

129 | CVPR 2018论文精读：如何从整体上对人体进行三维建模？

2018-07-30 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 06:27 大小 2.96M



今天我们来分享 CVPR 大会的最佳学生论文，标题是《全方位捕捉：用于跟踪面部表情，手势和身体运动的 3D 变形模型》（[Total Capture: A 3D Deformation Model for Tracking Faces, Hands and Bodies](#)）。

很多学术会议都利用最佳学生论文这个奖项来鼓励学生参与学术研究活动，所以这个奖项的一般要求是第一作者必须是在校学生。

这篇论文的作者群来自卡内基梅隆大学。

第一作者周寒星（Hanbyul Joo）是来自韩国的学者，目前在卡内基梅隆大学机器人学院（The Robotics Institute）攻读博士。他的博士论文方向是“计算行为科

学”（Computational Behavioral Science）。他已经在计算机视觉方向发表了多篇 CVPR、ICCV 论文。

第二作者托马斯·西蒙（Tomas Simon）也是卡内基梅隆大学机器人学院的博士生。他的研究方向是“三维运动的时空建模”（Spatiotemporal Modeling of 3D Motion）。

最后一位作者是这两位学生的导师亚瑟尔·舍艾克（Yaser Sheikh），是机器人学院的教授。

论文的主要贡献

这篇论文想要解决的问题很直观，那就是希望对人体进行三维建模，并且能够跟踪（Track）人体的行为以及活动。

这个任务看似简单，但有不少难点。

首先，过去绝大多数的对人体进行三维建模的工作，都是针对人体的不同部分分别进行的，比如对脸部、对身体和对手部分别建模。在对这些部位进行建模的时候，整体的设定都不太一样。例如，对脸部的建模一般是近景（Close-Up），而对身体则主要是看身体的整体行动。也就是，对于人体不同部位的建模经常在不同的尺度下进行，那就无法把各个部分的建模很容易地对接上。

其次，还有一些人体的部位，过去并没有太多专门的建模工作，比如针对头发和脚，但这些在全身的建模中也是必不可少的部分。

这篇论文就加入了对头发和脚这些部分建模的讨论，提供了对人体从整体上进行建模的一个框架。确切地说，论文提供了两个模型：一个叫“弗兰肯斯坦”（Frankenstein），一个叫“亚当”（Adam）。

“弗兰肯斯坦”主要还是依靠对人体不同部分的建模，然后把这些模型连接起来，通过一些处理，能够让最终呈现的三维建模符合现实。在这个模型的基础上，“亚当”则加入了头发和脚的元素，抛弃了“弗兰肯斯坦”的一些特殊处理的地方，从模型的角度来说更加简单，并且达到了更加逼真的程度。

论文的核心方法

首先，我们来看一看这个叫“弗兰肯斯坦”的模型。这个模型的思路是尽量把能够对人体各个部分建模的最好的模型先拼接到一起。总体来说，每一个部分基本上都由三组参数组成：**运动参数**（Motion Parameters）、**形状参数**（Shape Parameters）和**全局翻译参数**（Global Translation Parameter）。

对于人的身体部分，作者们采用了**SMPL 模型**[1]。这个模型根据**人体形状的均值和形状的变化量**进行线性的叠加，然后经过一个**LBS 变换**来得到对身体部分的最终建模。

对人脸的部分，作者们采用了一个叫**FaceWarehouse 的模型**[2]。这个模型是根据**人脸形状的均值、形状的变化量，以及动态的变化量**来进行线性的叠加。

对于手而言，目前并没有模型可以直接用。作者们在这篇论文中**提出了自己的模型**，总的来说就是对手的骨架和关节进行建模，然后进行类似于身体的 LBS 变换。同样，也对人体的脚进行了类似的建模。

当我们有了身体、人脸、手和脚的模型以后，下面的工作就是把这些部分衔接上。首先，作者们保留了人体模型，移除这个模型上的人脸、手和脚。然后利用人脸模型、手的模型以及脚的模型相应的全局翻译参数，使得这些部分的坐标能够拼接上。最后，作者们还应用了一个“**融合函数**”来构建出一个平滑的人体结构。

“弗兰肯斯坦”的模型有一个**四项的目标优化函数**。这个函数的第一项是拟合“**关键点**”（Key Points），让人体的躯干骨架能够拟合上运动轨迹。第二项是拟合“**三维点云**”（3D Point Cloud），也就是让人体大部分躯体的实体能够拟合运动轨迹。第三项就是作者们附加的一个“**小技巧**”（Heuristic）用来把人体的每个部分连接在一起。这一项解决的就是整个模型设计带来的问题，也就是每个部分都是单独的形状参数，而并没有完全在模型上连接起来。最后一项是**高斯先验概率**，用来帮助模型找到唯一的解。

在“弗兰肯斯坦”的基础上，作者们开发了“**亚当**”模型。为了构建“**亚当**”，他们捕捉了70个人体的形态数据，首先构建这些人体的“**弗兰肯斯坦**”模型。在这些模型基础之上，作者们加入了人体的**头发和衣服的形态**，并且重新定义了整个模型的框架。

和“弗兰肯斯坦”相比，“**亚当**”是对所有的人体部件直接进行建模。这个模型和我们前面描述的某个具体部分的模型其实很类似，也是把人体的形态均值、形态的变化值和人脸表现值进行线性叠加，然后进行 LBS 变换。

因为“亚当”在模型上进行了简化和创新，所以在目标优化函数中只有三项变量。而我们刚刚讲过的用于“弗兰肯斯坦”模型的小技巧在“亚当”中就变得不需要了。

实验结果

在实验中，作者们使用了 140 个 VGA 照相机对三维身体的关键点进行重建，用了 480 个 VGA 照相机，对脚进行重建，31 个高清照相机用于脸部和手部关键点的重建以及三维点云的构建。

作者们显示了利用“弗兰肯斯坦”和“亚当”这两个模型对人体的三维运动进行建模。总体来说，这两个模型的表现非常相似。“亚当”因为有了头发和衣服的形态，在运动中显得更加真实。只是在有一些情况下，“亚当”构建的腿会显得有一些不协调的瘦，出现这种情况的原因，作者们归结于数据的缺失。

不过，从总体效果上来讲，这篇论文作为第一个对于人体的全身进行三维建模并动态跟踪的工作，应该算是达到了满意的结果。

小结

今天我为你讲了 CVPR 2018 的最佳学生论文。

一起来回顾下要点：第一，我们详细介绍了这篇文章要解决的问题，就是从整体上对人体的运动进行三维建模；第二，我们简要介绍了文章提出的两个模型，“弗兰肯斯坦”和“亚当”核心内容；第三，我们简单介绍了这篇论文所取得的不错的实验结果。

最后，给你留一个思考题，如果我们需要对“亚当”这个模型进行改进，你认为下一步应该做什么？

欢迎你给我留言，和我一起讨论。

参考文献

1. M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. **SMPL: A Skinned Multi-Person Linear Model**. In TOG, 2015.

2. C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou. **FaceWareHouse: A 3D Facial Expression Database for Visual Computing**. In TVCG, 2014.

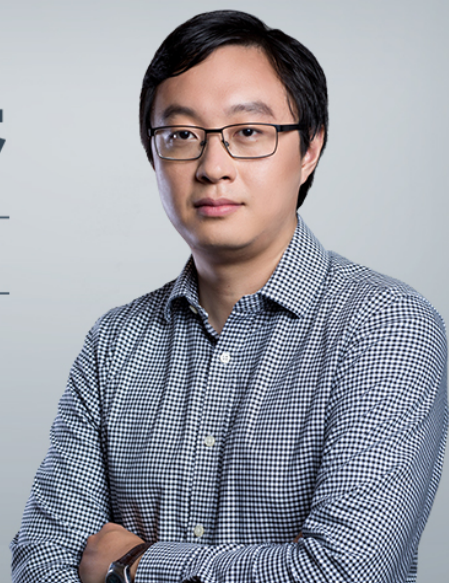


AI 技术内参

你的360度人工智能信息助理

洪亮劼

Etsy 数据科学主管
前雅虎研究院资深科学家



新版升级：点击「👤 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 128 | CVPR 2018论文精读：如何研究计算机视觉任务之间的关系？

下一篇 130 | CVPR 2018论文精读：如何解决排序学习计算复杂度高这个问题？

精选留言

💬 写留言

由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。