

Homework 1 - Learning about CG

15331416 数字媒体技术方向 赵寒旭

目录

1. 结合上述参考链接（或其他参考资料），谈谈自己对计算机图形学的理解。	2
1.1 计算机图形学的简单定义	2
1.2 计算机图形学的主要内容	2
1) 建模(<i>Modeling</i>)	2
2) 渲染(<i>Rendering</i>)	3
3) 动画(<i>Animation</i>)	3
4) 人机交互(<i>Human-Computer Interaction, HCI</i>)	3
1.3 与计算机图形学相关的研究方向	4
2. 结合上述参考链接（或其他参考资料），回答什么是 OpenGL? OpenGL ES? Web GL? Vulkan? DirectX?	4
2.1 OpenGL	4
2.2 OpenGL ES	4
2.3 Web GL	5
2.4 Vulkan	5
2.5 DirectX	5
3. gl.h glu.h glew.h 的作用分别是什么?	5
4. 使用 GLFW 和 freeglut 的目的是什么?	6
5. 结合上述参考链接（或其他参考资料），选择一个 SIGGRAPH 2017 上你最喜欢的专题，介绍该专题是做什么的，使用了什么 CG 技术？（不少于 100 字）	6
1) 合成奥巴马	7
2) 音频记录的文本插入和替换	8

1. 结合上述参考链接（或其他参考资料），谈谈自己对计算机图形学的理解。

1.1 计算机图形学的简单定义

计算机图形学(Computer Graphics, 简称 CG) 是一种使用数学算法将二维或三维图形转化为计算机显示器的栅格形式的科学。

简单地说, 计算机图形学的主要研究内容就是研究如何在计算机中表示图形、以及利用计算机进行图形的计算、处理和显示的相关原理与算法。虽然通常认为 CG 是指三维图形的处理, 事实上也包括了二维图形及图像的处理。

狭义地理解, 计算机图形学是数字图象处理或计算机视觉的逆过程: 数字图象处理是把外界获得的图象用计算机进行处理的学科, 计算机视觉是根据获取的图像来理解和识别其中的物体的三维信息及其他信息, 计算机图形学是用计算机来画图像的学科。

在学科开创之初, 计算机图形学要解决的是如何在计算机中表示三维几何图形, 以及如何利用计算机进行图形的生成、处理和显示的相关原理与算法, 产生令人赏心悦目的真实感图像。随着近 40 年的发展, 计算机图形学的内容已经远远不止这些了。广义的计算机图形学的研究内容非常广泛, 如图形硬件、图形标准、图形交互技术、光栅图形生成算法、曲线曲面造型、实体造型、真实感图形计算与显示算法, 以及科学计算可视化、计算机动画、自然景物仿真、虚拟现实等。

1.2 计算机图形学的主要内容

计算机图形学主要包含四大部分的内容: 建模(Modeling)、渲染(Rendering)、动画(Animation)和人机交互(Human-computer Interaction, HCI)

1) 建模(Modeling)

要在计算机中表示一个三维物体, 首先要有它的几何模型表达。因此, 三维模型的建模是计算机图形学的基础, 是其他内容的前提。表达一个几何物体可以用数学上的样条函数或隐式函数来表达; 也可以是用光滑曲面上的采样点及其连接关系所表达的三角网格来表达(即连续曲面的分片线性逼近)。

表 1 三维建模的主要方法

方法	简要介绍
NURBS (非均匀有理 B-样条、Bezier 曲线曲面) 方法	计算机辅助设计(CAD)中的主流方法是采用 NURBS 方法 (已成为 CAD 工业领域的标准), 这也是计算机辅助几何设计(CAGD)所研究的主要内容。
细分曲面(Subdivision surface)造型方法	一种离散迭代的曲面构造方法, 由于其构造过程朴素简单以及实现容易, 是一个方兴未艾的研究热点。
利用软件的直接手工建模	这些软件需要建模人员有较强的专业知识, 而且需要一定时期的培训才能掌握, 建模效率低而学习门槛高, 不易于普及和让非专业用户使用。

基于笔划或草图交互方式的三维建模方法	符合人类原有日常生活中的思考习惯，交互方式直观简单，是最近几年研究的热点建模方法。
基于语法及规则的过程式建模方法	特别适合具有重复特征和结构化的几何物体与场景，比如建筑、树木等。
基于图像或视频的建模方法	是传统的计算机视觉所要解决的基本问题。该方法的问题是物体本身已经存在，而且重建的三维模型的精度有限。
基于扫描点云（深度图像如 Kinect、结构光扫描、激光扫描、LiDAR 扫描等）的建模 (Reconstruction)方法	单纯的重建方式存在精度低、稳定性差和运算量大等不足，远未能满足实际的需求。
基于现有模型来合成建模的方法	随着三维模型的逐渐增多，可以利用现有的三维模型通过简单的操作，比如 cut and paste，或者分析及变形等手段，来拼接或合成新的三维模型。

2) 渲染(Rendering)

渲染是为了将已有的三维模型或场景绘制出来并产生效果较好的真实感图像。现在的渲染技术已经能够将各种物体，包括皮肤、树木、花草、水、烟雾、毛发等渲染得非常逼真，然而已知的渲染实现方法，仍无法实现复杂的视觉特效，离实时的高真实感渲染还有很大差距。

3) 动画(Animation)

动画是采用连续播放静止图像的方法产生物体运动的效果。计算机动画借助于编程或动画制作软件生成一系列的景物画面，是计算机图形学的研究热点之一。研究方向包括：人体动画，关节动画，运动动画，脚本动画，具有人的意识的虚拟角色的动画系统等。另外，高度物理真实感的动态模拟，包括对各种形变、水、气、云、烟雾、燃烧、爆炸、撕裂、老化等物理现象的真实模拟，也是动画领域的主要问题。

4) 人机交互(Human-Computer Interaction, HCI)

人机交互是指人与计算机之间以一定的交互方式或交互界面，来完成确定任务的人与计算机之间的信息交换过程。简单来讲，就是人如何通过一定的交互方式告诉计算机来完成他所希望完成的任务。

在早期(上个世纪 60-70 年代)，只有以键盘输入的字符界面；到了 80 年代，以 WIMP(窗口、图符、菜单、鼠标)为基础的图形用户界面(GUI)逐渐成为当今计算机用户界面的主流。近年来，以用户为中心的系统设计思想，增进人机交互的自然性，提高人机交互的效率是用户界面的主要研究方向。陆续提出了多通道用户界面的思想，它包括语言、姿势输入、头部跟踪、视觉跟踪、立体显示、三维交互技术、感觉反馈及自然语言界面等。

1.3 与计算机图形学相关的研究方向

表 2 相关研究方向

研究方向	简要介绍
虚拟现实(Virtual Reality)	利用计算机图形产生器，位置跟踪器，多功能传感器和控制器等有效地模拟实际场景和情形，从而能够使观察者产生一种真实的身临其境的感觉。
可视化(Visualization)	利用计算机图形学和图像处理技术，将数据转换成图形或图像在屏幕上显示出来，并进行交互处理的理论、方法和技术。
可视媒体计算与处理 (Visual Media Processing)	当计算机图形学正向地图像生成方法和计算机视觉中逆向地从图像中恢复各种信息方法相结合，可以带来无可限量的想象空间，构造出很多视觉特效来，最终用于增强现实、数字地图、虚拟博物馆展示等多种应用中去。
医学图像处理 (Medical Imaging)	医学图像处理由生物医学成像(X 射线、CT、MRI)和生物医学图像处理两部分组成，在生命科学研究、医学诊断、临床治疗等方面起着重要的作用。医学图像分析中涉及的两个最为重要的内容为图像分割与图像配准。
计算机艺术 (Computational Arts)	计算机图形学的发展也提供给了艺术家发挥和实现想象的丰富的技术手段。

2. 结合上述参考链接（或其他参考资料），回答什么是 OpenGL? OpenGL ES?

Web GL? Vulkan? DirectX?

2.1 OpenGL

OpenGL 一般被认为是一个 API(Application Programming Interface, 应用程序编程接口), 包含了一系列可以操作图形、图像的函数。然而，OpenGL 本身并不是一个 API，它仅仅是一个由 Khronos 组织制定并维护的规范(Specification)。

OpenGL 规范严格规定了每个函数该如何执行，以及它们的输出值。至于内部具体每个函数是如何实现(Implement)的，将由 OpenGL 库的开发者自行决定（这里开发者是指编写 OpenGL 库的人）。因为 OpenGL 规范并没有规定实现的细节，具体的 OpenGL 库允许使用不同的实现，只要其功能和结果与规范相匹配（亦即，作为用户不会感受到功能上的差异）。

2.2 OpenGL ES

OpenGL ES (OpenGL for Embedded Systems) 是 OpenGL 三维图形 API 的子集，针对手机、PDA 和游戏主机等嵌入式设备而设计。该 API 由 Khronos 集团定义推广，Khronos 是一个图形软硬件行业协会，该协会主要关注图形和多媒体方面的开放标准。

OpenGL ES 是从 OpenGL 裁剪的定制而来的，去除了 glBegin/glEnd，四边形 (GL_QUADS)、多边形 (GL_POLYGONS) 等复杂图元等许多非绝对必要的特性。

OpenGL- ES 是免授权费的，跨平台的，功能完善的 2D 和 3D 图形应用程序接口 API，主要针对多种嵌入式系统专门设计——包括控制台、移动电话、手持设备、家电设备和汽车。它由精心定义的桌面 OpenGL 子集组成，创造了软件与图形加速间灵活强大的底层交互接口。OpenGL ES 包含浮点运算和定点运算系统描述以及 EGL 针对便携设备的本地视窗系统规范。OpenGL ES 1.X 面向功能固定的硬件所设计并提供加速支持、图形质量及性能标准。OpenGL ES 2.X 则提供包括遮盖器技术在内的全可编程 3D 图形算法。

2.3 Web GL

WebGL (Web Graphics Library) 是一种 3D 绘图协议，这种绘图技术标准允许把 JavaScript 和 OpenGL ES 2.0 结合在一起，通过增加 OpenGL ES 2.0 的一个 JavaScript 绑定，WebGL 可以为 HTML5 Canvas 提供硬件 3D 加速渲染，这样 Web 开发人员就可以借助系统显卡来在浏览器里更流畅地展示 3D 场景和模型了，还能创建复杂的导航和数据视觉化。显然，WebGL 技术标准免去了开发网页专用渲染插件的麻烦，可被用于创建具有复杂 3D 结构的网站页面，甚至可以用来设计 3D 网页游戏等等。

2.4 Vulkan

Vulkan 是一个低开销、跨平台的二维、三维图形与计算的应用程序接口，最早由 Khronos 组织在 2015 年游戏开发者大会（GDC）上发表。与 OpenGL 类似，Vulkan 针对全平台即时 3D 图形程序（如电子游戏和交互媒体）而设计，并提供高性能与更均衡的 CPU 与 GPU 占用，这也是 Direct3D 12 和 AMD 的 Mantle 的目标。与 Direct3D（12 版之前）和 OpenGL 的其他主要区别是，Vulkan 是一个底层 API，而且能执行并行任务。除此之外，Vulkan 还能更好地分配多个 CPU 核心的使用。

Khronos 最先把 Vulkan API 称为“次世代 OpenGL 行动”（next generation OpenGL initiative）或“glNext”，但在正式宣布 Vulkan 之后这些名字就没有再使用了。Vulkan 基于 Mantle 构建，AMD 将其 Mantle API 捐赠给 Khronos 组织，给予该组织开发底层 API 的基础，使其像 OpenGL 一样成为行业标准。

2.5 DirectX

DirectX (Direct eXtension, DX) 是由微软公司创建的一系列专为多媒体以及游戏开发的应用程序接口。旗下包含 Direct3D、Direct2D、DirectCompute 等等多个不同用途的子部分，因为这一系列 API 皆以 Direct 字样开头，所以 DirectX（只要把 X 字母替换为任何一个特定 API 的名字）就成为这一巨大的 API 系列的统称。目前最新版本为 DirectX 12，随附于 Windows 10 操作系统之上。

DirectX 被广泛用于 Microsoft Windows、Microsoft Xbox 电子游戏开发，并且只能支持这些平台。除了游戏开发之外，DirectX 亦被用于开发许多虚拟三维图形相关软件。Direct3D 是 DirectX 中最广为应用的子模块，所以有时候这两个名词可以互相代称。

DirectX 主要基于 C++ 编程语言实现，遵循 COM 架构。

3. gl.h glu.h glew.h 的作用分别是什么？

gl.h：按照约定，所有的 OpenGL 函数、类型和宏的原型都包含在头文件 gl.h 中。GL 是

OpenGL 的核心库，包含必需的 OpenGL 函数。

glu.h：OpenGL 工具函数库 (OpenGL Utility library)，主要定义 OpenGL 中的一些工具函数。GLU 是 OpenGL 的实用库，包含许多的新函数。

glut.h：OpenGL 实用工具库 (OpenGL Utility Toolkit)。作用是实现和窗口系统的交互。GLUT 表示 OpenGL 工具箱，体现了现代窗口系统所共有的功能函数库。GLUT 的目的就是隐藏平台的细节，glut.h 已经包含了 gl.h 和 glu.h。使用 GLUT 是因为 OpenGL 没有包含输入和窗口命令，而输入和窗口命令是由平台所决定的，与平台的相关性较大。

4. 使用 GLFW 和 freeglut 的目的是什么？

在我们画出出色的效果之前，首先要做的就是创建一个 OpenGL 上下文(Context)和一个用于显示的窗口。然而，这些操作在每个系统上都是不一样的，OpenGL 有目的地从这些操作抽象(Abtract)出去。这意味着我们不得不自己处理创建窗口，定义 OpenGL 上下文以及处理用户输入。

有一些库已经提供了我们所需的功能，其中一部分是特别针对 OpenGL 的。最流行的几个库有 GLUT，SDL，SFML 和 GLFW。

GLFW 是一个专门针对 OpenGL 的 C 语言库，它提供了一些渲染物体所需的最低限度的接口。它允许用户创建 OpenGL 上下文，定义窗口参数以及处理用户输入。

GLUT (OpenGL Utility Toolkit) 代表 OpenGL 应用工具包，是一个和窗口系统无关的软件包，可以跨平台使用，而 freeglut 是 GLUT 的一个完全开源替代库，对 GLUT 进行了扩展。

使用 GLFW 和 freeglut 的目的是节省了我们书写操作系统相关代码的时间，给我们提供用于渲染的窗口和上下文，帮助我们处理用户输入。

5. 结合上述参考链接（或其他参考资料），选择一个 SIGGRAPH 2017 上你最喜欢的专题，介绍该专题是做什么的，使用了什么 CG 技术？（不少于 100 字）

Speech and Facial Animation 语音和脸部动画

本专题涉及 4 篇论文，分别是：

- ① A Deep Learning Approach for Generalized Speech Animation
广义语音动画的深度学习方法。
- ② Audio-Driven Facial Animation by Joint End-to-End Learning of Pose and Emotion
联合端到端的姿势和情绪学习的音频驱动的脸部动画
- ③ Synthesizing Obama: Learning Lip Sync from Audio
合成奥巴马：从音频学习唇部同步
- ④ VoCo: Text-based Insertion and Replacement in Audio Narration
音频记录中的文本插入和替换

专题介绍了音频合成和脸部动画技术的相关进展，使用的 CG 技术包括：建模，渲染，动画制作。

下面选择两个较为有趣的篇目进行简单介绍：

1) 合成奥巴马

基于总统奥巴马的音频，这个团队合成了一个高质量的视频，视频中奥巴马的唇部动作与音频的内容有明显可见的准确同步。

通过对奥巴马每周数个小时录像的训练，循环神经网络习得了从原始音频特征到嘴型的映射。考虑到每个时刻的嘴型变化，他们合成了高质量的嘴巴纹理，并将其与合适的 3D 姿势匹配相结合，从而改变他在目标视频中的真实表达，让视频的内容与输入音轨相符合。

原始视频：



目标视频：



合成结果：



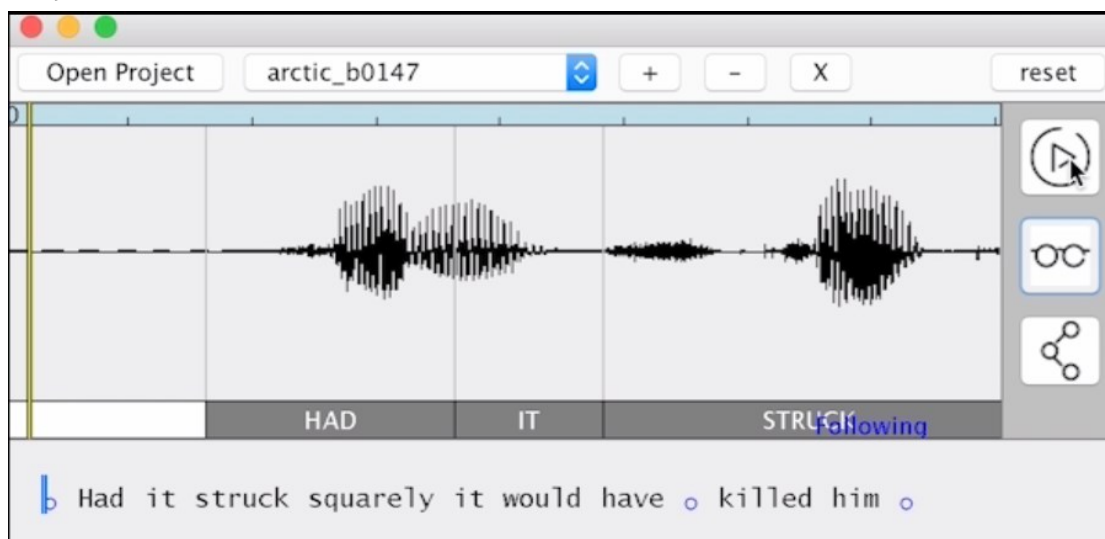
2) 音频记录的文本插入和替换

使用传统软件编辑音频叙述通常涉及许多艰苦的低级操作。一些最先进的系统允许编辑者直接在文本记录中执行选择，剪切，复制和粘贴操作，然后以直接方式自动应用于波形。然而，基于文本界面操作的明显缺陷是不能输入未出现在记录中的新单词，例如插入新单词或替换错误的单词。虽然现在已有高质量的语音合成器，但在以与语音叙述的其余部分相匹配的语音合成新词这方面仍然存在挑战。

这篇文章提出了一个可以合成一个新的词或短语并使其融合在现有叙述的环境中的系统。方法是使用文本到语音合成器以通用语音说出该词，然后使用语音转换将其转换为与叙述相匹配的语音。界面支持全自动合成，在备选发音集合中进行选择，对编辑位置和音高轮廓进行精确控制，甚至编辑自己的语音指导。文章提出的研究表明，此方法的输出优于基线方法，且通常与原始语音无法区分。

视频中截图如下：

通过修改文本，可以得到和原始发音意义截然相反的语音，但听上去仍像是由同一人说出的。



Open Project

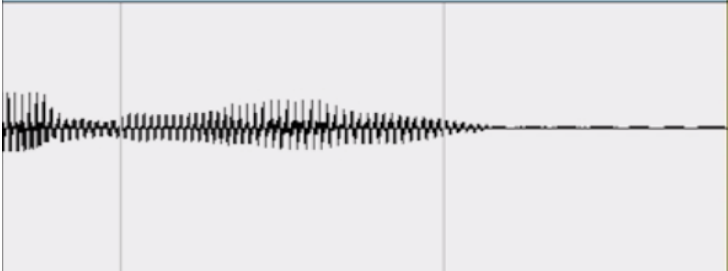
arctic_b0147

+

-

X

reset



HIM

Following

◦ Had it struck squarely it would have ◦ saved him ◦

