# 实习任务描述

## 实习任务内容

当前基于深度信息的动作识别主要利用深度数据集中的骨架和点云信息，在本课题中我们将体感信息定义为如以上两者，能够从深度图像中提取的，且能够表现人体动作和行为特征的信息。

在对最近几年提出的基于深度信息识别动作的相关方法进行回顾的基础上，我们需要总结当前动作识别方法，以骨架关节特征、三维模型特征、空-时特征和学习特征作为思路对现有动作表示方法进行分类，并以此讨论不同方法在不同环境（背景变化、视角变换、噪声和遮挡）中的适用性。最后，为了应对视频监控的实时预警所产生的大量数据，本文使用在线学习的方式对动作识别框架进行优化，并提出对未来研究方向的相关建议。

## 预期成果

实现不同动作识别模型，并利用不同模型在不同深度图像数据集上分别采用跨目标测试和交叉验证的方法进行对比测试。通过对不同的动作识别方法得出的测试结果（准确率和混淆矩阵），评价不同动作识别模型的优缺点。具体使用的数据应包含较长的视频样本与较短的视频样本（如：MSR Action3D和CAD-60）以及在一些的特性（背景、自遮挡和噪声等），如：RGBD-HuDaAct、MSR DailyActivities3D等。

## 进度时间安排：

|  |  |
| --- | --- |
| 2019.2-2019.3 | 机器学习与深度图像处理等相关文献的研读 |
| 2019.3-2019.4 | 搜集不同数据集，研究动作识别技术的评价体系 |
| 2019.4-2019.5 | 对多个动作识别方法实现并测试 |
| 2019.5- | 毕业论文的撰写与答辩 |

# 实习任务原理

## 实验方案

### 数据集划分

对于动作识别精确度的判断，目前主要采用跨目标验证和交叉验证的方法。

跨目标验证的思想是：训练样本与测试样本分别来自不同动作执行者的动作序列。此方法便是为了解决上一节提到的，同类型的动作不同动作执行者模型评价问题。

表 1 部分深度数据集资料[19]

|  |  |  |  |
| --- | --- | --- | --- |
| 数 据 集 | 类别数 | 样本数 | 特 性 |
| MSR Action3D | 20 | 567 | 10个演员，每类动作每个演员执行2~3次；提供20个关节点的3D坐标数据、深度图像与RGB图像；视频序列为无背景的纯人体运动目标 |
| UTKinct Action | 10 | 200 | 10个演员，每类动作每个演员执行2次；提供20个关节点的3D坐标数据 |
| MSR DailyActivities3D | 16 | 320 | 10个演员；大部分样本涉及到人和物体的交互；捕获的3D关节点坐标受噪声污染严重 |
| Florence 3D Action | 9 | 215 | 10个演员，每类动作每个演员执行3次；动作相似性大，包含人与物体的交互，同类动作具有不同的执行方式 |
| RGBD-HuDaAct | 12 | 1189 | 30个演员，每类动作每个演员执行2~4次；提供深度图像与RGB图像，样本中混有随机背景动作 |
| MSR ActionPairs | 6 | 180 | 10个演员，每类动作每个演员执行3次；每个动作对有相似的运动和形状 |
| CAD-60 | 12 | 60 | 4个演员，在5个不同的场景中执行动作；提供15个关节点的3D坐标数据、深度图像与RGB图像 |

交叉验证是用来验证分类器性能的一种常用统计分析方法，基本思想是按照一定的划分方式将原始数据集进行分组，一部分作为训练集，另一部分作为验证集。首先用训练集对分类器进行训练，再利用验证集来测试训练得到的模型。评价分类器的性能指标将使用验证集中的测试结果得出。

表 2每批训练的动作执行者划分表

|  |  |  |
| --- | --- | --- |
| 分组序号 | 训练动作执行者序号 | 测试动作执行者序号 |
| 1 | 3、1、10、5、2 | 4、6、7、8、9 |
| 2 | 10、5、2、4、3 | 1、6、7、8、9 |
| 3 | 3、5、6、2、7 | 1、4、8、9、10 |
| …… | …… | …… |

跨目标验证实验分多批进行，每批测试使用不同的动作执行者作为测试集与训练集，正如表2所示。在所有批次测试结束后，对正确率和混淆矩阵求均值，作为最终的模型评价指标。在我的实验中，总共有8批这样的训练。

### 模型的评价

在实验和论文中，主要使用正确率和混淆矩阵对模型进行评价。以下给出正确率和混淆矩阵的定义。

对一批二分类样本进行分类后，对于每一个样本，其分类结果必然属于表3四种情况之一，模型分类的准确率如公式1所示。

表 3 二分类结果可能出现的情况

|  |  |  |
| --- | --- | --- |
| 真正例（True Positive，简称TP） | 将一个正例正确判断成一个正例 | |
| 伪正例（False Positive，简称FP） | 将一个反例错误判断为一个正例 | |
| 真反例（True Negtive，简称TN） | 将一个反例正确判断为一个反例 | |
| 伪反例（False Negtive，简称FN） | 将一个正例错误判断为一个反例 | |
|  | | (1) |

将四种情况以如表4表 4 分类准确率的二维分布表二维表格形式表示，便可以清晰地表示出模型分类性能，以及哪些类更加容易混淆。

表 4 分类准确率的二维分布表

|  |  |  |  |
| --- | --- | --- | --- |
|  | | 预测分类 | |
| 0 | 1 |
| 实际分类 | 0 |  |  |
| 1 |  |  |

除去标签，便是二分类的混淆矩阵定义。对于分类问题，混淆矩阵为一个的矩阵。

## 动作识别的原理

### NBNN分类器原理

NBNN分类器使用的是类似于最邻近方法对视频序列进行分类，即找出与待分类样本“距离”最近的样本，即为样本所属的类别。对每一个可能的分类求出“距离”后，找到“距离”最近的样本。该样本所属的类别便是待分类样本的类别。该距离的算法如下：

|  |  |
| --- | --- |
|  | (2) |

下面开始介绍“距离”的计算方法：为每个视频样本所拥有的帧数，为视频序列每一帧的描述符，指的是描述符在类内的最近邻。即是姿态特征的抽象表示，本节中使用特征关节作视频每帧的描述符。在利用公式2求出所有可能的分类的“距离”后，可得待分类视频样本 所属的分类。

### 隐马尔可夫模型原理

使用隐马尔可夫模型的动作分类大致包含三个过程，即特征抽取、向量量化和离散隐马尔可夫建模。特征抽取使用线性判别分析法（Linear Discriminant Analysis，简称LDA），也叫做Fisher线性判别(Fisher Linear Discriminant，简称FLD)。线性鉴别分析的基本思想是将高维的模式样本投影到最佳鉴别矢量空间，以达到抽取分类信息和压缩特征空间维数的效果，投影后保证模式样本在新的子空间有最大的类间距离和最小的类内距离，即模式在该空间中有最佳的可分离性。向量量化则使用K聚类算法，将训练集中的每一个视频样本中的每一帧的特征描述符量化为视觉词。

我们将三元组视作一个HMM模型，对向量量化后的视频特征训练个HMM模型。对于一个输出序列，其分类方法如公式3所示：

|  |  |
| --- | --- |
|  | (3) |

# 实现及应用

## 数据集样本的标准化

当前用于动作识别的深度图像数据集种类繁多，且形式不一。由于所用数据集中的视频序列存在长短不一的情况，例如，本文使用的MSR Action3D数据集，每个样本最多只有50帧，而小样本只有20多帧。作为分类器样本的输入，每个样本应该有相同的大小。此问题的对应解决办法有删去长度多余的部分或者采用插值的办法调整样本的大小。如果采用删去长度多余部分，且数据集视频序列长度本身就偏小，会造成用于表示动作的视频序列长度不足，使分类器模型欠拟合。

## 动作的表示

本课题使用骨架关节特征、三维模型特征、空-时特征和学习特征作为思路，对现有方法进行分类，并对每种分类中具有代表性的方法进行对比和评价。

|  |  |
| --- | --- |
| **方法分类思路** | **具体方法名称** |
| 骨架关节特征 | 绝对关节位置、相对关节位置、特征关节和三维关节直方图 |
| 三维模型特征 | 随机占用模式和局部占用模式 |
| 时-空特征 | 傅里叶时间金字塔和空时占用模式 |
| 学习特征 | 卷积神经网络 |

# 任务完成情况

骨架关节的动作表示方法和对应的分类器均已实现。对不同分类器和不同骨架关节特征进行分类后，可以得到其正确率如表5所示。表中隐马尔可夫模型的实验结果很不理想，初步猜测是模型参数（如状态数与输出数等）设置的不够合理。可尝试将状态数加大，而输出数是对每一帧动作视频的特征描述符聚类后的类数，应将聚类后的类数减少。

表 5不同分类器与不同骨架关节特征搭配的正确率

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型名称 | 关节绝对位置 | 关节相对位置 | 特征关节 | 关节直方图 |
| 朴素贝叶斯最近邻 | 0.7251 | 0.7802 | 0.8408 | 0.8365 |
| 隐马尔可夫模型 | 0.7189 | 0.7396 | 0.7855 | 0.7714 |

# 实习体会

前期在论文准备方面，对自己所写论文类型以及内容安排没有清晰的认识，因此在论文框架编排、相关文献的收集和摘要的撰写上存在困难。其中具体的问题是：论文标题用词不准确且不具有代表性；相关文献不能够找到具有代表性的著作与方法；摘要不能够达到老师的“言简意赅”和“切中要点”的要求。在老师的帮助下，这些困难最终得以解决。

在实验实现多种算法时，由于数学知识和机器学习知识的匮乏，多次出现实验结果不够理想，却不能获知原因的窘境。

因此，需要在接下来的时间内认真研究论文涉及算法的原理，学习相关知识并完善知识体系。同时，论文撰写以及篇章结构的设计应多参考同类型课题相近的论文，做到表述清晰、言简意赅。