

**实践报告**

|  |  |
| --- | --- |
| 综合设计题目： | 普通RGB视频的动作识别 |
| 姓 名： | 周铭昊 |
| 学 院： | 计算机科学与技术学院 |
| 专 业： | 计算机科学与技术 |
| 班 级： | CS1503班 |
| 学 号： | U201514559 |
| 指导教师： | 陈凯 |

|  |  |
| --- | --- |
| 分数 |  |
| 教师签名 |  |

2018 年 7 月 8 日

# 摘 要

人类的动作捕捉、注意力识别。。。

**目 录**

[摘 要 I](#_Toc518862894)

[1. 绪论 1](#_Toc518862895)

[1.1 课题背景 1](#_Toc518862896)

[1.2 课题研究意义 1](#_Toc518862897)

[1.3 国内外研究情况 1](#_Toc518862898)

[2. 系统需求分析及设计 3](#_Toc518862899)

[2.1 需求分析 3](#_Toc518862900)

[2.2 概要设计 3](#_Toc518862901)

# 绪论

## 1.1 课题背景

行为识别(action recognition)是计算机视觉领域中的一个重要且基础的研究方向，对于运动分析、手势识别、视频内容搜索等方向都有广泛的应用。

尽管近几年图像分类和检索领域实现了很大突破，但是识别视频中的人类动作仍然是一个巨大挑战。原因在于动作本质上没有物体那么明确，这使得我们很难构建精确标注的动作视频数据集。

深度相机，例如微软公司的Kinect。这类RGB-D摄像机虽然能够提供有价值的深度数据，简化了人物动作姿态的重建过程。但是在户外等阳光较强的场景下存在一定的失效，而且体积较大，功耗较高，使用场合受到的约束较大，在当前普遍使用普通RGB摄像头的环境下不够便捷。

构建一个能够识别不带有深度信息的普通RGB视频中人物动作、姿态的系统，可以极大的扩展行为识别在实际生活中的应用范围，而不必受制于Kinect等专业工具。

## 1.2 课题研究意义

Action recognition方向的主要目标是判断一段视频中人的行为的类别，虽然这个问题是针对视频中人的动作，但基于这个问题发展出来的算法，大都不特定针对人，也可以用于其他类型视频的分类。

计算机对人类进行实时的姿态检测，并追踪人物的动作变化。通过深度学习算法达到理解人类的动作姿态，分析人类当前的状态的功能。在机器人应用方向，这项技术可以帮助机器人判断人类现在需要什么，能否被打扰，让机器也能够“察言观色”，可以让机器人在社会空间中更好地服务人类。

而今为了更为精准地为用户提供贴心服务，往往需要获取更多的用户信息，其中用户的动作、姿态信息颇为重要，便捷快速的行为识别是一项必需的技术。

## 国内外研究情况

1. CMU: openpose 研究多人的姿态估计，开源

运行环境：caffe

自下而上，关键点被分组到人的实例

时间：2.8-3.4fps

Github:  [https://github.com/CMU-Perceptual-Computing-Lab/openpose](https://github.com/mpatacchiola/deepgaze)

1. 谷歌：deepgaze 研究头部姿态和注意力方向，开源

主要为头部姿态估计，先框出人脸区域，再进行姿态估计。

运行环境：tensorflow+python

运行速度较快，基本满足实时性，准确度略差。

Github: <https://github.com/mpatacchiola/deepgaze>

1. 马克斯普朗克：Vnect 用2D摄像头实现低成本动捕，开源

# 2. 系统需求分析及设计

## 2.1 需求分析

通过2D摄像头记录的视频进行人物动作识别，头部姿势估计和分类。当眼睛区域可见时，可以估计凝视方向，这可以给出用户当前注意力方向的良好指示。

主要功能要求包括：

1. 获取RGB摄像头的拍摄内容并且实时显示；
2. 对RGB视频进行人物的行为识别；
3. 将识别出的内容进行人物姿态重建，并实时显示；
4. 在眼睛区域可见时，估算出当前的注意力方向；
5. 有功能清晰的用户交互界面。

## 2.2 概要设计

本系统的设计目的是结合深度学习，利用普通RGB摄像头, 达到RGB-D摄像机等级的实时人体姿态估计效果，并对识别的结果进行有效分析。

首先获取RGB视频信息，然后对视频进行行为识别。行为识别的过程主要由两个组件组成。第一是卷积神经网络(CNN)，用于进行人脸检测，头部姿态估计和分类，从而从2D图像空间中得到联合位置；第二是将识别出的关键点信息与关节位置相拟合，以确保时间的平滑追踪，提高姿态的稳定性。

最后再将行为识别的结果通过图形界面输出给用户，识别出包括人物的动作类别、姿态类别、注意力方向等信息。

## 2.3 --设计

Deepgaze是一个基于OpenCV和Tensorflow，用于人机交互、人物检测和跟踪的库，包含有用的包：

* 头部姿态估计（Perspective-n-Point，卷积神经网络）
* 人脸检测（Haar Cascade）
* 皮肤和颜色检测（范围检测，反投影）
* 基于直方图的分类（直方图交叉点）
* 运动检测（帧差分，MOG，MOG2）
* 运动跟踪（粒子滤波器）
* 显着性图（FASA）

OpenPose是第一个在单个图像上联合检测人体，手和面部关键点（总共135个关键点）的实时多人系统。主要功能有：

* 2D实时多人关键点检测：
  + 15或18或25关键点的身体/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。

通过对Deepgaze中关键技术的使用，可以实现RGB视频中人脸信息的识别，但本系统不仅限与人脸的识别，还包括动作姿态的捕捉。因此需要用到OpenPose来进行人物关键点的捕捉，从而重建人物骨架，以达到动作识别的目的。