视频序列动作识别遇到的问题

样本大小不一

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。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。对于不同的样本，需要进行样本大小的标准化。为了使样本总数不减少，过小的样本不能直接丢弃。在此，我提出两种简单的办法。

1) 从后往前直接截取

对于样本S，从前往后提取前t帧的子序列。如，一个含有最小样本 20 帧，最大样本 50 帧的数据集，从后往前截取前 20 帧，超过 20 帧的深度图像直接丢弃。这样便可以在不减少样本数量的同时，将样本大小标准化。其缺点是，若原数据集样本不均衡，则处理后的数据会保留此问题。

2) 滑动窗口模式

对于样本S，设定步长s，子样本S'。则有$S' = \{S_i, \dots, S_{i+l}\}$，不足$l$的部分直接丢弃。这样做的结果是，对于原先均衡却大小不一的数据集，处理后的数据集会破坏原数据的均衡性。

识别精度与数据冗余的平衡

动作识别过程中大量使用了数据降维方法，如：主元素分析等。其具有降低数据冗余度的方法依赖样本特征与所属分类的相关性的功能，而往往样本容量越大越容易正确地分析出特征和分类的相关性。同时，样本容量大又必将导致数据冗余度上升，模型运行效率变低。反之，样本容量减小使模型运行效率增高的同时，识别精度将会下降。因此需要在识别精度和数据冗余之间找到一个平衡点，达到我们需要的效果。

如图 1 所示，对于目前使用的 MSR Action3D 数据集，每个样本的大小达到 20 帧时，识别的精确度快速收敛。对于样本大小大于 30 帧的数据集，精确度不再随样本大小变化而变化。由此可以得出，对于视频序列的识别，含有 20 帧左

右的样本最为高效。

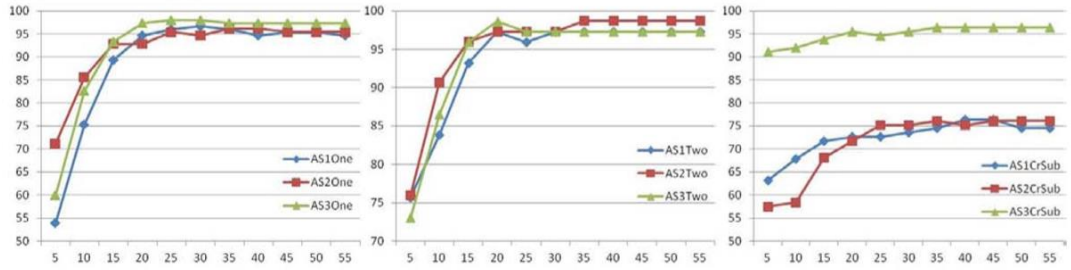

图 1 交叉检验精确度随样本大小的变化[1]

主元素分析目的维数的约束

对于每一帧图像中含有维度$n \times p$的特征f进行主元素分析，并将其降维至p'维，则有$1 \leq p' \leq p$。本课题使用的实验环境为 Scikit-Learn 的 Python 模块，其默认的 PCA 方法，使用奇异值分解，并默认$1 \leq p' \leq n$。而参考文献中的目的维数满足$n \leq p' \leq p$。故需要对该方法重新实现。

视频序列的数据冗余度降低

对于从每个帧t中提取多种类型的特征，如：3D 关节位置特征、LOP 特征和 ROP 特征等。在这个小节中，我们介绍文献[2]提出傅立叶时间金字塔模型，来表示这些帧级特征的时间模式。

傅立叶时空金字塔是一种能够良好展示动作时间结构的描述性特征。为了捕捉动作的时间结构，除了全局傅里叶系数之外，我们递归地将动作划分为金字塔，并对所有分段使用短时傅里叶变换，如图所示。最终特征是来自所有段的傅里叶系数的组合。

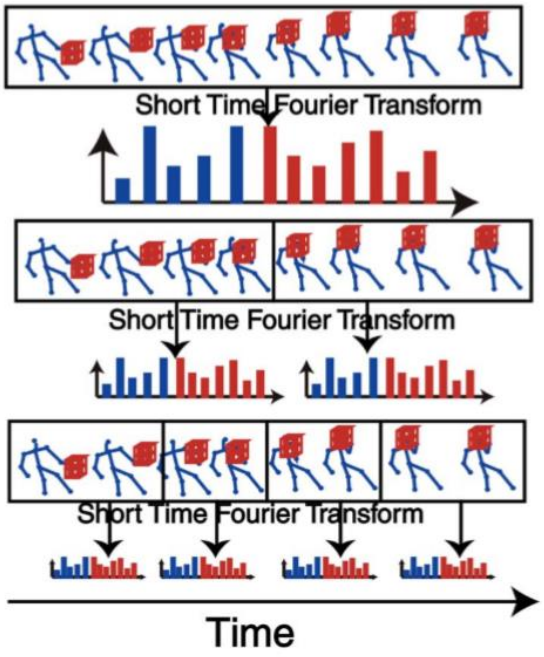

图 2 时间金字塔原理示意图[2]

对于每个关节j，令$g_j(p_j, o_j)$表示其整体特征向量，其中p_j是其 3D 成对位置向量而o_j是其 LOP 向量。令N_j表示g_j的维数，即$g_j = (g_1, g_2, \dots, g_{N_j})$。每个元素$g_n$是时间的函数，我们可以将其写为$g_j[t]$。对于每个金字塔等级的每个时间段，我们将短傅里叶变换应用于元素$g_n[t]$，并获得其傅里叶系数，并将它们用作慢速频率系数作为特征。关节j处的傅立叶时间金字塔特征是定义为金字塔所有层次的低频系数，并表示为G_j。

使用傅立叶时空金字塔特征有几个好处。首先，通过丢弃高频傅立叶系数，所提出的特征对噪声具有鲁棒性。其次，该特征对时间错位不敏感，因为时间转换的时间序列具有相同的傅里叶系数幅度。最后，动作的时间结构以金字塔结构为特征

引用文献

[1] Yang X , Tian Y L . EigenJoints-based action recognition using Naïve-Bayes-Nearest-Neighbor[C]// 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops). IEEE Computer Society, 2012.

[2] Wang J , Liu Z , Wu Y , et al. Learning Actionlet Ensemble for 3D Human Action Recognition[J]. IEEE Transactions on Software Engineering, 2013, 36(5):914-927.