

硕 士 研 究 生 读 书 报 告



题目 3D动画的连续性编辑

作者姓名 曾强

作者学号 21860401

指导教师 李启雷

学科专业 计算机技术

所在学院 工程师学院

提交日期 二○一八 年 十二 月

Continuity Editing for 3D Animation

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Computer Science and Technology

Major Subject: Computer Science and Technology

Advisor: Qilei Li

By

Qiang Zeng

Zhejiang University, P.R. China

2018

摘要

我们描述了一种基于优化的方法，用于从3D动画中自动创建编辑良好的电影。虽然以前的工作主要集中在放置相机以产生漂亮的动作视图的问题上，但是切割和粘贴所有镜头的问题可用的相机从未得到广泛的解决。在本文中，我们回顾了文献中编辑错误的主要原因，并提出了一种依赖于这种错误最小化的编辑模型。我们做出了似乎合理的半马尔可夫假设，从而产生了一种计算效率高的动态编程解决方案。我们还展示了我们的方法可以生成具有不同编辑节奏的电影，并通过用户研究验证结果。结合最先进的电影摄影技术，我们的方法因此有望显着扩展虚拟电影制作的表现力和自然性。

**关键词**：3D动画，连续性编辑工具，电影摄影技术, 虚拟电影制作

Abstract

We describe an optimization-based approach for automatically creating well-edited movies from a 3D animation. While previous rk has mostly focused on the problem of placing cameras to produce nice-looking views of the action, the problem of cutting and pasting shots from all available cameras has never been addressed extensively. In this paper, we review the main causes of editing errors in literature and propose an editing model relying on a minimization of such errors. We make a plausible semi-Markov assumption, resulting in a dynamic programming solution which is computationally efficient. We also show that our method can generate movies with different editing rhythms and validate the results through a user study. Combined with state-of-the-art cinematography, our approach therefore promises to significantly extend the expressiveness and naturalness of virtual movie-making.

**Keywords**: 3D animation, lack continuity editing tools ,cinematography, virtual movie-making

1引言

由于低成本的输入设备的使用，高分辨率3D模型的广泛可用性和创建新的几何和动画内容的便捷，开启了许多成为数字3D故事讲述者的可能性。一个令人信服的插图是引擎电影社区的兴起，包括创造电影依靠内容和渲染引擎，而这种引擎借助于ad-hoc技术从电脑游戏中借鉴或专用建模工具。

然而，迄今为止，明显缺乏可用工具来轻松创建电影摄影（定位摄像机以创建镜头）并执行此类故事的编辑（在创建的镜头之间选择适当的剪辑）相机。制作电影需要知识大量的经验规则和既定的惯例。特别是连续性编辑-创建确保视觉连续性的一系列镜头-是一项复杂的工作。大多数3D动画包缺乏连续性编辑工具，需要自动方法，至少部分地支持用户的创作过程。

以前在自动电影编辑方面的贡献集中在混合人工智能和计算机图形技术的生成方法上。但是，评估一下电影剪辑的质量（无论是由机器还是由艺术家生成）是一个众所周知的难题。一些贡献提到了启发式选择多个编辑解决方案没有进一步的细节，而其他编辑解决方案最小化了成本函数。此外，切割的精确时间尚未实现解决了切割节奏（每分钟拍摄次数）及其在确定电影节奏中的作用。大多数这些方法产生了一种被专业编辑所知的反应式编辑风格作为拉网风格，机械切入新的扬声器或动作。

在本文中提出了一个连续性编辑模型3D动画，为自动创建电影序列提供通用解决方案。这个模型将连续性编辑过程编码为通过编辑图形搜索最佳路径。在这个编辑图中，anode表示时间步长（镜头的时间片段），并且弧表示两个摄像机之间的过渡，从摄像机到相同的摄像机（无切割）或另一个摄像机（切割）。优化过程使用动态编程，在半马尔可夫假设下最小化所产生的错误沿着三个标准：镜头的质量（关于正在展开的行动），尊重连续性编辑规则和尊重有充分根据的模式节奏（削减节奏）。半马尔可夫模型之前已被用过信息提取，演讲一代和计算机视觉。这是他们第一次被建议作为电影编辑的计算模型。此文的贡献是：（i）3D动画连续性编辑的详细形式化，包括彻底的视觉特性和连续性规则的数量（ii）用于自动编辑的最佳算法，其中参数例如可以控制起搏，从而显着提高编辑工具的表现力，以及（iii）通过用户评估比较模型验证使用最佳编辑和原始编辑现有电影采用退化的方法。

**2 相关工作介绍**

从三维动画生成电影的问题可以被分解成（i）选择镜头的问题放置相机（电影摄影问题），以及（ii）决定何时将这些镜头组合成电影（电影编辑问题）。在本节中，按时间顺序回顾以前的关于电影编辑问题的工作。

Christianson等介绍了声明性相机控制语言作为一般框架为电影摄影和电影制作基于习语的解决方案电影编辑问题。电影成语是获得良好的电影摄影和在任何给定情况下编辑，类似于基于案例推理的情况。因此，DCCL使用会话习语拍摄对话，拍摄战斗的战斗习惯等然后，DCCL通过拍摄每个电影树来构建电影树所有可用习语中的场景，包括一个启发式评估器，用于检查序列的视觉质量，选择质量最高的那个。评估标准是只是简单地提到了，相比之下，本文提供了一个完整的可以用于以原则和可重现的方式使用所有可用相机找到最佳编辑解决方案的评分函数的描述。

Tomlinson，Blumberg和Nain2000提出了一个完整的建议用于产生表现性电影的自动方法。他们的系统使用情感自动相机状态和目标，在多个视觉之间进行选择目标-例如特定字符，或彼此交互的两个字符，或与之交互的一个字符它的环境。虽然他们的系统有趣的激励在情感和视觉上的镜头，蒙太奇的规则没有强制执行，并且通过临时规则解决了镜头的节奏。KennedyandMercer直接解决计划完整序列的问题必须知道的给定行动序列的镜头提前。用户可以选择快速编辑还是慢速编辑有两种不同的风格。作者使用深度优先前锋链式规划器，只能评估少数可能的计划并强制执行单一连续性编辑规则（“不跨越行动”）。

与以前的作品不同，Darshak将输入作为输入的显式声明表示故事目标，包括所有行动，行动之间的因果关系和排序约束。故事目标定义于少数戏剧性情境的术语，以及分层次序因果链接计划器计算摄像机计划，由一系列实现的镜头组成理想的故事目标。Darshak在激励镜头方面走了很长的路，但实际的摄影和编辑并未得到评估。Cambot是一个电影制作系统，其中发现的镜头选择使用动态编程解决优化问题。场景表现为一系列不重叠的戏剧性节拍，并且他们的方法评估不同位置的字符（阻塞）和相机每个节拍的选择。虽然作者也利用动态编程，他们的方法与（Elson和Riedl2007）。首先，搜索一组更大的可能解决方案，通过评估更多的镜头过渡以更精细的粒度（每一帧，而不是每一拍）。因此，方法可以选择拍摄之间的精确切割点，并保持对镜头变化率（切割速度）的控制。其次，详细描述了实现一个的成本函数广泛的电影剪辑规则。成本函数有一个半马尔可夫属性，允许有效地优化在这个更大的解决方案空间第三，作者的方法没有要求将场景分解成线性序列节拍，并且与重叠动作同样有效对话。

Assa，Wolf和Cohen-Or提出了一种全自动编辑过程，可以在多个摄像机之间实时查看人体运动。通过测量每个摄像机3D场景中人体运动之间的相关性来计算镜头之间的排名，和这些动作的屏幕2D投影（更大相关性越强，射击效果越好）。衡量侵蚀的一种方法目前的观点被用来激励两者之间的削减观点，同时强制执行连续性规则。在实时电影摄影的类似背景下，Lino等人提出了一种系统自动计算每个帧的集合通过空间分区的不同视点。编辑然后通过将连续性规则编码为来执行处理过滤删除不一致视点的运算符。

Markowitz等人扩展了电影成语方法通过用行为树替换有限状态机。他们假设，执行180度规则和30度规则单行

**3 系统概览**

在本文中，作者将电影编辑问题作为半马尔可夫链空间的优化问题。此系统将3D动画场景作为输入，包括a世界事件的流动，以及从不同摄像机拍摄并覆盖整个场景的一系列冲动。然后对关于三个关键方面的可能编辑进行排名：（i）有多少镜头传达展开的动作，（ii）多少连续性编辑规则强制执行和（iii）输入切割节奏是多少尊重。

给定具有任意复杂动作的3D动画场景a和从中选择的匆匆（即未编辑的镜头）M相机，半马尔可夫链是一系列状态（镜头）具有持续时间dj的sj，根据sj，dj和a上的概率模型选择。下一个概率拍摄的持续时间为dj+1且从时间tj+1开始的j+1取决于仅在上一次拍摄sj和段中的动作a[tj+1，tj+1+dj+1]。

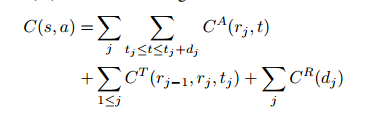
特别介绍编辑图形的概念，图形所有可能的镜头和过渡。在此图中，一个节点表示匆忙的时间步长（一帧），弧表示从匆忙（相机）的帧i到帧的过渡i+1匆忙（相同或不同）。系统的输出然后是场景的完整编辑列表，计算为通过编辑图表的连续路径，最大限度地减少错误这三个关键方面。

系统的第一个输入是一个有序世界事件的有序列表（称之为行动），用表示形式（主语，动词，宾语）。在下文中，作者认为所有动作的主语和宾语都是字符，将所有字符的集合称为C.作者使用四个主要类别的动作：由角色的嘴执行的说话动作，对所执行的动作做出反应通过角色的眼睛，操纵所执行的动作角色的手，以及由此执行的移动动作人物的脚。结果，角色可以成为主题在任何给定时间最多四个不同的行动和无限数量的行动的对象。

系统的第二个输入是来自的M列表不同的摄像机拍摄场景的总持续时间N个视频帧。虽然在本文中有许多相机由人类专家手动放置，他们同样可以由自动相机规划师计。

系统的输出是一部电影，描述为编辑决策列表（EDL）定义为一系列镜头三胞胎的形式（rj，tj，dj）。请注意，在本文中，只考虑保留时间的按时间顺序排列的EDL（tj+1=tj+dj）。在这种有限的背景下，EDL可以使用急速选择函数r（t）重构给出紧急指数作为时间的函数。

作者建议将编辑过程转换为符合三个标准的数学模型：（i）多少镜头传达展开的动作，（ii）连续性编辑原则和（iii）电影的切割节奏。为此，作者使用对数线性模型，其中选择的概率特定的镜头序列sj被认为是线性代价函数C（s，a）的指数。成本函数C（s，a）进一步分解为三个项，分别测量（i）传递展开动作的误差每次拍摄，（ii）违反连续性编辑规则切割和（iii）选择射击持续时间的错误。

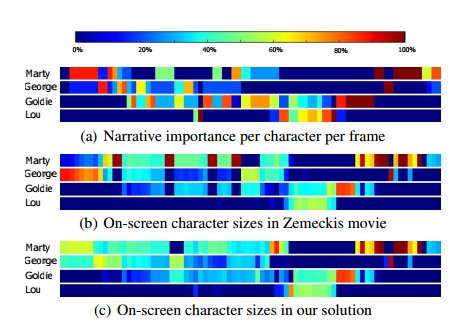


在该等式中，第一项是所有帧的总和与行动相关的成本函数CA的所有镜头现场。第二个术语是所有削减的总和与镜头之间的过渡相关的成本函数CT。这两个词进一步分解为加权特征之和。第三个词是a将与编辑相关的成本函数CR的所有镜头相加韵律。

**4 象征性映射**

在拍摄虚拟场景期间，所有摄像机都会捕获图像是世界场景的透视投影进入自己的参考框架。与图形管道执行的计算并行，作者执行a保持记录的行为的象征性投射每一次冲刺中都可以看到多少动作帧。

执行过程如下。首先，每个动作都被分解为其成分-动词，主语和宾语。基于在动词类别（说话，反应，移动或操纵）上，然后计算所涉及的边界框受试者的身体部位（例如嘴，眼睛，手和脚）和对象字符，然后计算屏幕大小他们在每一帧的投射中投射。其次，为了评估这些行为中有多少是可见的，计算了可见和遮挡的字符区域。为此，每个人都这样做在角色的每个身体部位b的面f，在冲刺r中的时间t计算其投影尺寸（或面积）S（f，r，t）。该投影尺寸是相对于屏幕尺寸测量的，并且包括f的屏幕上和屏幕外投影。然后按如下方式定义f的可见和遮挡大小。其遮挡大小O（f，r，t）对应于累积其区域大小被遮挡或出现在屏幕外，而其可见尺寸V（f，r，t）是互补值计算使得S（f，r，t）=V（f，r，t）+O（f，r，t）。最后定义每个的投影尺寸和可见尺寸字符c作为每个面上相应值的总和其身体部位。这种方法更容易扩展到的情况非角色对象（使用它们的边界框）和多个字符。



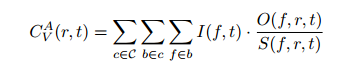
5 叙述的重要性

对于每个动作持有时间t，叙述重要性I（a）取决于a是否是前景或后台行动。然后，首先根据行动中的不同角色分配重要性（主题和对象）。请注意，这定义了叙述的重要性在时间t，每个字符I（c，t）。其次分发每个角色的重要性按比例分配到充当角色的角色的个体身体部位。例如，在说话动作中主体和物体的视觉目标是他们的头;在操纵动作中的主体视觉目标是它的头部和胸部;在移动动作中，主体的视觉目标是其全身。最后不均匀地分散身体部位在其正面，背面的重要性和侧面以获得每张脸的叙事重要性I（f，t）是时刻t的一个角色。

**6 镜头选择**

基于与给定帧的展开动作相关的符号和几何数据，作者对每个帧的每个帧进行排序拍摄三个方面：动作可见性，动作接近度和动作排序。

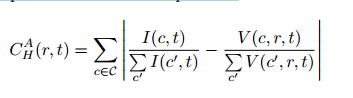
操作可见性术语评估有多少展开操作可见。为了完全满足这个标准，角色的每个重要身体部分都参与展开动作应该在屏幕上并且完全可见。与动作可见性相关的成本计算为每个面上的总和封闭的每个角色c的每个身体部位b的f由叙事重要性加权的面部比例：



动作接近度术语评估了如何沉浸其中相机处于展开状态，即屏幕多少由行动填补。与穷人相关的成本（或罚款）然后通过屏幕的比例给出动作接近度由人物填充：



动作排序术语评估角色的屏幕重要性与其叙述重要性的匹配程度。这也被称为希区柯克原理，这表明角色的大小应该与故事中的叙事重要性成正比。作者的实现考虑了场景中存在的所有角色，而不仅仅是每个镜头中出现的角色，或参与主要动作的角色。这有利于轻松排除不重要角色的突出镜头，并支持重要角色的突出镜头，专注于它重要的身体部位（说话时嘴巴，反应时眼睛，操纵时双手，移动时双脚）。该与希区柯克原理相关的成本计算为与其叙述重要性相比，角色的屏幕可见性的所有偏差的总和：



**7 连续性编辑**

作者根据所有可能的转换排序连续性编辑的计算模型。连续性是电影制作中最常用的编辑风格。它依赖于一套完善的规则：避免跳跃削减，强制执行屏幕，动作和凝视连续性保持屏幕角色的从左到右排序。

**跳跃切割。**当连续两次拍摄中出现相同的角色时（即切割前后），其表观尺寸或其轮廓角必须有足够的变化在两个镜头中，让观众感知切割为观点的变化，而不是突然的变化角色的姿势，也称为跳跃。为了防止这样的削减，作者通过成本函数惩罚这些错误所有出现在连续镜头中的角色，其公式详见章节跳转部分。

**屏幕，动作和凝视连续性。**连续性编辑的主要目标是为了强化切割中所有角色的屏幕位置、运动方向和注视方向的连续性。为了防止不连续，作者会通过非线性组合的差异屏幕位置、注视方向和运动方向连续镜头中出现的所有角色惩罚他们。这些成本函数的公式详见附录的截面屏幕连续性、运动连续性和凝视连续性。

**从左到右排序**。字符的从左到右排序是强制视觉连续性的另一个重要因素。相对屏幕位置反转的字符切割之后似乎是跳来跳去，这引起了对切割的注意。这个标准也是众所周知的作为180度规则。对于每对字符（c，c`）如果它们在屏幕上的位置差ΔPx（c，c0）（在x轴上取得）的方向差异是相反的符号，那么会在切割之前和之后出现在屏幕上，然后惩罚回复（成本=1）。那么从左到右连续性的总成本就是计算为所有字符对的此惩罚的总和切割前后出现在屏幕上。

**8 限制和未来工作**

模型目前仅限于线性编辑的情况，其中保持事件的时间顺序。在未来工作，作者希望通过允许删除此限制时间椭圆和事件的重新排序。另一个限制是仅限于有限的相机选择。未来的工作需要优化相机位置和构架。此外，所提出的模型只能实现控制起搏。其他风格参数，如镜头构图（平面与深度舞台），相机移动（静态）镜头，小车拍摄，起重机拍摄），过渡（反向拍摄）和照明偏好有利于用户的创造力。镜头选择的模型是基于的边界框角色的身体部位和他们的原始分类动作。场景中的物体和场所，以及角色的情绪和意图，也应该发挥作用。最后，应该注意到半马尔可夫假设有其自身的局限性。重要的电影编辑模式如书籍结束，分离或平行动作，不能被这种无记忆模型考虑在内。该调查高阶马尔可夫模型或无上下文将继续使用语法来克服这些限制。

**9 结论**

在本文中提出了一种连续性编辑方法，用于自动创建3D动画的电影序列。文中已经介绍了这个概念编辑图形并显示动态编程如何用于在半马尔可夫假设下计算最优编辑。作者提供了手段的详尽描述对镜头和编辑进行排名，并测量距离指定的切割节奏。本文的解决方案得到了在感知用户研究中获得的主观评估的支持。该所提出的方法从现有技术（例如基于成语的表示）中明确转变，其水平为以前的贡献没有解决的表现力。最后，这项工作为解决自动化电影摄影中的新挑战提供了基础，例如学习从真实电影中复制电影风格。

**致谢**

本文提出的研究得益于许多人的帮助。 Laura Paiardini和Estelle Charleroy为示例序列创建了资源和动画。 Adela Barbulescu帮助创建了面部动画。 Anne-Helene Olivier和Michael Gelicher在实验设计和用户研究评估方面给予了宝贵的帮助。 还要感谢参与用户研究的志愿者。这项工作的初步版本由第二作者开发，作为Xtranormal Technologies获得专利的Text-to-Movie应用程序的一部分。这项工作由法国研究机构资助 （ANR）通过项目EXPRESSIVE通过项目CHROME和CINECITTA以及欧洲研究理事会（ERC）。

参考文献

[1]Markowitz, D.; Jr., J. T. K.; Shoulson, A.; and Badler, N. I.2011. Intelligent camera control using behavior trees. In Motion in Games (MIG), 156–167.

[2]Mitchell, C.; Harper, M.; and Jamieson, L. 1995. On the complexity of explicit duration hmm’s. IEEE Transactions on Speech and Audio Processing 3(3):213–217.

[3]Murch, W. 1986. In the blink of an eye. Silman-James Press.

[4]Murphy, K. P. 2002. Hidden semi-markov models. Technical report, MIT AI Lab.

[5]Salt, B. 2009. Film Style and Technology: History and Analysis (3 ed.). Starword.

[6]Sarawagi, S., and Cohen, W. W. 2004. Semi-markov conditional random fields for information extraction. In Advances in Neural Information Processing Systems.

[7]Sharff, S. 1982. The elements of cinema. Towards a theory of cinesthetic impact. Columbia University Press.

[8]Shi, Q.; Wang, L.; Cheng, L.; and Smola, A. 2008. Discriminative human action segmentation and recognition using. semi-markov model. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[9]Smith, T. J. 2005. An Attentional Theory of Continuity Editing. Ph.D. Dissertation, University of Edinburgh.

[10]Thompson, R., and Bowen, C. 2009. Grammar of the Edit. Focal Press.

[11]Tomlinson, B.; Blumberg, B.; and Nain, D. 2000. Expressive autonomous cinematography for interactive virtual environments. In International Conference on Autonomous Agents, AGENTS ’00, 317–324. ACM.

[12]Truffaut, F., and Scott, H. G. 1967. Truffaut/Hitchcock. Simon & Schuster.Yu, S.-Z. 2010. Hidden semi-markov models. Artificial Intelligence 174(2):215 – 243.  
Zen, H.; Tokuda, K.; Masuko, T.; Kobayasih, T.; and Kitamura, T. 2007. A hidden semi-markov model-based speech synthesis system. IEICE - Trans. Inf. Syst. E90-D(5):825–  
834.