

1 文献基本信息

作者: Liuhao Ge^{1*}, Zhou Ren², Yuncheng Li³, Zehao Xue³, Yingying Wang³, Jianfei Cai¹, Junsong Yuan⁴; 文献题目: 基于单 RGB 图片的 3D 手形和姿势估计; 刊物: 计算机视觉与模式识别; 刊号: ISSN(1063-6919); 时间: 2019; 格式引文: Ge, Liuhao et al. “3D Hand Shape and Pose Estimation from a Single RGB Image.” CVPR (2019).

2 文献讨论的问题及其主要观点

写作目的 基于视觉的 3D 手分析是非常重要的题目，因为它在虚拟现实（VR）和增强现实（AR）中有许多应用。尽管有多年的研究，但由于手的形状，姿势，手势，咬合等的多样性和复杂性，这个题目仍然有待进一步研究。最近的一些研究从单眼 RGB 图像中进行 3D 手型分析，它们主要集中在估计稀疏 3D 手部关节位置，忽略了密集 3D 手型。许多身临其境的 VR 和 AR 应用通常需要对 3D 手的姿势和 3D 手的形状进行精确估计，这促使本文提出了更具挑战性的任务：如何联合估计 3D 手关节位置，而且还要联合估计 3D 手的完整 3D 网格从单个 RGB 图像获取手表面？

研究方法 基本思想是通过迭代优化将生成的 3D 手模型拟合到输入深度图像。考虑从单眼 RGB 图像估计 3D 手形，但 RGB 图像中缺乏明确的深度提示使得此任务难以通过迭代优化方法来解决。所以这项工作应用了经过端到端训练的深度神经网络，可以直接从单个 RGB 图像中恢复 3D 网格。具体地说，输入是一个以手为中心的单个 RGB 图像，通过两层式沙漏网络推断 2D 热图。通过使用包含八个残留层和四个最大池化层的残差网络，将估计的 2D 热图与图像特征图结合起来，编码为潜在特征向量。然后将编码后的潜在特征向量输入到图 CNN，以推断 3D 手网格中的 3D 坐标。通过使用简化的线性图 CNN，从重建的 3D 手网格顶点线性回归 3D 手关节位置。首先在异步数据集上训练网络模型，然后在真实数据集上对其进行微调。在包含 3D 手部网格的真实性和 3D 手部关节位置的综合数据集上，通过使用 2D 热图损失，3D 网格损失和 3D 姿态损失，以完全监督的方式对网络进行端到端训练。

主要结论 本文开发了基于 Graph CNN 的模型，从输入的 RGB 图像重建手表面的完整 3D 网格。为了训练模型，作者创建了大规模的 RGB 图像数据集，其中包含 3D 关节位置和 3D 手工网格的地面真相注释，并在此模型上以完全监督的方式训练模型。为了在没有 3D 地面真实性的现实数据集上微调模型，作者将生成的 3D 网格渲染到深度图，并利用观察到的深度图作为弱监督。在作者提出的新数据集和两个公共数据集进行的实验表明，该方法可以实时准确地恢复 3D 手部网格和 3D 关节位置。

3 文献阅读后的分析

以前的方法通过迭代优化将可变形手模型拟合到输入深度图，从深度图像估计 3D 手形和姿势。在深度学习继续发展之后，研究人员通过 CNN 估计深度图像中的姿势和形状参数，他们使用网格和姿势损失以端到端的方式训练 CNN。但这种方法恢复的手部网格的质量受到其简单的 LBS 模型的限制。随着单面 RGB 相机的广泛应用，许多最新的著作使用深度神经网络从单个 RGB 图像估计 3D 姿势。它们主要集中在估计稀疏 3D 手部关节位置，忽略了密集 3D 手型。本文基于单 RGB 图像，通过图卷积神经网络（Graph CNN）重建完整的 3D 手表面网格来估计完整的 3D 手形和姿势，且实现了卓越的 3D 手部姿势估计精度。