

# 基于图像运动的三维重建及虚拟化应用

李林尧 张兆翔 王蕴红 王 超

(北京航空航天大学计算机学院 北京 100191)

**摘 要** 提出了一种稳定、快速地获取摄像机视频运动图像的三维重建方法,并对该运动图像做适当的虚拟化处理以展示重建效果。采用基于尺度不变特征点匹配的摄像机标定进行三维重建。尺度不变特征对于视频图像中的特征具有优秀敏锐的匹配能力,极大地放宽了摄像机标定对于设备上的限制,拓宽了实时三维重建的适用范围。通过对系统的一系列优化,不但提升了三维重建的精度,减少了错误匹配对摄像机标定的影响,而且进一步提升了处理速度。通过在三维重建的基础之上进行虚拟化处理,展示了本系统的三维重建效果。实验结果表明,该系统适用范围广,处理速度较快,重建精度高,实现了基于视频运动图像的三维重建。

**关键词** 特征匹配,摄像机标定,三维重建,虚拟现实,基于运动恢复结构

## Motion-based 3D Reconstruction and Applications to Virtual Reality

LI Lin-yao ZHANG Zhao-xiang WANG Yun-hong WANG Chao

(School of Computer Science and Engineering, Beihang University, Beijing 100191, China)

**Abstract** This paper presented a method to robustly and rapidly reconstruct general 3D scene from 2D images in camera, with the appropriate virtual application to show the effects of reconstruction. We presented novel camera calibration and scene reconstruction using scale-invariant feature points. Scale invariant feature has impressive matching ability in camera's video, which greatly reduces the restriction of facility in camera calibration, broadens the area of video reconstruction. With some optimization, we enhanced the accuracy of reconstruction and speed. At last, we processed the visualized application, which shows the effects of our system. The result illustrates that our system is capable of producing accurate and fast scene structure from video in a very few seconds.

**Keywords** Feature matching, Camera calibration, 3D reconstruction, Virtual reality, Structure from motion

## 1 引言

三维重建是计算机视觉的最主要的研究方向,在很多重要的应用领域,比如医疗、军事等,一直都是一个充满挑战的难题。在目前的研究体系中,三维重建方式种类繁多,按照基本实现原理,可以将这些方法划分为基于机器学习与图像理解和基于特征提取与摄像机标定两类方法。

Saxena 等人使用 Markov Random Field 和监督学习的方法估算图片中的深度信息<sup>[1]</sup>,它是一种典型的基于图像理解的三维重建系统。这种方法的优势在于无需硬件设备支持,也不需要进行预处理,直接计算就可获得场景三维信息;其缺点在于系统处理效率低(一般每幅图像的处理时间在 2~5 分钟),三维重建效果受场景类型限制且精度较低,另外需要在最初使用前进行数百张图片的机器训练。

三维重建的另外一种主要方式是摄像机标定。摄像机标定是为了获取摄像机的参数(其中内参数包括焦距、光心位置、畸变系数;外参数包括旋转矩阵和平移矩阵)。根据外参

数矩阵就可以计算获得场景三维信息,因此完成摄像机标定等价于三维重建。在近 20 年的发展中,产生了一系列具有代表性的标定算法,例如直接线性变换法 DLT<sup>[2]</sup>、径向一直约束法 RAC<sup>[3]</sup>、主动视觉标定法<sup>[4]</sup>以及张正友的自标定法<sup>[5]</sup>等。在这些算法中,张正友法以其较高的灵活性和稳定性获得了广泛的应用。

特征提取与匹配是摄像机标定过程中不可缺少的重要过程。传统的方法是用灰度角点算子进行匹配,而这种方法依赖于图形样式单一规则的平面模板,限制了场景信息。本文的目标就是通过改进标定算法中的特征匹配方式,消除标定模板对场景的限制,拓展其应用范围。

## 2 尺度不变特征变换技术

特征提取是计算机视觉领域极其重要的研究内容之一。Lowe 提出的 SIFT(Scale-Invariant Feature Transform)特征算子<sup>[6]</sup>是一种基于尺度空间的、对图像缩放、旋转和仿射变换都具备良好不变性的局部特征描述算子,被广泛应用于图像

到稿日期:2010-08-17 返修日期:2010-12-08 本文受国家重点基础研究发展 973 计划(2010CB327902),国家自然科学基金(60873158, 61005016, 61061130560)和中央高校基础科研业务费资助。

李林尧(1988—),主要研究方向为图形图像;张兆翔(1983—),博士,讲师,主要研究方向为模式识别、计算机视觉、数字图像处理、机器学习;王蕴红(1968—),博士,教授,博士生导师,主要研究方向为模式识别、计算机视觉、数字图像处理和信号处理;王超(1983—),博士生,主要研究方向为模式识别、计算机视觉, E-mail: ableblaze@163.com(通信作者)。

中的特征提取和目标识别。本文利用 SIFT 在图像特征提取与匹配上的优秀特性,实现摄像机下的特征匹配。

对于给定的图像  $f(x, y)$ , 高斯尺度空间被定义为  $f(x, y)$  和高斯核函数  $g(x, y; t) = \frac{1}{2\pi t} e^{-(x^2+y^2)/2t}$  进行卷积的结果, 这个过程被称为高斯模糊, 即为

$$L(x, y; t) = (g(\cdot, \cdot; t) * f)(x, y) \quad (1)$$

式中,  $x$  和  $y$  是卷积的变量, 参数  $t$  称为尺度(scale)。图 1 展示了尺度  $t$  逐渐增大的图像效果。

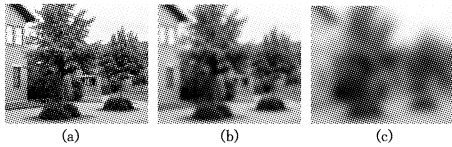


图 1 当尺度(a) $t=0$ , (b) $t=4$ , (c) $t=64$  时, 高斯模糊处理图像的变化情况; 当  $t=0$  时, 图像为原始图像

图 1

由上面这样一组尺度逐渐增加, 经过高斯模糊处理的图像, 可以构成一个图像的高斯金字塔。通过对相邻尺度的图像直接相减获得高斯差分金字塔, 如图 2 所示。

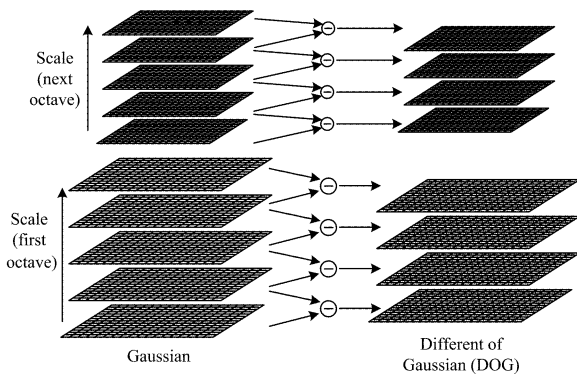


图 2 高斯金字塔与高斯差分金字塔<sup>[6]</sup>

金字塔上的每一个像素点(排除最顶层和最下层的点), 都有 26 个像素点与之相邻, 包括相邻层的 18 个点和所在层的 8 个点。将采样点与其相邻点作比较, 如果其在这个局部区域上明显不同于其他点, 则被定义为极值点。另外, 对极值点进行泰勒展开, 以获得更精确的极值点坐标和尺度, 并且尽可能消除边界响应点。

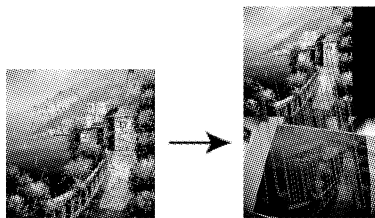


图 3 图中的图像由于灰度信息复杂, 匹配较为稳定, 其作为模板图像被用于接下来的摄像机标定

图 3 SIFT 匹配一例

每个关键特征点都具有 3 个属性: 位置、尺度、方向。对这些点进行方向赋值, 使这些点具备旋转不变性。在计算关键点的主方向时, 首先需要计算梯度直方图。这个直方图由特征点领域内的采样点的梯度方向形成, 分 36 条直方条, 每条覆盖 10 度的方向。每个采样点加入直方图由其梯度幅值和高斯窗加权产生, 即进行高斯平滑处理。得到的方向直方

图的峰值就是局部梯度的主方向。另外, 取与特征点主方向的差值来计算描述字, 从而获得方向不变性。

在 SIFT 特征提取之后, 需要建立一个特征数据库, 这个数据库将用于接下来的匹配。利用每个特征点独特的特征向量进行匹配, 可以获得精确度很高的匹配结果, 如图 3 所示。

### 3 改进的张正友摄像机标定法

摄像机标定的目的是确定摄像机的图像坐标系与物体空间中的三维坐标系之间的对应关系, 即摄像机参数。只有摄像机在正确标定后, 才能根据图像平面中的二维坐标推导出对应物体在三维空间下的实际位置, 或反之。由张正友提出的摄像机标定法<sup>[5]</sup>(简称张正友法)是一种介于传统标定方法和自标定方法之间的摄像机标定方法。这种方法只需同一标定板在不同姿态下的一组图片, 而标定板的运动方式不需知道, 避免了传统标定法设备要求高、操作繁琐等缺点; 又较自标定法精度更高、鲁棒性更好, 因此在应用上非常灵活。

张正友标定法的摄像机模型如图 4 所示。  $M(X, Y, 0)$  为标定模板上的一点, 假设  $Z=0$ 。  $m(u, v)$  为图像上与  $M$  点对应的像素点坐标。  $M$  和  $m$  的关系如下所示

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} [R \quad t] \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} \quad (2)$$

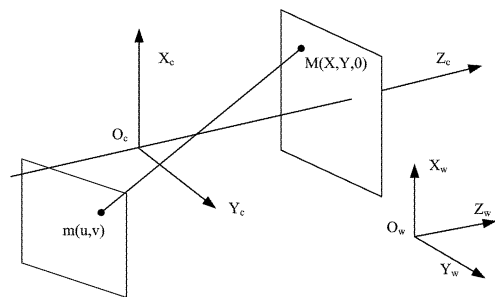


图 4 张正友算法的摄像机模型<sup>[7]</sup>

定义等号右边第一个矩阵为内参数矩阵 (Intrinsic Parameter Matrix), 第二个矩阵为外参数矩阵 (Extrinsic Parameter Matrix)。

为了求出摄像机的内参数和外参数矩阵, 需要至少 3 张以上的拍摄模板的图像建立的大量点对应关系, 这里通常采用角点特征匹配方法。这种方法依赖于规则特征的平面模板, 这样的标定板需要特别定制, 且在实际应用中, 与场景不相协调。本文通过使用 SIFT 特征建立匹配关系, 实现无标志模板的摄像机标定。系统整体流程如下

预处理

1. 打印一张标定模板图片, 贴在一个平面上
2. 从不同角度拍摄若干张模板图像

特征匹配

1. 提取 SIFT 特征点
2. 建立特征点映射关系

标定摄像机

1. 计算内参数
2. 计算外参数

优化求精

#### 4 三维重建过程

从本质意义上来讲,完成摄像机标定等价于三维重建。因为本文处理的目标是视频,所以对于每一帧图像都需要计算摄像机外参数,即旋转矩阵和平移矩阵。限于计算机的处理速度,场景完全的三角化重建将耗费大量的运算时间,在目前条件下很难满足视频的实时性需求。考虑到场景中的平面模板将占据大多数空间,本文仅对该平面进行重建,并在此基础上进行虚拟化处理。

旋转矩阵决定了平面模板相对于摄像机坐标系的空间姿态,因此需要计算出模板相对摄像机坐标系三轴的欧拉角 $\alpha, \beta, \gamma$ ,这里 $\alpha, \beta, \gamma$ 分别对应 $x, y, z$ 三轴。标定求得的旋转矩阵 $R$ 可以表示为

$$R = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix} = \begin{bmatrix} c_y * c_z & s_x * s_y * c_z - c_x * s_z & c_x * s_y * c_z + s_x * s_z \\ c_y * s_z & s_x * s_y * s_z + c_x * c_z & c_x * s_y * s_z - s_x * c_z \\ -s_y & s_x * c_y & c_x * c_y \end{bmatrix} \quad (3)$$

式中,

$$\begin{aligned} \gamma &= a \tan(r_4 / r_1) \\ s_z &= \sin(\gamma) \\ c_z &= \cos(\gamma) \end{aligned} \quad (4)$$

则有:

$$\alpha = a \tan \frac{r_3 * s_z - r_6 * c_z}{r_5 * c_z - r_2 * s_z}, \beta = a \tan \frac{-r_7}{r_1 * c_z + r_4 * s_z} \quad (5)$$

至此得到了三轴旋转角,即得到了平面模板相对摄像机

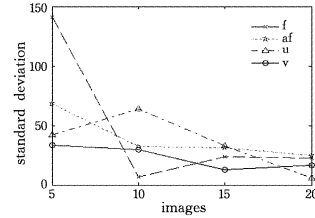
的空间姿态。

对于平移矩阵,可以直接获得图像模板相对于摄像机坐标系的平移量。以 $320 \times 240$ 的窗口为例,模板上的一点 $(x, y, z)$ ,可以通过求在窗口内平移的比例大致知道模板的位置,这个比例矩阵即为 $(x/320, y/240, z/\delta)$ 。因为窗口坐标系的深度 $\delta$ 是无穷大的,为了方便显示,可以为 $\delta$ 指定一个固定值。

#### 5 虚拟化应用与实验结果

摄像机标定是一种三维重建方式,因此,评价该方式的最重要的指标是重建的精度与效果。另外,尽可能达到实时处理的要求,对方法的推广也具有十分重要的意义。为了验证本方法良好的三维重建效果,在场景的平面模板上注册虚拟物体,通过统计不同图像下单帧的处理时间,评价本方法的实时处理性能。

张正友法的标定结果随序列图像数量的增加向实际参数逼近<sup>[5]</sup>。为了验证这一点,分别采用5/10/15/20张图像用于标定测试。实验数据如表1所列。 $\delta$ 代表标准差,从它的数值变化可以看出,随着图像数量的增加,标准差逐渐减小,即标定结果逐渐向实际参数逼近,如图5和表1所示。



$a$  是图像的长宽比,  $f$  是摄像机的焦距,  $(u, v)$  是摄像机的初始光心位置

图5 用改进的张正友标定法得到的摄像机内参数变化情况

表1 利用改进的张正友法得到的摄像机内参数(训练图像分别为5,10,15,20张)

Num	5 images			10 images			15 images			20 images		
	initial	final	$\delta$	initial	final	$\delta$	initial	final	$\delta$	initial	final	$\delta$
f	929.1	1210.9	140.9	839.91	826.47	6.721	627.70	674.83	23.56	990.25	945.07	22.59
af	764.53	901.27	68.37	689.35	753.73	32.19	618.86	556.29	31.28	754.05	705.12	24.46
u	196.41	280.40	41.99	107.05	234.36	63.66	148.41	214.44	33.02	189.33	201.01	5.840
v	172.7	239.81	33.53	113.61	173.11	29.75	115.08	140.80	12.86	133.94	166.71	16.38

在处理速度测试中,分别对SIFT特征点为271个和943个的模板进行标定,如图6所示。前者平均每帧处理时间为978ms,后者平均每帧处理时间为2121ms。从测试结果来看,处理速度并不理想。在实时化的处理时,为了保证帧图像的连续性,要求处理速度至少达到10帧/s,也就是说单帧处理时间低于100ms。系统的计算量主要来自于特征提取与摄像机标定两个步骤,涉及了大量的浮点计算,但目前可以使用的浮点运算优化手段很多,例如SSE, GPU加速等,因此提升的空间还是很大的。

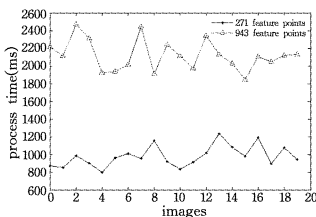
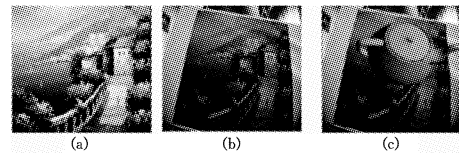


图6 单帧图像的处理时间(分别对应271和943个特征点的图像)

图7是实验的运行效果图,可以看到,实时注册的虚拟茶壶的空间姿态与位置能够很好地与平面模板相协调,真实感效果是令人满意的。



(a)模板图像 (b)摄像机下的模板 (c)注册虚拟物体

图7 本方法的实验效果

**结束语** 针对现有摄像机标定技术的标定模板进行研究,提高了张正友法的灵活性。文中用SIFT特征算法代替张正友法使用的角点算子进行匹配,使得无需特制用于标定的模板图像,就可提高其灵活性。实验表明,得益于SIFT算子在视频图像下的优秀不变性,任何具有稳定特征的图像都可以用于摄像机标定,改进后的张正友标定法大大提升了其实用性,并且仍然保持了较高的精度。此外,由于SIFT算法

良好的改进潜力,利用其改进算法——SURF<sup>[8]</sup>和PCA-SIFT<sup>[9]</sup>,可以进一步提高运行效率,对于摄像机视频场景的实时三维重建具有非常重要的意义。

参 考 文 献

[1] Saxena A. Make3D: Learning 3D Scene Structure from a Single Still Image[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(5): 824-840

[2] Abdel-Aziz Y I, Karara H. M-Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry[C]// Proceedings Symposium on Close-Range Photogrammetry. 1971: 1-18

[3] Tsai R. An efficient and accurate camera calibration technique for 3D machine vision[C]// Proceedings of Computer Vision and

Pattern Recognition. 1986: 364-374

[4] 杨长江,汪威,胡占义. 一种基于主动视觉的摄像机内参数自标定方法[J]. 计算机学报, 1998, 21(5): 428-435

[5] Zhang Z. A flexible new technique for camera calibration[J]. IEEE-PAMI, 2000, 22(11): 1330-1334

[6] David G L. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110

[7] 吴朝福. 计算机视觉中的数学方法[M]. 北京: 科学出版社, 2003

[8] Bay H. SURF: Speeded up robust features[C]// Computer Vision-ECCV 2006, PT 1, Proceedings. 2006, 3951: 404-417

[9] Ke Y, Sukthankar R. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors[C]// Proc. Conf, Computer Vision and Pattern Recognition. 2004: 511-517

(上接第 272 页)

此外,将本文方法绘制效果与文献[6,8]方法进行了对比,结果如图 11 所示。

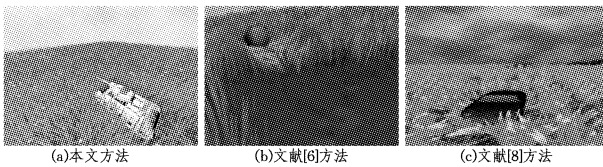


图 11 不同方法绘制效果比较

由于文献[6]方法是根据一个受影响区域来判断草叶是否与碾压物发生碰撞,没有实际的碰撞检测计算,其结果过于粗糙,因此难以反映真实的交互效果。但该方法绘制速度较快,平均帧速达到了 35fps。文献[8]方法采用基于包围盒的碰撞检测,碰撞精度不高,平均帧速为 45fps。本文提出的可控精度草体碾压算法,在真实感和实时性间获得了较好的平衡,在碰撞精度为 4 的情况下,平均帧速达到了 30fps。

从草体表达方式、碰撞检测、碰撞物外形要求、GPU 加速、草叶间相互作用以及层间动态过渡等几个方面,将本文方法与已有方法进行了详细比较,结果如表 2 所列。

表 2 本文方法与其他方法比较

	草体 表达	碰撞 检测	碾压物 外形要求	GPU 加速	草叶间 相互作用	动态 过渡
Guerraz <sup>[6]</sup>	混合式	×	无限制	×	×	×
牛立新 <sup>[7]</sup>	粒子系统	低精度	球体	×	√	×
Zhao <sup>[8]</sup>	混合式	低精度	无限制	√	×	×
本文方法	混合式	可控精度	无限制	√	√	√

由表 2 看到,与文献[6-8]相比,本文方法提供了更高精度的碰撞检测,使模拟效果更加逼真,同时碰撞检测的精度可根据实际需要进行调控,具有较大的灵活性。其次,与文献[7]类似,本文方法也考虑了草叶间的相互作用,但我们并不进行叶间碰撞检测,而是采用受力向量传递的策略,提高了计算速度。此外,与现有方法不同,本文方法考虑了不同草叶表达方法间的动态过渡,有效消除了层间动画不连续的问题。

**结束语** 草地场景的真实感动态模拟一直是计算机图形学领域的研究热点和难点。本文提出了一种草体碾压效果模

拟方法,该方法采用混合式草地场景表达方法构造出不同细节的局部草地场景,利用基于 GPU 的碰撞检测算法、受力向量传递策略、billboard 斜率控制等手段,对不同类型的草体实施变形计算,有效克服了传统方法处理速度慢、模拟精度差以及对碾压物外形存在限制等缺陷,在视景仿真、娱乐游戏、军事仿真等领域具有较好的应用价值。

参 考 文 献

[1] Boulanger K, Pattanaik S N, Bouatouch K. Rendering grass in real time with dynamic lighting[J]. IEEE Computer Graphics and Applications, 2009, 29(1): 32-41

[2] Wang Y M, Jiang X F, Ge L. Modeling and rendering of dynamic grassland scene in the wind[C]// Proceedings of the International Conference on Information Engineering. Washington DC: IEEE Computer Society, 2009: 69-74

[3] Habel R, Wimmer M, Jeschke S. Instant animated grass [J]. Journal of WSCG, 2007, 15(3): 123-128

[4] Bakay B. Animating and lighting grass in real-time[D]. Vancouver: University of British Columbia, 2003

[5] 王佳声,唐好选,孙振义. 自然风力作用下草体运动的模拟方法[J]. 哈尔滨工业大学学报, 2008, 40(3): 466-470

[6] Guerraz S, Perbet F, Raulo D, et al. A procedural approach to animate interactive natural sceneries[C]// Proceedings of the 16<sup>th</sup> International Conference on Computer Animation and Social Agents. Washington DC: IEEE Computer Society, 2003: 73-78

[7] 牛立新. 真实感图形算法与实时绘制技术研究[D]. 北京: 首都师范大学硕士学位论文, 2007

[8] Zhao X K, Li F X, Zhan S Y. Real-time animating and rendering of large scale grass scenery on GPU[C]// Proceedings of the 2009 International Conference on Information Technology and Computer Science. Washington DC: IEEE Computer Society, 2009: 601-604

[9] 范昭炜,万华根,高曙明. 基于流的实时碰撞检测算法[J]. 软件学报, 2004, 15(10): 1505-1514

[10] Philip J S, David H E. 计算机图形学几何工具算法详解[M]. 周长发,译. 北京: 电子工业出版社, 2005