

硕 士 研 究 生 读 书 报 告



题目 照片旅游：在3D中探索照片集

作者姓名 王国军

作者学号 nb15033

指导教师 李启雷

学科专业 移动互联网与游戏开发技术

所在学院 软件学院

提交日期 二○ 一五 年 十二 月

Photo Tourism: Exploring Photo Collections in 3D

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: Qilei Li

By

Guojun Wang

Zhejiang University, P.R. China

2015

摘要

本文提出一个系统，这个系统采用了新颖的3D界面，交互式浏览和探索大型非结构化场景照片集合。这个系统由一个基于图像的建模前端，它会自动计算每一张照片的角度，以及稀疏模型对应的场景和形象的三位模型。该系统主要是使用基于图像的渲染技术使照片之间能够顺利的转换，同时也可以实现全景3D导航和一套图像和几何世界的探索，并伴随产生一些辅助信息，如架空地图。这个系统也可以很容易地构建旅游风景区或历史地点，以及注解照片的图像细节。这些细节会自动的转移到集合中的其他有关图像上。该系统应用了几个大的个人摄影集以及从互联网上照片共享网站中收集的图像。

**关键词**：基于图像的渲染， 基于图像的建模，图片浏览，结构体

Abstract

We present a system for interactively browsing and exploring large unstructured collections of photographs of a scene using a novel 3D interface. Our system consists of an image-based modeling front end that automatically computes the viewpoint of each photo- graph as well as a sparse 3D model of the scene and image to model correspondences. Our *photo explorer* uses image-based rendering techniques to smoothly transition between photographs, while also enabling full 3D navigation and exploration of the set of images and world geometry, along with auxiliary information such as overhead maps. Our system also makes it easy to construct photo tours of scenic or historic locations, and to annotate image details, which are automatically transferred to other relevant images. We demon- strate our system on several large personal photo collections as well as images gathered from Internet photo sharing sites.

**Keywords：**image-based rendering, image-based modeling, photo browsing, structure from motion

1引言

基于图像的绘制的一个中心目标是要基于场景的照片集合唤起存在的内心的场景。过去几年见证了朝着这个目标取得的重大进展，通过在研究界和商业产品的图像合成方法的研究，有如全景工具的商业产品，我们的梦想之一是这些方法有一天会允许虚拟旅游世界的有趣和重要场所。

与此同时，数字摄影，连同互联网结合在了一起，真正意义上实现了照片的共享。例如，在谷歌图片中搜索“巴黎圣母院”，会返回超过15000张照片，从无数的观点来看，获取现场的详细级别，照明条件，如海的儿子、几十年等等。但是不幸的是，共享照片超过了技术的扩散等浏览收藏，谷歌（www.google.com）和Flickr（www.flickr.com）返回了一页又一页的缩略图，用户必须进行梳理。

在本文中，我们提出了一个系统，用于浏览和组织的热门网站，它利用底层的现场常见的3D几何形状的大型照片集。我们的方法是基于计算，从图像本身，在摄影师的位置和方向，随着场景的稀疏三维的几何代表，采用先进设备，最先进的是基于图像的建模系统。该系统处理广泛不同条件采用不同摄像机无组织的光子集合。使用了以下的功能：

场景可视化：通过照片可以看到遍布世界的各个城市；

基于对象的图片浏览：可以让人看到包含此对象的场景或者更多部分的图像；

我在哪：可以告诉我我在哪里时拍下了这张图片；

我究竟在看什么：可以告诉我有关对象可见的图像传输注释从类似的图像中获得更多的信息。

**2 系统概述**

所谓在本节中，主要是为该系统的特定功能进行概述，有关功能的视觉展示的介绍。

在拍摄中，我们是基于使用照相机的姿势（位置、方向和视野）以及稀疏3D场景信息来创建新的接口用于浏览大量的照片。知道相机的姿势使把图像放置在一个共同三维坐标系统中，并且允许用户由三维空间中的一个图像移动到另一个使用我们的相片浏览搜索现场。我们使用变形技术提供照片之间的平滑过渡。此外，重新构造特征提供了一个稀疏但是有效的三维可视化的场景作为背景。除了展示这些功能作为一个简单的展示之外，该系统为使用者提供了一个更好的现场感非常真实的，外观新颖的渲染技术。

同时标准的3D导航控件——teraction，并不适合其他任务。例如，假设您希望看到图像场景中的一个特定的对象或区域。支持这种基于对象的浏览，我们允许用户周围绘制在一个图像中的感兴趣的对象的箱，该系统然后平稳的移动到“最好的”视图的对象。我们提供被以几何体的较为稳定的一组相同视图的对象，为了更有效的比较场景的外观，比如天、季节、天气条件、时间不同的能力。

通常情况下，我们有兴趣了解更多有关图像的内容，例如，“这雕刻的是什么？”或“什么时候建这栋大楼”，大量的这种形式已经存在于旅游指南中，地图注解图像内容，和互联网资源，如维基百科（www.wikipedia.org）和Flickr。但是，你可能会观看在任何特定时间的图像（例如，从您的手机摄像头）可能不会有这样的效果。但是我们的系统的一个主要的特点是，图像之间的自动传输的注释能力，因此大约在一个图像中的对象的信息被链接到包含相同对象的所有其他图像。

我们系统的核心是一个坚固的结构，从运动方式上为重建所需的3D信息。我们使用的方法首先是计算图像，US－ING描述符是稳健相对于变化的姿势、尺度和照明之间特征相对应，并运行一个优化来恢复相机参数和这些特征的3D位置。结果，spindences和3D数据启用的具有上诉所有特征的系统。

**3 相关工作**

在我们主要有三个主要类别的相关工作：基于图像的建模，基于图像的渲染和图像浏览、注释。

**3.1 基于图像的建模**

基于图像的建模（IBM）是从输入图像到创建三维模型的过程。我们基于图像的模型系统是基于运动（SFM），其目的是恢复摄像机最近工作中的参数和从图像序列获得稀疏三维几何场景。特别是，我们的可持续森林管理的做法是于布朗使用各种数据集，进行若干的修订，以提高固定性。这些包括初始化新的相机使用姿势估计，以帮助避免局部最小值，不同的启发式选择最初的两张图片，检查重建点状态良好的，添加到现场，并且使用焦距标记信息的图像。Schaffalitaky和Zisserman[2012]提出了另一个相关技术，用于校准无序的图像集，有效匹配图像之间的结合点。虽然两种方法解决同样的SFM的问题，我们所做的，他们是更简单的测试数据集更有限的成像条件的变化。

IBM的一个特定的应用程序创建大规模的建筑模型，著名的例子包括半自动外交事务委员会ADE系统，该系统用于重建引人注目的加州大学伯克利分校的摄像机；自动架构重建系统诸如迪克等人；以及麻省理工学院城市扫描项目，成千上万的校准图像从一个捕获——strumented钻机来计算三维模型的麻省理工学院的校园，还有几个正在进行的学术和商业项目——大规模的城市场景重建，包括4D城市项目（www.cc.gatech.edu），其目的是打造亚特兰大从histor-iCal的照片的时空模型；斯坦福CityBlock项目，它采用了城市街区的视频以创建多角度的图像和Pollefeys和Niste R中的城市布局项目。

**3.2 基于图像的渲染**

基于图像的绘制（IBR）现场致力于从一组输入光子的合成场景的新视野的概率LEM。阿斯MovieMap项目是这个领域的先行者，具有开创新的作用，其中数千科罗拉多州阿斯彭的图像是从行驶的汽车上捕获的，注册到一个城市的街道地图，并储存在激光视盘。甲烯体健的用户界面交互式地通过图像的移动作为用户的期望路径的功能。其他功能包括导航地图的覆盖全市的图像显示，并以触摸的观点和跳跃当前场的任何建筑物到建筑物的立面的能力。该系统该允许安装元数据，如餐厅的菜单和历史图像与个别建筑物。我们的工作可以看作是一种能够自动创建图像的无组织的集合的MovieMaps。（相比之下，阿斯MovieMap涉及一个团队工作了十几年）一些可视化、导航和AN－符号功能类似于原MovieMap的工作，但是改为了广义的形式。

最近的工作中IBR一直专注于技术的新试图合成。在应用中，阿里亚加等方面工作最接近于我们在其使用的大集合在整个建筑空间拍摄图像的。同一作者的计算解决跨多个图像一致性比赛为IBR的目的问题，然而，我们的图像被不同的摄影师获得，而不是被采取在固定网格带有引导的机器人。

在对比中IER大多数现有工作，我们的目标是合成从不同视点本身逼真的图像，但以浏览在一个三维空间上下文给出的几何感照片特定集合底层的现场。因此，我们的方法使用了一个近似平面为基础的试图内插法和非真实感的背景场景结构呈现。我们侧重重建整个表面模型的更多挑战性的问题、光场或像素精确视图插值。这样做的好处是，我们能够与输入的图像进行操作。

**3.3 图片浏览、检索和注释**

有很多技术和商业产品用于浏览照片，同时大量的研究表明人们往往会整理照片，许多这些技术使用了元数据，例如关键字，摄影师，或时间，作为光组织的基础。

同时出现了使用地理定位信息以促进照片浏览的兴趣，特别是，世界媒体的广泛交流安排上的2D地图图像，PhotoCompas基于时间和地点的集群的图像，RealityFlyThough使用界面的想法与我们相似的探索，从摄像机仪器与全球定位系统和倾斜传感器的视频，以及Kadobayashi和田中提出来的检索使用接近虚拟摄像机图像的接口，在这些系统中，某一地址被从GPS获得的或手动指定，因为我们的方法不需要GPS或其他仪器，所以它具有适用于现有的图像数据库，并从因特网光子tographs的优点。此外，我们的许多做法的导航功能利用图像特征的对应和稀疏的三维几何计算，因此超越了有可能在这些以往的基于位置的系统。许多的技术也存在于从数据库中检索年龄的相关任务，与我们工作的一个特定的系统是视频谷歌（不要与谷歌自己的视频搜索混淆），这允许用户在视频的一帧选择一个查询对象和有效地找到在其他帧的对象。我们基于对象的导航模式使用了类似的想法，但是扩展到了3D领域。

一些研究人员已经研究技术的自动和半自动图像注释，尤其是注释转移，本地化系统采用接近传输地理参照照片之间的标签。我们的系统的注释功能的优点在于，我们的特征对应使转印在更精细的粒度；我们可以把图像之间传输特定对象的注释和地区，同时考虑到闭塞和这些对象的动作下，进行改变。这个目标是类似的增强现实（AR）的方法，其中还寻求来注释图像。虽然大多数AR方法注册一个三维计算机生成模式的图像，我们不是2D图像标注转移到其他图像。生成的注释内容，因此容易的多，注释转移也已探索了视频序列。

最后，约翰逊和奇波拉已经开发出一种系统，其中用户可以拍一张照片，将其上传到它相对应的服务器上，并获得位置信息。我们的系统还支持另外还有许多其他功能（可视化、导航、注释等）应用程序。

**4 重建相机和稀疏几何形状**

我们的系统需要大约相对的单元中，取向、和固有参数，如焦距为集合中的每一张照片，以及稀疏三维场景几何形状准确的信息。我们系统的某些功能需要摄像机的绝对位置，在一个地理参考坐标系，一些信息可以提供有GPS设备和电子的COM通行证，但现有照片绝大多数缺乏这种信息。在我们的系统中，我们并不依赖相机或任何其他设备为我们提供的位置，方向或几何体。相反，我们计算利用计算机视觉技术本身的图像信息。我们首先检测特征点在的每个图像，然后匹配图像之间的特征，最后运行一个反复的，稳健的可持续性的森林管理过程来恢复相机的参数。

**4.1 关键点检测和匹配**

第一步是找到特征点中的每个图像，我们使用，由于它的不变性图像转换的SIFT关键点检测。一个典型的图像包含几千个SIFT关键点，其他特征检测器也可能被用来进行检测，除了关键点的位置本身，SIFT提供了局部描述符对应的每个关键点。接着，对于每队图像，我们匹配关键点描述符中的一对之间，使用近似最近邻的Arya包，在每次迭代RANSAC，我们使用八点AL-gorithm，其次是非线性重新禁闭计算的候选人基本矩阵，最后，我们会删除异常值，恢复ERED基本矩阵匹配，如果剩余的数量小于20，可能就考虑删除所有的数据。

**4.2 基于运行的构造**

接下来，我们恢复出一组的照相机参数和一个三维位置的每个轨道，该回收的参数应该是一致的，在该重投影误差之间，每个轨迹和其对应的图像特征之间距离的总和，是MIN-imized。这个最小化问题是配置为非线性的最小二乘法问题，并解决了算法，如列文伯格－马夸尔特。而不是估计一次参数，所有的摄像机和轨道，我们采用循序渐进的方式，在同一时间将在一台摄像机。

首先，我们估计一台摄像机的参数，这最应该有大量的数据，同时也应该有大量的基准，从而使该观测点的三维位置是最好的条件，因此，我们选择了对图像的数量最多的分析。接下来，我们增加一个摄像头的优化，我们选择观察人数最多的曲目，其某些地区已被记录在3D摄像头中，并采用直接线性变换，初始化新相机的外部参数。最后，我们添加了新摄像头观察到优化轨道，如果它是由至少一个其他回收相机观察到的，和如果三角磁道给出了其位置的良好条件的估计，重复这个过程，一个摄像机的时间，直到没有剩余的相机观测到任何重新构建3D图像。以在每次迭代的最小化目标函数，我们使用Lourakis和Argyros的稀疏调节库中，重建场景后，我们选择性地运行后处理步骤，以检测场景中的3D线段使用线段重建技术。

**4.3 地理注册**

在SFM的过程中，需要估计摄像头的位置，位置估计的过程的最后一步是使模型与地理参考图图像或地图（如卫星图像、平面图、或数字高程地图），以使每个照相机的绝对地心坐标可以确定，这个步骤是不需要我们的浏览器系统工作的，但是需要人来操作（如现实鸟瞰图） 。

所估计的相机位置，在理论上，与由一个相似的绝对位置相变换，在一些情况下，回收的场景不能被对齐到地理位置利用相似变换引用的坐标系，如果SFM过程失败了，以获得所诉场景的充分度量重新构造，就会发生这种情况，由于在回收点和相机位置的低频漂流而引起误差，这些误差源不具有显著的作用，对许多在浏览器界面中使用的导航控制的，因为错误通常不是局部的显示，但是有问题的，当一个精确的模型是期望的。

其中回收场景的一种方式是通过添加约束到SFM，牵制一组地面控制点或相机稀疏已知的某一地址的阳离子进行优化，或者用户可以手动制定点或照相机的位置。

**5 结论**

我们使用了多个数据集评估我们的系统。前两组采取更多的控制设置（即一个人用一台摄像机和镜头）：布拉格，在两天内拍摄的，捷克共和国的布拉格古镇广场的一组197的照片和长城，沿长城拍摄的一组120张照片（其中82最终注册）。

我们还用包含从Flickr下载的图片的图片集来实验。在每种情况下，我们的系统在整个图片集上检测和匹配特征点并且自动鉴别和注册与场景一个练习部分对应的字集。有四组照片集如下：Notre Dame是一套597张的巴黎圣母院大教堂的照片。这些照片是从2635与notredame为关键字搜索的结果匹配的照片开始注册。Yosemite是一个325张约塞米蒂国家公园半圆顶的照片集，从1882张匹配halfdome AND Yosemite 搜索的照片开始注册。TreviFountain是一个360张罗马特雷维喷泉的照片集，从466张匹配tre AND rome搜索的照片开始注册。Trafalgar Square是一个278张特拉法尔加广场的照片集，从1893张匹配trafalgarsquare搜索的照片开始注册。

平均每个数据集的投影误差是1.516像素（我们注册的长短图像的平均尺寸是1611个像素，最近成功1128）。这些数据集的可视化，他们展示了photo exporer的功能，包括在几个数据集的对象选择，相关的图像选择，变形，注解转移。

对于半圆顶数据集，初步构建模型后，我们用之前描述的方法把它对齐数字高程地图。然后，我们注册Ansel Adams的“Moon and Harf Dome”这张历史照片到数据集，拖动并把它放在模型上，以及上文中提到的Ansel Adams拍摄照片的位置上看到的场景的合成绘制。

参考文献

[1]Photo Tourism: Exploring Photo Collections in 3D. Noah Snavely. Steven M.Seitz Richard Szeliski. SIGGRAPH，2006．