

硕 士 研 究 生 读 书 报 告



题目 动作驱动的三维室内场景演化

作者姓名 黄冠积

作者学号 21951086

指导教师 李启雷

学科专业 软件工程

所在学院 软件学院

提交日期 二〇一九 年 十二 月

Action-Driven 3D Indoor Scene Evolution

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: LiQiLei

By

HuangGuanJi

Zhejiang University, P.R. China

2019

摘要

介绍了一个动作驱动的3D室内场景演化框架，其目标是模拟场景是如何被人类动作改变的，特别是由动作所需要的对象位置。为此，开发了一个动作模型，每种类型的动作结合一个或多个人体姿势、一个或多个对象类别和属于这些类别的对象的空间配置信息，这些信息总结了动作的对象-对象和对象-人的关系。重要的是，所有这些信息都是从带注释的照片中获得的。通过分析学习动作之间的相关关系来指导动作图的构建。从一个初始的3D场景开始，从动作图中对一系列动作进行概率抽样，以推动渐进的场景演化。每个动作触发都基于对象的共现和动作模型的空间配置来确定对象的位置。

**关键词**：动作驱动、动作模型、3D演化

Abstract

This paper introduces an action-driven 3D interior scene evolution framework, which aims to simulate how the scene is changed by human actions, especially the object positions required by the actions. For this purpose, an action model is developed, in which each type of action combines one or more human postures, one or more object categories and the spatial configuration information of objects belonging to these categories, which summarizes the object-object and object-person relationships of actions. Importantly, all of this information is taken from annotated photos. Through the analysis of the correlation between learning actions to guide the construction of action diagram. Starting with an initial 3D scene, probability sampling of a series of actions from the animation diagram is performed to promote progressive scene evolution. Each action trigger determines the location of the object based on the co-occurrence of the object and the spatial configuration of the action model.

**Keywords：**Action driver, action model, 3D evolution

1引言

我们生活在一个3D世界中，我们不断地对周围的3D场景环境进行操作和交互。随着时间的推移，由人类活动引起的物体运动推动着场景的演变。数字3D场景能否以这种动作驱动的方式处理，这似乎是一个很自然的问题。从新兴的VR/AR应用到数据驱动的场景分析，随着对三维场景数据特别是室内场景数据需求的不断增长，场景生成技术越来越受到图形和视觉领域的关注。

以往的场景生成是通过从3D场景示例中学习到的概率分布而产生，然而大多数可以从公共数据仓库(如3D仓库)获得的室内场景都具有展览的组织性和清洁度，这些场景大多是精心设计的，而在现实生活中，我们的办公室、实验室和卧室通常比较乱，这种随机的杂乱性是很难通过学习3D场景摆布概率学到的，这就导致了场景生成效果一定程度的失真。

为此引入动作模型，支持动作驱动的3D场景演化，其目标是模拟场景是如何被人的动作改变的，特别是被动作所需要的物体位置改变的，每个动作触发都基于对象的共现和动作模型的空间配置来确定对象的位置，这就使得场景的生成是连续的，是一个个动作的结果，最终得到的效果更加自然。

**2 动作模型**

2.1 数据准备

使用了Microsoft COCO (Common Objects in Context)数据库，其中包含了大量的室内场景和日常人类活动的照片，每张照片中都有大量的对象分段、标签和文字说明来描述人类行为。首先分析其中的照片来收集与某些动作相关的照片。然后，相同动作的photos根据照片中的对象类别和人类姿势进行聚类。通过检查相关物体的重叠来计算动作过渡概率。

3D人体姿态恢复。COCO数据库只提供了人体的二维图像区域，而动作模型中的人体姿态定义是三维的。因此，需要从2D图像中恢复最合理的3D人体姿态。为此，首先手动注释每个动作实例图像中的人类骨骼关节。为了获得一个稳定的3D姿态估计，还需要手动提供尽可能多的缺失关节的似是而非的位置，比如在部分遮挡的情况下，输出是一个三维人体骨架和一个弱摄像头投影矩阵的输入图像。

2.2 动作模式定义

描述一个室内场景中的动作模式为：< T K H; C, D >, 其中T是动作类型,K是关键对象指定动作发生的地方,H是一个代表性的3维人体姿势, C存储每个对象发生时间的概率分布, D指定组成的空间配置:为每一个对象,总结其相对于人类的姿势和其他对象位置信息。根据定义，一个动作模型可以通过(T, K, H)的组合来唯一标识，(T, K, H)表示“在(K)处执行什么动作(T)，当时的姿势的怎样的（H）”。通过考虑这五个因素，可以在丰富的环境下模拟动作，比如躺在床上看书和坐在桌子上看书。

2.3 行动图

为了保证动作的平滑性，两个动作之间并非随机的，而应该是一个动作序列中的部分，序列中两个相邻的动作应该共享某些组成对象，为此建立了行动图。在学习阶段，从Microsoft COCO数据库中学习动作节点，构造动作图。从COCO数据库开始，首先为每种动作类型找到一组实例映像。然后，在人工标记关节的帮助下，为每个实例恢复三维人体姿态，在此基础上，利用恢复后的三维人体姿态，推导出每个实例的关键对象和三维对象布局。然后，通过分析对象的发生和空间布局，构建每个簇的动作模型。最后，在行动节点上构造一个行动图，并根据行动节点之间的相关性计算动作相互变换的概率。

由动作i变换成动作j概率定义为：

其中表示动作i和动作j之间的关联性，计算公式为：

其中表示动作i和动作j的场景对象的重合度：

其中， 当时返回1，否则返回0；

后半部分，o表示的是动作场景中的对象，C表示的是直方图，，取直方图的较小值，表示两个动作场景交集，而，取直方图的较大值，表示的是两个动作场景的并集。

2.4 图像适应

动作图是从COCO数据库中学习到的所有类型的动作，并涵盖了各种场景的动作:卧室、厨房、办公室等等，将其应用于特定场景需要一个图形适应的预处理过程。也就是说，给定一个有关键对象的输入场景，如果关键对象在场景中丢失，需要修剪某些动作节点。例如，如果场景中没有银器，那么用餐的动作节点可能在这个动作的实现会导致银器的插入;但是，如果餐桌(关键对象)不见了，就不应该允许吃饭。关键对象通过对上面公式中动作i到动作j的概率影响，使得不合理的行为不发生，而对于细枝末节，由于其所占比重不高，因而不能直接决定动作的发生与否，相反会在确定行为发生后，补全细节，以使得动作自然合理。

2.4 动作实现

在一个动作完成后，所涉及的对象保留在场景中，这意味着一个操作是在其所有祖先在操作序列中创建的上下文中执行的。为了从序列中实现一个新的动作，首先将关键对象K、人体姿态H放入场景中;然后通过对发生模型C进行采样，收集活动对象O的集合，并将其置于场景中，使其空间布局符合分布D;最后，重新定位非活动对象O，使之拟合人体姿态(H)，这样它们就不会妨碍当前的操作。

放置活动对象。给定动作的人姿态和关键对象，需要找出要插入的组成对象集O，以及它们在场景中的位置和方向。首先，根据学习到的模型C生成组成对象集。对于场景中已经存在的活动对象，有两个选项—重用它或插入一个新对象。为整个场景中每个对象类别的出现时间预先定义了一个上界。例如，一个场景最多可以包含一个显示器，一个键盘，但是两个咖啡杯，十本书。新对象在其出现时间到达上限之前插入场景;否则，只允许重用对象来实现新的操作。

然后，在第二步中按出现概率和大小递减的顺序一次放置一个对象。对象的概率计算公式如下：

其中L(o)是碰撞惩罚项，如果o与场景中任何其他物体发生碰撞则为0，否则为1。悬垂惩罚项S(o)阻止o从支承面边缘悬挂，将o的边界投影到支撑面上，计算投影与支撑区域之间的交集面积A(o)，如果A(o)>=0.5则S(o)=1,否则S(o)=0，也就是说，如果o不到一半的体积悬挂在支撑面上，则认为o的位置是不合理的。

其中P(o, H)为对象-人分布概率，P(,)为对象-对象分布概率，f(,)是对象和共现的概率。

放置非活动对象。对于非活动对象，除了碰撞和悬挂拒绝之外，还必须保证不妨碍动作的进行，并且在动作过程中应尽可能接近原来的位置。与活动对象类似，按照概率大小依次将非活动对象插入场景，对于每一个非活动对象，统一尝试10000个可能的位置，挑选使得以下式子最大化的位置为最佳位置。

其中是碰撞惩罚项，悬挂惩罚项，表示是否影响动作的发生，如果在运动区内则为0，否则为1；新生成的对象位置距离原位置的距离，用于防止非活动对象的剧变。

**3 动作模型的创新性**

以往的3D场景模拟合成，采用了场景的整体视图，导致生成的场景必须与整体示例相似，并且可能属于相同的场景类别。相反，动作驱动的进化是一种程序化的、更原子化的场景生成形式，与场景类别无关，它的训练拟合标准是在给定的场景环境中什么动作是适用的。例如，“坐在桌子上看书”的动作就适用于任何有桌子的场景。一次应用一个动作允许更多的本地控制并生成更高粒度的场景。主要创新性在：（1）渐进的场景生成方法，在不影响真实性和自然性的前提下，生成了一组不断演化的粒状3D场景，显示出比以前作品更高的场景复杂性和混乱程度。（2）从注释的照片训练学习，而不是三维场景的样本。这使能够利用更丰富的数据源进行动作驱动的场景处理。由于动作模型是对动作进行分析，不依赖于整体的3D环境，只需对普通的图片添加好动作标签，即可进行训练，简化了训练条件（3）一个更完整的动作模型，它考虑了群体交流，以及多个对象的共同出现和联合放置，允许对象重新定位和插入。

动作模型是从只有空间约束的带注释的照片中学习的，是在人-物关系和物体共现的条件下发展出一个物体放置的场景，因此对于不同于训练环境的场景同样使用，具有较强的推广性。

**4 应用**

从新兴的VR/AR应用到数据驱动的场景分析，随着对三维场景数据特别是室内场景数据需求的不断增长，场景生成技术越来越受到图形和视觉领域的关注。应用动作驱动的场景演化方法，可以复制室内场景在现实生活中的演化过程，产生连续的一系列逼真的虚拟三维场景。正是由于其是靠动作驱动的，除了以往整体场景布置的真实性外，某一动作地发生导致场景发生变化这一过程，不论是与动作相关的对象，还是其他无关对象，都会相对平滑地过渡到下一场景，使得总体场景变化更加真实、自然。

参考文献

[1] Ma R , Li H , Zou C , et al. Action-driven 3D indoor scene evolution[J]. Acm Transactions on Graphics, 2016, 35(6):173.