大小, 在  $x$  方向上取相应长度的段, 恢复原图象的分辨率。图 6 中 EPI 上不同粗细的黑线表示了每帧图象所取信息的长度。在基于 XPI 的遮挡区域恢复时, 应在相应的  $x$  方向和  $t$  方向上处理(见图 5(b))。



(1) 原始深度图 (2) 深度边界的二次定位

图 7 FG 序列:深度图

(7) 深度分层段表示 在增加遮挡区/侧面区并恢复图象原分辨率后, 就无法用数组图象表示 PVI, 因此构造了以不同深度段为基础的变分辨率非数组 PVI 表示, 深度分层段表示 (DLL, Depth Layer Line), 每段的结构为

```
struct DepthLayerLine {
    unsigned char type; /* FRONT, SIDE, OCCLUDED
    */
    int x, y; /* DLL 的空域图象坐标(x,y) */
    float ts, te; /* DLL 的起止帧 */
    int length; /* DLL 的长度 */
    unsigned char *I; /* DLL 内各点的灰度 */
    float *v; /* DLL 内各点的图象速度 */
}
```

在每一段中各点的图象速度值  $v$  没有突变; 由各点的速度递推可计算出各点所处的帧  $t$ , 进而计算出在  $t=0$  帧坐标系中的 3D 坐标:

$$Z = F \frac{V}{v}, Y = y \frac{V}{v}, X = x \frac{V}{v} - Vt \quad (3)$$

在 PVI 上计算出方向可度量性函数  $B(y, t)$  后, 对每一  $y$  坐标对应的外极面图象的基本处理是独立的, 因此可以用大粒度的并行结构实现, 如果有  $H$  个并行处理器, 则可同时完成  $H$  幅  $W \times F$  外极面图象的处理。

## 4 深度图和灰度图的融合

单纯长时序列运动分析(如 EPI 分析), 可以对明显的特征点和较强或高频的纹理区域的运动及运动边界得到较好的结果, 失效或发生错误的情况包

括:

(1) 均匀区域。在 EPI 处理时, 将对应于 PVI 上某一  $y$  处的均匀区域根据左、右明显轨迹线进行深度插值, 这可能发生错误, 如在图 7 的树枝之间无纹理的天空错误的深度赋值。

(2) 水平边。在 EPI 分析中, 采用的方向可度量性准则将其定义为不可度量, 因此其结果类似于均匀区域, 如图 8(b) 图象右中明显的水平亮线, 它对应于图 2(d) 中外极面图象的处理结果。因此在全景外极面图象运动分析第(1)~(5)步生成的深度图象的基础上, 我们将 PVI(或 XPI) 灰度信息和深度信息进行融合。融合方法使用了以下基本的事实: 深度边界(运动边界)几乎一定发生在灰度边界(纹理边界)处, 在灰度均匀的区域(或纹理几乎一致的区域)一般不可能有深度突变。

目前实现了一个依据上述原理的简单算法, 其步骤为:

(1) 对深度中值滤波。虽未对景物深度做任何限制, 但实际景物中的深度符合两种类型: 局部深度连续(或不变), 局部深度突变。为保持深度边界的位置, 并去除在全景外极面分析中产生的噪声, 在水平和垂直方向上分别进行中值滤波, 产生较少噪声的深度图。

(2) 垂直方向上边界修正: 在 XPI 的垂直方向上, 检测出灰度突变的灰度边界和深度突变的深度边界。如果在深度边界处无灰度边界, 则将深度边界移到合理的灰度边界处, 并将改变边界的区域重新赋值。

经过上述 2 步并未解决所有的问题, 但有很大的改善。图 8 显示了 MB 序列的全景图、原始深度图和空域滤波后的深度图。图中的深度图是直接将方向角的值用亮度形式表示, 距离越近亮度越大。在图 8(a) 中, 叠加在 PVI 上的黑线表示了深度边界, 可以发现对树和楼房之间, 楼房不同深度之间有较精确的深度边界。

在扩展全景图象 XPI 上处理了 FG 序列, 图 9(a) 是对图 6(b) 深度图进行空域滤波的结果, 图 9(b) 中将方向角之差为  $1^\circ$  的深度变化线覆盖在 XPI 上。可看到草地由近到远渐变的深度, 树干和背景深度的明显界限, 及房前明显可见的支架。图 9(a) 中黑色区域对应无纹理的天空, 因其处于图象边框无法进行内插, 故赋值为最小角度  $0^\circ$  (对应无穷远)。

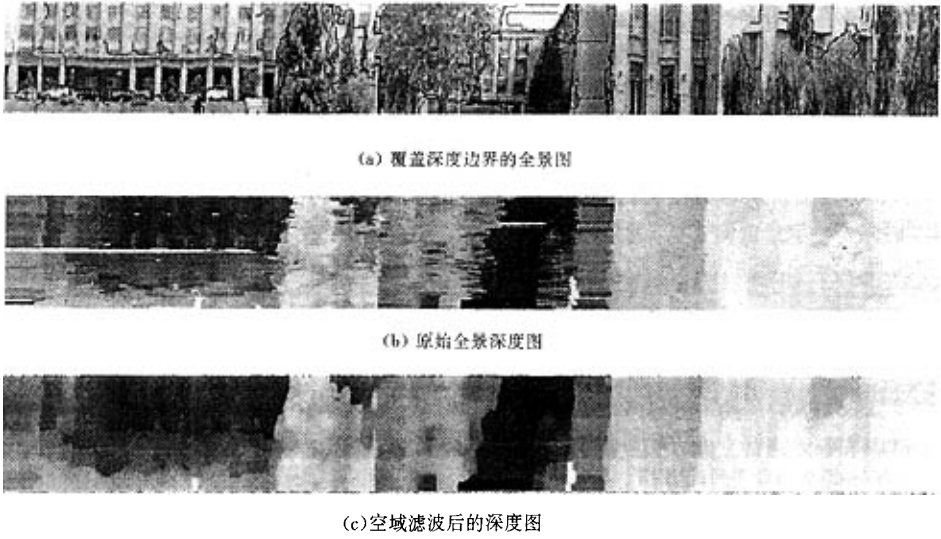


图 8 MB 序列: 全景图、原始深度图、空域滤波后的深度图



图 9 FG 序列: 空域滤波

### 5 深度分层表示

在形成对应于 PVI 或 XPI 的基本深度图后,为得到更完整的景物表示,首先进行了深度区域的划分,分辨率的增强和遮挡/侧面区域的检测和恢复,得到深度分层段 DLL(第 3 节第 6.7 步),在此基础上研究了 2 种表示方法:深度分层区域表示(DLR, Depth Layered Region)和深度分层布景表示(DLS, Depth Layered Setting)。

#### 5.1 深度分层区域表示

对于 PVI 及其对应的深度图,我们构造出深度的分层区域表示 DLR,每个深度分层区域表示具有连续深度的区域,它是将具有连续深度的分层段聚类的结果。针对识别和显示的需要,DLR 具有以下几个据域:区域类型、几何特征、纹理特征、区域原始数据 DLL 和区域的相互关系:

```
struct DepthLayerRegion {
```

```
    unsigned char type; /* 区域类型(包括 FRONT,
    OCCLUDED, SIDE) */
    GEOMETRY geometry; /* 几何属性(包括中心(X0,
    Y0, Z0),体积范围(ΔX, ΔY, ΔZ)表面方向等 */
    TEXTURE texture; /* 包括纹理类型,灰度平均值  $\bar{I}$ ,
    等 */
    DepthLayerLine * dll; /* 由分层段组成,每个分层段
    包括各点的深度 I 和图象速度 v */
    RELATION relation; /* 其它区域的关系(如同一深
    度、遮挡关系、侧面) */
}
```

图 10 比较显示了遮挡恢复前后的合成图象,在图 10(b)中,被树冠遮挡的在阴影中的部分楼房显露出来了。



图 10 MB 序列: 遮挡恢复

#### 5.2 深度分层布景表示

本文对不同深度层次和具有遮挡关系的景物提出和实现了一种深度分层布景表示 DLS,这种表示类似于舞台上的布景,但不同的是每层除了包含表示布景层纹理表面的灰度图(Intensity Map)外,还

有一幅用运动纹理的方向角表示的深度图 (Depth Map)。DLS 的数据用简单的阵列图象表示,并有以下特点:

(1) 分层布景表示以扩展全景图象 (XPI) 为基础,是一时空组合图象 (在长序列 PVI 中为 yt 图象,在水平方向上为平行投影,在垂直方向上为透视投影);而已有的分层表示采用某一视图为参考帧的透视图,当序列较多时难以表示。

(2) 分层布景表示只按遮挡关系分层,之所以分层是为了表示被遮挡的区域和表示不同的分辨率,而未将图象简单划分为近似的仿射平面区域。

(3) 分层布景表示了每个层次上点的灰度和深度,因此可以由此生成任意视角的图象,由于分层,也可选择不同的层次进行图象合成。

图 11 显示花园图象序列 FG 的 2 个布景层,即背景 (草地,房子和天空),及前景的物体 (树干、树枝),注意到在背景上填补了在原 XPI 上丢失的纹理和深度信息,虽然在此例子中花草地由近到远,基本为一水平面,楼房基本为一垂直面,但由于没有明显的遮挡关系和深度突变,仍划归为同一背景层。

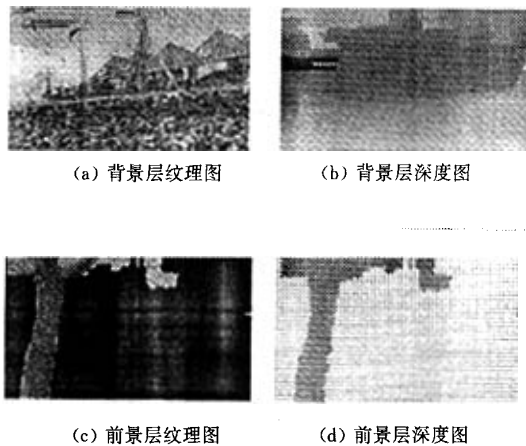


图 11 FG 序列:深度分层布景表示

观察的干扰。有以下 3 种显示方式:

(1) 全景图象显示:可以显示每层或多层的全景表示,如直接显示原始 PVI (图 2(b)) 和不同深度层次的 PVI (图 11) 等,由于具有三维数据,也可显示不同视角的全景图,如顶视图。

(2) 任意视点图象显示。可方便实现虚拟漫游,分层布景表示简化了透视投影时的遮挡消隐问题,对于某一视角,可根据各层在视点坐标系的位置,由远及近顺序渲染生成 (rendering),从而得到正确的遮挡关系,图 12(a) 显示了在远离远拍摄路径视点上的合成 FG 图象,与图 4(a) 比较可发现树干和房前小树的图象位置关系有变化。

(3) 选择不同的层显示,可以显示不同的深度层次,图 12(b) 显示了没有前景物体 (树干) 的 FG 序列的合成图象。

### 6.2 选择机器人定位的显著路标

3D 全景建模方法可应用于机器人全局定位的路标提取。显著的路标可以选择不同深度层次的物体,如 MB 序列中的树丛和主楼的深度突变处,FG 序列中的树干。在深度分层区域上表示了区域属性,可以采用区域间的相互关系定位机器人,如深度变化的关系或表面颜色、纹理的关系等。

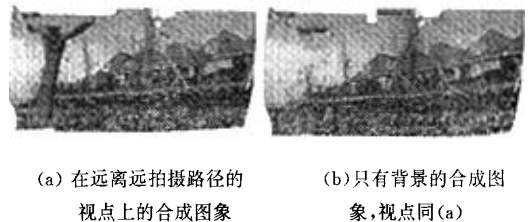


图 12 FG 序列:合成图象

### 参考文献

- 1 Kanade T, Rander P, Narayanan P J. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE Multimedia*, 1997; 34~47.
- 2 Szeliski R. Image mosaicing for tele-reality application. *Proc. 2nd IEEE Workshop on Application of Computer Vision*, 1994, 44~53.
- 3 Zheng J Y, Barth M, Tsuji S. Panoramic representation of scenes for route understanding. In: *Proc. IEEE 10th ICPR*, 1990; 161~167.
- 4 Wang J, Adelson E H. Representation moving images with layers. *IEEE Trans. on Image Processing*, 1994, 3(5): 625~638.
- 5 Sawhney H S, Ayer S. Compact representation of videos through dominant and multiple motion estimation. *IEEE Trans. PAMI*,

## 6 3D 全景模型的应用

### 6.1 任意视角图象的合成

在建立了 3D 全景模型后,可以合成任意视角的图象,用于自然场景真实环境再现。由于采用虚拟摄像机,所以可以不受视点、视角、视野的限制,生成适合于人观察和理解的场景图象;也对场景中不同深度层上的物体选择显示,从而避免不重要物体对

- 1996, 18(8): 814~830.
- 6 Baker H H, Bolles R C. Generalizing epipolar-plane image analysis on the spatio-temporal surface. In: Proc. CVPR, 1988: 2~9.
  - 7 Ishiguro H, Yamamoto M, Tsuji S. Omni-directional stereo for making global map. In: Proc. ICCV-90.
  - 8 Costeira J, Kanade T. A multi-body factorization method for motion analysis. In: Proc. ICCV'95, : 1071~1076.
  - 9 Tomasi C, Kanade T. Shape and motion from image streams under orthography: A factorization method. Int CV, 9, 1992, 137~154.
  - 10 Dalmia A K, Trivedi M. High speed extraction of 3D structure of selectable quality using a translating camera. CVIU, 1996, 64(1): 97~110.
  - 11 Murray D W. Recovering range using virtual multicamera stereo. CVIU, 1995, 61(2): 285~291.
  - 12 Irani M, Ananda P, Hsu S. Mosaic based representation of video sequence and their applications. In: Proc ICCV'95, 605~611.
  - 13 朱志刚. 视觉导航中环境建模的研究[博士论文]. 北京: 清华大学计算机系, 1997. 6.
  - 14 杨雨东, 徐光祐, 朱志刚. 2.5 维帧间运动估计方法. 清华大学学报(自然科学版), 1997, 137(9).



**朱志刚** 副教授, 1997 年获清华大学计算机博士学位, 现为 IEEE 会员, 中国图象图形学学会常务理事。从事专业研究方向为计算机视觉、虚拟现实建模和智能机器人。在国内外刊物、学术会议上发表论文 50 余篇。



**徐光祐** 教授, 博士生导师。1963 年毕业于清华大学自动控制系, 现为 IEEE 会员, 中国图象图形学学会常务理事。研究兴趣为计算机视觉、分布式多媒体和人机交互技术。已发表专(译)著 5 部, 在国内外重要刊物上发表论文近 80 篇。



**林学荫** 教授, 1962 年毕业于清华大学自动控制系。1983~1985 赴美从事计算机视觉研究工作。从事专业研究方向为计算机视觉和智能机器人。曾主持和承担 863、国防预研项目 6 项, 发表论文 60 余篇。

## Building 3D Panoramic Model of Natural Scene from Fluctuated Image Sequences

Zhu Zhigang, Xu Guangyou, Lin Xueyin

(Department of Computer Science, Tsinghua university, Beijing 100084)

**Abstract** The problem of virtualized reality, or view synthesis of real scenes, has received increasing attention in recent years. This paper presents a systematic approach to automatically construct the 3D natural scene from video sequences captured by a vibrated camera with only approximately known motion. A two stage method is presented for the 3D panoramic scene modeling which consists of (1) image stabilization by motion filtering and (2) depth estimation and depth boundary localization. The two stage method not only combines Zheng and Tsuji's panoramic image method with Baker's epipolar plane image analysis, resulting the so called panoramic epipolar plane image method, but also generalizes them to handle image sequence vibrations due to the uncontrollable fluctuation of the camera. The two-stage method bypasses the correspondence problem and ill-posed problem encountered in the general motion analysis, and avoids the local minimum problem in the spatial-constraint-based iteration method. This paper mainly describes the robust algorithms of spatio temporal orientation detection of spatio temporal textures, boundary localization of kinetic occlusion and depth layering of the panoramic scene. Two kinds of depth layered representations, depth layered region and depth layered setting, have been developed, which found a ground base for landmark selection of global localization and image synthesis of virtualized reality.

**Keywords** Image-based VR, Image stabilization, Epipolar plane image, Panoramic view image, Layered representation