

3D-human-pose 项目中期报告

智能 16 唐超

1. 概述

人体姿态识别有很多应用：比如行为识别、人机交互、游戏、动画以及 3D 试衣等等,目前世界上多个实验室在进行相关的研究。在尝试对 Human Body Pose Estimation、Convolutional Pose Machines、openpose 三个人体姿态估计模型进行复现后，只有 openpose 模型成功且姿态估计效果最佳，计划后期主要以 openpose 模型为研究对象，主要学习 3D 的人体姿态重建、面部手部估计等功能的实现及原理、对于不同应用场景（如估计的重点部位不同）参数的调整。

2. 项目进展

从项目开题至今，所作的工作如下：

- (1) 前期为了实现 Human Body Pose Estimation（基于 tensorflow）、Convolutional Pose Machines（基于 pytorch），花了大量时间和精力实现搭建环境，包括建虚拟机、配置多种深度学习框架、配置 gpu 计算框架 cuda、cudnn 等等，但最后程序都没有运行成功。不过在这一过程中学习了 git 的使用和深度学习的基础知识，认识了深度学习中经常用到的框架等。
- (2) 尝试实现 CMU 认知计算实验室的 openpose 模型，最终在自己的电脑上成功实现图片 2D 人体姿态的识别及实时身体、脸部、手部的识别。
- (3) 对该模型 2D 人体姿态识别的算法原理有了浅显的认识。

识别结果如下：

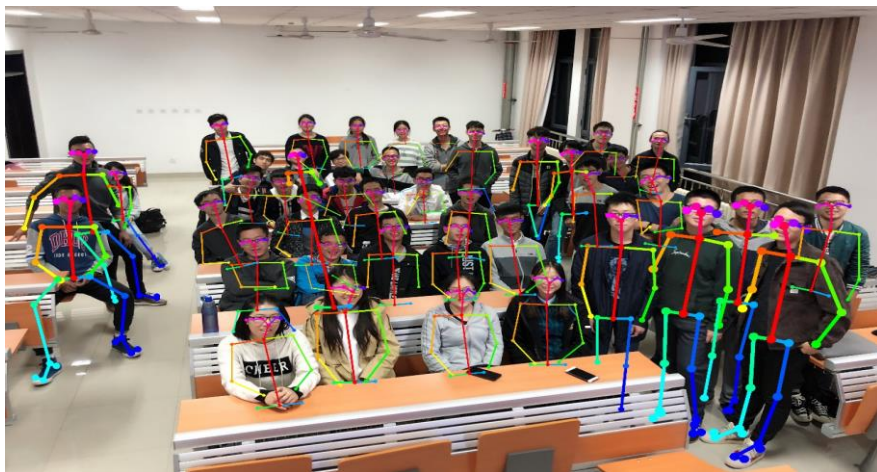


图 1 多人 2D 人体姿态识别

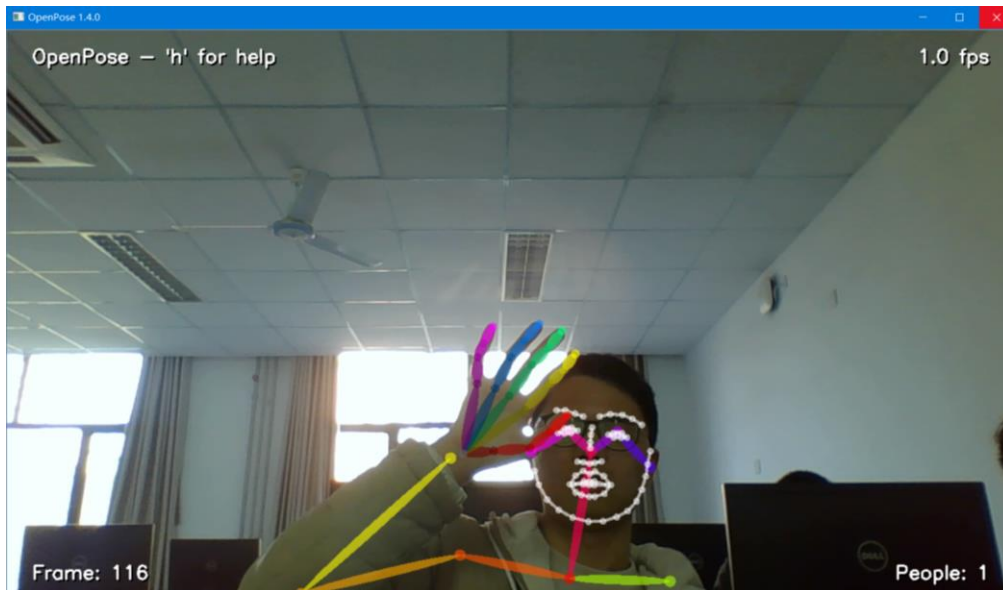


图 2 实时身体、脸部、手部识别

3. 存在的问题

- (1) 对于实时的识别，运行速度较慢，需要对参数进行优化，如牺牲图像像素换取更高的计算速度，或换用显卡性能较高的电脑。
- (2) 得到人体的 3D 姿态需要同一对象的多视角图片作为输入，实时的 3D 人体姿态估计需要一个立体的摄像系统。

4. 关于 openpose

OpenPose 是由卡内基梅隆大学认知计算研究室提出的基于卷积神经网络和监督学习并以 caffe 为框架写成的开源库，可以实现人的面部表情、躯干和四肢甚至手指的跟踪，不仅适用于单人也适用于多人，同时具有较好的鲁棒性。OpenPose 可接受的输入有很多种，可以是图片、视频、网络摄像头等。同样，它的输出也是多种多样，可以是 PNG、JPG、AVI，也可以是 JSON、XML 和 YML。输入和输出的参数同样可以针对不同需要进行调整。OpenPose 提供 C++API，以及可以在 CPU 和 GPU 上工作。其具体功能如下：

2D 实时多人关键点检测：

15 或 18 或 25 关键点的身部/足部关键点估计。运行时间对检测到的人数不变。

2x21 关键点手关键点估计。目前，运行时间取决于在检测到人的数量。

70 关键点面部关键点估计。目前，运行时间取决于在检测到人的数量。

3D 实时单人关键点检测：

来自多个单一视图的三维三角剖分。

处理 Flir 相机的同步。

与 Flir / Point Grey 相机兼容，但提供 C++ 演示以添加您的自定义输入。

校准工具箱：

轻松估计失真，内在和外在的相机参数。

单人跟踪进一步加速或视觉平滑。

其 2D 人体姿态识别主要流程如下：

- (1) 输入一幅图像，经过卷积网络提取特征，得到一组特征图，然后分成两个岔路，分别使用 CNN 网络提取 Part Confidence Maps 和 Part Affinity Fields；
- (2) 得到这两个信息后，使用图论中的 Bipartite Matching（偶匹配） 求出 Part Association，将同一个人的关节点连接起来，由于 PAF 自身的矢量性，使得生成的偶匹配很正确，最终合并为一个人的整体骨架；
- (3) 最后基于 PAFs 求 Multi-Person Parsing—>把 Multi-person parsing 问题转换成 graphs 问题—>Hungarian Algorithm(匈牙利算法)

5. 后期工作计划

