

硕 士 研 究 生 读 书 报 告



题目 MoSculp：形状和时间的交互式可视化

作者姓名 王有伟

作者学号 21851473

指导教师 陈实

学科专业 软件工程

所在学院 软件学院

提交日期 2018 年 12月

MoSculp: Interactive Visualization of Shape and Time

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: Shi Chen

By

Youwei Wang

Zhejiang University, P.R. China

2018

摘要

本文章提出了一个可视化的通过3D动作雕塑表现2D动作的系统，系统从输入视频计算运动方式行为，然后将其以3D的形式嵌入到场景中，或者将其直接3D打印出来，为了实现这一目标，系统引入了一种算法，估计由2D图像随时间变化的3D几何形状，并开发了基于3D图像渲染的将3D雕塑插回原视频的方法。通过传达3D信息，用户可以获取到肉眼难以观察的运动状态，文章中也验证了此方法的有效性，并进行了用户研究，发现此方法比现有的频闪视频法和时空可视化方法可以表达更多的信息。最后文章也对用于人的此方法进行了拓展研究，将其用于其他事物，也很有效果。

**关键词**：运动估计和可视化，基于图像的渲染

Abstract

This paper presents a visual system for representing 2D motion through 3D motion sculpture. The system calculates the motion behavior from the input video and then embeds it into the scene in 3D, or directly prints it in 3D.in order to achieve this, the system introduces an algorithm that estimates the 3D geometry of a 2D image over time and develops a method for inserting 3D sculptures back into the original video based on 3D image rendering. By conveying 3D information, the user can obtain the state of motion that is difficult for the naked eye to observe. The paper also verifies the effectiveness of this method, and conducts user research, and finds that this method can express more information than the existing stroboscopic video method and space-time visualization method. In the end the article also expands on this method for humans, it is also very effective for other things.

**Keywords：**Motion estimation and visualization, image-based rendering

1简介

复杂的动作，如摆动网球拍或跳舞芭蕾舞，很难通过静态照片传达给观众。 为了解决这个问题，研究人员和艺术家开发了许多运动可视化技术，如计时摄影，频闪摄影和多重曝光摄影但是，因为这样方法完全在2D中运行，它们无法传达运动的底层3D结构。因此，他们在对象的某些部分被遮挡时，会产生杂乱的结果而且它们通常需要特殊的捕获程序，环境，或照明设备。在此文章中，介绍了系统MoSculp，它将视频作为输入并生成动画雕塑：一个可视化的穿过空间的雕塑。 运动雕塑有助于可视化人体轨迹，并揭示其3D形状随着时间的推移而演变的动作，一旦系统计算出来这个雕塑，便可以将其插回到源视频，并渲染在合成场景中，或物理3D打印。

该项目开发了一个交互式界面，允许用户：（i）在3D模式中浏览运动雕塑，即在它们周围导航并从其他视角来看待它们，从而揭示出来一些无法访问到的信息，以及（ii）自定义各种渲染设置，包括灯光，雕塑材料，渲染的身体部分，场景背景等。 这些工具使得用户可以灵活的表达他们的艺术设计，进一步促进他们对人体形状和运动的理解。

该项目设计了一个便于新人用户使用的系统，其核心部分是估计人体姿态和姿态随时间变化的方法，该估算算法是基于现有技术的基础上，旨在恢复所需的3D信息来建立运动雕塑模型，同时支持用户的改错。要将雕塑插回原始视频，开发了一种保留深度排序的基于图像的3D感知渲染方法。此系统实现了高品质，适用于各种人类行为的无伪影复合材料的运动动作，如芭蕾舞，击剑和其他。

文章也简要回顾了艺术渲染领域的相关工作，比如自动化艺术渲染，图像中的运动效果，人体姿势估计，视频编辑和摘要方法，以及物理可视化。

**2 系统过程及演示**

为了生成运动雕塑，用户首先加载一个视频进入系统，之后MoSculp系统会检测2D关键点并将它们叠加在输入帧上。然后用户浏览检测结果并确认开启几个（3-4个）随机选择的帧，用户并且通过单击“左/右更正”按钮来更正识别的关键点。 添加标记后，用户点击“Done Annotating”，触发MoSculp纠正时间上不一致的检测，并将这些标记的框架作为锚点。 然后MoSculp在离线过程中生成运动雕塑，包括估计所有框架中的人体形状和姿势渲染雕塑。

处理后，生成的雕塑被加载到MoSculp，用户可以在3D中虚拟探索它。 这通常会揭示从原始相机角度看不到的有关形状和运动的信息，以及了解不同的身体部位的互动。最后，渲染的动画雕塑以新的窗口显示，用户可以在其中自定义控制渲染设置，比如场景，光照，渲染的身体部位，材质，透明度以及人物，通过这些渲染属性，用户可以自定义可视化并且选择最能够表现时空信息的设置。

接下来文章演示了一些实际的系统使用效果，该项目从大部分网络视频网站上收集了大多数视频，也自己用佳能相机拍摄了两个视频，一个是跳跃视频，另一个是U型走路视频。对于每一个演示例子，文章中都将运动雕塑嵌入到原视频和合成背景中。从论文中的图片展示来看，该系统在实现效果方面表现良好，实现了从2D到3D的转化，增添了动作细节，实现了方便人对动作进行细微观察的需求。

**3 系统实现**

实现图像的转换及渲染是此系统的核心，MoSculp背后的算法包括几个步骤（a）首先在每一帧中检测人体及其2D姿势（由一组关键点表示），（b）恢复表现人形态的3D身体模型（c）从3D模型中提取3D骨架将其扫描到3D空间并创造一个初始运动雕塑，最后，（d-f）与人类一起渲染不同风格的雕塑，同时保持深度排序。

（a）2D关键点检测

使用OpenPose估算由一组2D关键点表示的每帧中的2D身体姿势。 每个关键点都是与关节标签和它在框架中的2D位置相连接的。虽然在单个图像中检测到的关键点通常是准确的，人体运动中固有的模糊性有时导致时间不一致，例如，左和右肩在相邻的框架之间的翻转。 通过在检测之间强加时间一致性来解决在相邻的帧中的这个问题。 具体的说，在每帧的检测中使用了隐马尔可夫模型（HMM）。在每个关节的特定时间戳的位置计算最大边际可能性估计，同时强加时间平滑。用户也将在系统界面上标记检测到的关键点是否正确，最后将标记正确的帧做HMM的约束，通常3或4个标签足以纠正100帧视频的所有错误。

（b）基于时间的从2D关键点的到3D模型的转化

有了检测到的2D关键点，这一步的目标是匹配每个框架中身体的3D模型。该项目选择最小化重新投影误差，即每个2D的距离关键点和网格顶点的3D到2D投影对应于相同的身体部位。使用SMPL身体模型网格，其包含一组控制体形，姿势的参数，和位置。 具体而言，移动体由下式表示形状参数b，每帧姿态qt和全局转换参数Tt呈现。 我们通过估计N帧中每一帧的这些参数最小化以下目标函数：

（c）生成雕塑

有了一系列3D体形，通过提取从每帧3D模型中重建的骨架并连接贯穿于所有帧的这些骨架，从而创建了人的时空扫描。 这个时空扫描形成了我们初始的动作雕塑。

（d）精制和渲染运动雕塑

虽然生成了初始的雕塑，但是为了实现无伪影和生动的渲染，还有一些问题需要解决，首先便是细节问题，从3D身体模型无法捕获重要的细节，如面部结构，头发，衣服等等，其次，遮挡下的纹理实现也是问题，对此的解决办法是将3D运动模型插回到原始的2D视频中。

（e）3D雕塑和2D视频的深度感知复合

天然的叠加到渲染视频上的3D雕塑会导致混乱的可视化，雕塑和物体之间完全无视3D空间关系，为了解决这个问题，此项目在每个视频帧中估计了人的深度图，来确定人是离相机更近还是更远。

（f）深度和雕塑的细化

虽然上一步的深度图会自动与雕塑进行关联，但是这个深度图却并不是完美对齐人的轮廓，因此还需要推断出人体在每个帧中的深度图来进行深度排序。使用初始深度贴图生成的渲染会导致视觉伪像，为了消除这种伪影，提取了跨越所有帧的前景蒙版，并细化人的初始深度。为了改善对象的深度，计算密集匹配，然后通过光流翘曲传播初始深度值到前景掩膜。如果一个像素翘曲后没有深度，则复制其最近的具有深度的邻居像素的深度。

**4 用户研究及技术评估**

项目还进行了几项用户研究以比较不同可视化中运动情况的感知，从而评估本项目提供的风格设置。结果显示该项目的可视化项目比替代品更受欢迎，在信息展示方面比造型时间摄影和频闪摄影做的更好。

在光照选择上也进行了用户调研，大多数用户选择本项目的反射与阴影的渲染和本地化照明。

在技术评估方面，主要评估了两个方面，一个是随时间推移的3D身体估计，另一个是基于流量的深度和雕塑细化。在随时间的身体估计方面，该项目考虑到多帧来建立形态，对比SMPLify的对每个帧独立的建立模型，具有更好的稳定性，因此效果更好。由于3D形状和姿势是使用低维基矢量编码的，因此其和2D的完美对其是无法实现的。但是，利用基于流量的细化方案可以减少这种伪影，通过计算IoU来量化这种贡献，经过分析后，发现IoU平均值从0.61增加到了0.94，效果得到了显著改善。

在人体模型的基础上，项目又对相机的运动和非人运动雕塑进行了动作雕塑构建研究。

参考文献

[1] Xiuming Zhang, Tali Dekel, Tianfan Xue et al. 2018. MoSculp Interactive Visualization of Shape and Time. In ACM Symposium on User Interface Software and Technology (UIST) 2018.