
上海大学 2023 ~ 2024 学年 秋 季学期研究生课程考试

课程报告

课程名称: 图像处理与理解 课程编号: 3XS081041

论文题目: 流体模拟与视频生成方法研究

研究生姓名: 郎志远 学 号: 23724590

论文评语:

成 绩: _____ 任课教师: 李晓强

评阅日期: _____

流体模拟与视频生成方法研究

郎志远

摘要 本文综述了最近几年在流体模拟和图像生成领域的部分论文，涵盖了不同的生成模型和应用领域。计算机流体模拟和渲染被广泛应用于影视特效、物理仿真、游戏制作等各个领域，且以离线计算为主。但随着图形卡对并行计算的支持提升，离线的流体模拟和渲染方法经过优化改进，虽然精度上具有小范围误差，效率上却能够满足实时计算的要求，可用于小规模流体模拟和渲染，因此在实时效率下提升流体模拟和渲染的真实感至关重要。在图像生成的基础上，视频生成是计算机视觉和人工智能领域的一个重要研究方向，旨在通过机器学习和深度学习技术生成逼真的视频。近年来，生成对抗网络和变分自编码器等生成模型的发展为视频生成任务带来了显著的进展。

关键词 流体模拟 视频生成 图像生成 生成对抗网络 卷积神经网络

Research on Fluid Simulation and Video Generation Methods

Zhiyuan Lang

Abstract This paper reviews a selection of papers from the last few years in the field of fluid simulation and image generation, covering different generative models and application areas. Computer fluid simulation and rendering are widely used in various fields such as film and television special effects, physics simulation, and game production, and are mainly computed offline. However, with the improvement of graphics card support for parallel computing, offline fluid simulation and rendering methods have been optimized and improved to meet the requirements of real-time computation for small-scale fluid simulation and rendering in terms of efficiency, although the accuracy has a small range of error, so it is crucial to improve the realism of fluid simulation and rendering in real-time efficiency. On the basis of image generation, video generation is an important research direction in the field of computer vision and artificial intelligence, aiming to generate realistic videos through machine learning and deep learning techniques. In recent years, the development of generative models such as generative adversarial networks and variational self-encoders has brought significant advances to the video generation task.

Key words Fluid Simulation; Video Generation; Image Generation; Generative Adversarial Networks; Convolutional Neural Networks

1 引言

流体模拟和渲染，是在计算机中还原真实世界中物理现象的工作之一。给定一个包含流体区域(如溪流、瀑布和海洋)的场景图像，人类可以想象静止图像中的液体是如何移动的。近年来，单一图像的流体动画问题逐渐成为一个热门话题。为了模拟流体，首先需要基于流体力学计算流体的运动规律；然后结合计算机图形学知识，使用基于物理的渲染从能量守恒的角度模拟光线的传播，合理绘制物体表面的颜色，提升渲染画面的真实感，最终得到模拟和渲染后的结果。渲染管线大致分为三个阶段：

应用阶段、几何阶段、光栅化阶段。本文对模拟和渲染的研究集中在应用阶段和光栅化阶段，基于虚幻引擎实现。它们的一般框架可以概括为两部分：(1)预测/计算表明液体运动的光流(2)基于运动预测/计算合成未来帧图像。尽管这些方法在简单的流体区域上获得了令人印象深刻的视觉效果，但如何处理透明度、碰撞和真实场景中的薄结构的复杂内容仍然是一个具有挑战性和开放的问题。

在图形学领域关于流体模拟的研究中，Muller M 等人在实时物理笔记中总结了经典的物理模拟和渲染实现方法。基于网格的流体模拟(Grid Based Fluid Simulation)章节中，介绍了求解纳维-斯托克斯方程(Navier-Stokes Equations)的方法，用于解

算流体的受力问题。之前工作的局限性的关键在于视频的生成过程是以全局的方式建模的。具体来说,后续的帧是基于对输入图像或神经特征的现有纹理进行变形而生成的。这种操作将场景视为一个整体,将场景中的所有内容扭曲在一起。另一方面,对单个流体图像进行动画化的另一个主要困难是预测流体的运动流动。存在两种类型的管道。第一类是交互式运动预测方法。它们需要稀疏的提示(例如,标记速度方向和相对振幅)在液体区域的随机像素作为附加输入。然后使用特征聚类模型或卷积网络形成密集运动场。由于精度取决于用户可以提供的运动细节水平,这种方法在复杂的流体场景中是耗时的。第二类是自动方法。这些工作将该任务视为一个域变换问题,并使用图像到图像的平移技术来预测由通道光流表示的运动。速度不确定(如平坦的河面)可以向左流动,也可以向右流动,在一段时间内由于与物体碰撞而引起的速度的巨大变化。

2 相关方法

2.1 视频生成

在早期的研究中,研究人员主要通过改变播放录制帧的属性来生成新的视频上下文[1]。传统的可移动物体视频运动生成方法是基于随机过程分析和颜色转移。给定一个场景的时间变化的参考视频,可以使用补丁对齐方法和基于能量的优化[2]生成纹理运动和序列图像。此外,将类似的方法应用于单图像输入的[3]。但由于可以检测到明显的伪影,且不能保证视频的流畅性,使得此类方法的使用局限于典型的场景样本。近年来,深度神经网络在图像生成领域[4]的发展,为提高图像的合成质量做出了大量的创新努力。生成对抗训练、感知丧失、部分卷积和循环一致性等技术已被广泛用作标准组件。

大量的工作集中在无监督和随机方式下的视频合成[5]。有大量的研究涉及从中间表示(如语义标签映射)生成视频。与这项工作相关的是单图像到视频合成的工作[6]。庄永裕等人通过允许用户将图像分解成几个层,每个层都需要以不同的方式进行动画处理,从而使图片动画化。熊伟等人提出了一种从单幅图像合成视频的两阶段方法,其中在第一阶段使用 3D-GAN 生成帧序列,在第二阶段使用另一 GAN 进一步细化帧序列。李翊君等人[7]首先从输入图像中预测未来帧的光流图序列,然后使用它

们来获得未来的 RGB 帧。Logacheva 等人提出了一种完全不同的方法,他们在 StyleGAN[8]潜在空间中对视频中的景观帧序列进行建模,同时加强时间一致性。与[14]类似,Holynski 等[9]首先估计了未来帧的光流,只不过该工作假设整个视频中有一个恒定的二维流图。Dorkenwald 等人提出使用残差表示在图像和视频之间进行一对一的映射,允许用户为视频生成提供单一的运动方向。[1]和[2]提出了一种方法,通过在运动的开始和结束位置定义的特定位置进行单个戳来控制图像中不同部分的动画。然

在图像生成成功的基础上,范思明、朴景潭等人开发了一些方法来迫使网络学习中间场景表示,以预测未来的帧或插值关键帧,一旦输入序列和相机轨迹给定。以单个图像作为输入,潜在生成方法如[10]提出了一个自回归模型,生成无限像素形成序列输出。InfinityGAN 通过结构合成器和纹理合成器来分解全局外观、局部结构和纹理,生成具有空间大小和细节级别的图像,这是以前无法达到的。[11]尝试用时变的傅里叶变换来表示流并生成运动视频。[12]努力挖掘图像的语义,并相应地生成视频,达到自然的效果。动画景观通过学习中间流场和颜色转换地图来解耦运动和外观。然而,这些方法要么关注空间上的一致性,要么关注粗级时差场景变化。没有一种方法能通过精细的运动来讨论或处理流体的特殊性。

2.2 流体模拟

2.2.1 物理流体模拟

给定初始运动,根据初始状态预测接下来的运动。传统方法起源于计算机图形学。多年来,为了实现这一目标已经做出了若干努力精确的运动近似各种类型的流体。拉格朗日-欧拉混合方法[13]成功地模拟了不可压缩流体在预先定义的 2D 或 3D 盒子环境中的运动,解决了压力在欧拉网格上的投影和在拉格朗日粒子上的平流。为了提高算法在复杂形状上的收敛能力和灵活性,材料点法(MPM)得到了广泛的应用,后来又提出了更多的方法来稳定迭代速度时的收敛。然而,在大型场景中,这些方法非常耗时。为了降低模拟大范围流体的成本,人们提出了只模拟小范围流[14]或流体表面[]的方法,通过避免 3D 体积求解器的高计算成本,从而产生更快的近似。虽然这些作品取得了惊人的模拟效果,但由于场景特定的建模,它们仍然局限于合成场景。

2.2.2 优先使用数据对流体进行动画处理

给定一个静态图像作为动画目标，一组真实世界的水视频候选集作为动画库，[15]建议利用视频检索从库中找到一个/一组与目标图像内容相似的视频作为参考。然后将参考点的运动和外观信息传递到目标图像上的每个感兴趣区域，形成水动画视频。这些方法打开了现实世界的流体动画与数据之前在野外。然而，动画结果与源输入图像的无缝对齐并不能得到很好的保证，需要大量的人工操作。随着深度学习的发展，最近的著作提出了利用真实世界的视频进行训练，通过从重构损失中学习自动预测每帧运动和外观。首先通过运动欧拉积分将运动估计部分简化为单帧运动预测，然后提出深度特征翘曲技术来缩小翘曲引起的空白区域的大小。最后，通过可学习参数 Z 自适应融合前向特征和后向特征，生成动画流视频。基于提出将流体运动回归到用户的稀疏引导条件下，通过运动速度聚类生成成对的训练数据。他们进一步提出使用多尺度表示来捕捉不同分辨率下的不同流体速度。尽管这些方法在真实世界的场景中取得了令人印象深刻的效果，但由于对整个场景的单一表示将不同的对象和纹理视为平等的，因此它们很难处理流体上复杂的上下文关系。与之前的工作不同，我们的系统结合了物理模拟和基于学习的管道的优点。对于纹理，为了避免流体和环境之间的相互作用，强制网络将场景解耦到一个新的双层表示单反中，其中包括静态背景和随时间变化的表面流体层。同时，由于之前的大尺度数据，网络可以在空出的区域产生合理的、真实感的纹理。对于动作，在保持物理合理性的前提下，范思明、朴景潭等人提出了一种三步运动预测方法。首先使用人类界面的稀疏标签来生成初始粗糙运动，然后通过引入 2.5D 的表面模拟 SFS 来计算合理的演化。之后，一个平滑的运动翻译器被用来细化运动趋势，以更好地匹配真实世界的例子与数据之前。通过大量实例的网络训练，可以生成自然的流体动画，灵活地编辑主要的流体区域和边界条件，创建各种视觉效果。

2.3 运动预测

初始速度场通常被描述为两幅序列图像的光流，由于速度模糊，很难从一幅图像中预测出初始速度场。从图像颜色相对于坐标的连续建模出发，利用颜色的估计差值来预测运动。由于卷积神经网络被广泛应用于各种任务中，基于深度学习的方法 [16] 可以在复杂图像上获得更稳定的结果。最近的

研究转向把这个任务看作是一个图像到图像的翻译任务。条件生成对抗网络和变分自动编码器以及类似的方法已经被提出来完成图像到光流映射的传输任务，这些方法尝试将过程建模为基于参数的概率匹配，并生成类似于目标流的概率分布。给定图像输入为条件。一些自监督方法在遮挡边界和语义蒙板的帮助下，自动学习给定视频的流，以解决复杂视频中对象边界的离散颜色变化。有些以单个图像为输入，使用 U-Net 形网络输出光流图，捕捉整个图像中不同尺度的运动。将动作的方向分为八个离散的方向，每个方向用一个绝对值表示快或慢，并预测概率，以简化训练。也有一些交互式的运动预测系统旨在在尽可能少的用户指导下生成更精确的运动。

3 总结与展望

本文综述了流体模拟和视频生成领域的研究进展，并探讨了它们在计算机视觉和人工智能领域的重要性和应用。流体模拟和渲染技术在影视特效、物理仿真和游戏制作等领域具有广泛的应用。传统的离线计算方法在实现高质量流体模拟和渲染方面取得了显著成果，但其计算效率相对较低。然而，随着图形卡对并行计算的支持不断提升，离线方法经过优化改进后已能够满足小规模流体模拟和渲染的实时计算需求。另一方面，视频生成作为计算机视觉和人工智能领域的研究方向，旨在通过机器学习和深度学习技术生成逼真的视频。近年来，生成对抗网络 (GAN) 和变分自编码器 (VAE) 等生成模型的发展为视频生成任务带来了显著的进展。这些生成模型能够学习并模拟真实视频的统计特征，从而生成具有逼真度和连续性的视频序列。然而，流体模拟和视频生成领域仍存在一些挑战。在流体模拟方面，处理透明度、碰撞和复杂内部结构等问题仍然是具有挑战性的任务。对于视频生成，需要进一步改进生成模型的质量和稳定性，以生成更逼真、流畅的视频序列。此外，流体模拟和视频生成的实时性和计算效率也是需要解决的问题。

未来的研究可以着重解决流体模拟和渲染领域的几个方面。首先，可以进一步提高流体模拟的精确度和真实感，尤其是在处理透明度、碰撞和复杂内部结构等方面。可以探索新的数值方法和算法，以提高模拟的效果和可视化的逼真度。其次，

可以研究实时流体模拟和渲染的方法，使其能够应用于更大规模和更复杂的流体场景。这需要在计算效率和模拟质量之间找到平衡，以满足实时计算的需求。对于视频生成领域，可以进一步改进生成模型的质量和稳定性。可以研究更加高级和复杂的生成模型，以提高生成视频的逼真度和流畅性。此外，可以将视频生成与其他领域的研究相结合，如语义标签映射和多层次处理，以进一步改善视频生成的效果和真实感。总体而言，流体模拟和视频生成领域有许多机遇和挑战等待着我们的研究。通过不断改进算法和技术，我们可以更好地应用流体模拟和视频生成技术到现实生活中。

参考文献

- [1] Erik Reinhard, Michael Adhikmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *Comput. Graph. Appl.*, 21:34–41, 2001.
- [2] Arno Schodl, Richard Szeliski, David H Salesin, and Irfan Essa. Video textures. In *SIGGRAPH*, 2000.
- [3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017.
- [4] Yung-Yu Chuang, Dan B Goldman, Ke Colin Zheng, Brian Curless, David H Salesin, and Richard Szeliski. Animating pictures with stochastic motion textures. *ACM TOG*, 24:853–860, 2005.
- [5] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *CVPR*, 2016.
- [6] Yuki Endo, Yoshihiro Kanamori, and Shigeru Kuriyama. Animating landscape: self-supervised learning of decoupled motion and appearance for single-image video synthesis. *ACM TOG*, 38:175:1–175:19, 2019. 2, 3, 5
- [7] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Flow-grounded spatial-temporal video prediction from still images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 600–615, 2018.
- [8] Yuanming Hu, Tzu-Mao Li, Luke Anderson, Jonathan Ragan-Kelley, and Fredo Durand. Taichi: a language for high-performance computation on spatially sparse data structures. *ACM TOG*, 38:201, 2019.
- [9] Aleksander Holynski, Brian L Curless, Steven M Seitz, and Richard Szeliski. Animating pictures with eulerian motion fields. In *CVPR*, 2021.
- [10] Chieh Hubert Lin, Hsin-Ying Lee, Yen-Chi Cheng, Sergey Tulyakov, and Ming-Hsuan Yang. Infinitygan: Towards infinite-resolution image synthesis. *arXiv preprint, arXiv2104.03963*, 2021.
- [11] Ekta Prashnani, Maneli Noorkami, Daniel Vaquero, and Pradeep Sen. A phase-based approach for animating images using video examples. In *Comput Graph Forum*, 2017.
- [12] Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays. Transient attributes for high-level understanding and editing of outdoor scenes. *ACM TOG*, 33:1–11, 2014.
- [13] Yongning Zhu and Robert Bridson. Animating sand as a fluid. *ACM TOG*, 24:965–972, 2005.
- [14] Vincenzo Citro, Paolo Luchini, Filippo Giannetti, and Franco Auteri. Efficient stabilization and acceleration of numerical simulation of fluid flows by residual recombination. *J. Comput. Phys.*, 344:234–246, 2017.
- [15] Makoto Okabe, Yoshinori Dobashi, and Ken Anjyo. Animating pictures of water scenes using video retrieval. *Vis. Comput.*, 34:347–358, 2018.
- [16] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In *CVPR*, 2017.