

基于头部姿态估计的听障生课堂 专注度监测方法及系统

卓 然¹, 高 飞²

(1. 浙江特殊教育职业学院 公共基础学院, 杭州 310023;

2. 浙江工业大学 计算机学院, 杭州 310023)

摘要: 对听障学生在特殊教育课堂中仅依赖手语教学情境下的专注度监测问题进行了研究; 针对该问题, 构建了基于头部姿态估计的课堂专注度评估框架; 采用欧拉角计算方法实现单人头部姿态的量化估计, 并结合余弦函数与采样策略设计归一化的课堂专注度评估计算模型; 经实验测试, 所提出的方法实现了对听障学生课堂专注度的有效估计, 系统运行稳定、响应及时; 经实际应用验证, 该系统满足特殊教育课堂中对非侵入式、实时专注度监测的需求, 具有一定的实用价值。

关键词: 特殊教育; 听障生; 头部姿态估计; 欧拉角; 课堂专注度

Method and System for Monitoring Classroom Concentration of Deaf Students Based on Head Pose Estimation

ZHUO Ran¹, GAO Fei²

(1. School of General Education, Zhejiang Vocational College of Special Education, Hangzhou 310023, China;

2. College of Computer, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract: Research was conducted on the attention monitoring for hearing-impaired students in special education classrooms relying solely on sign language instructions. To address this issue, a classroom attention assessment framework based on head pose estimation was established, which employed the Euler angle-based method to achieve the quantitative estimation of individual head pose, and presented a normalized attention assessment model by integrating a cosine function with a sampling strategy. Experimental evaluations demonstrate that the proposed approach effectively estimates classroom attention levels of hearing-impaired students, with stable operation and prompt response. Practical applications verify that the system meets the requirements of non-intrusive, real-time attention monitoring in special education classrooms, demonstrating a certain practical value.

Keywords: special education; deaf students; head pose estimation; Euler angles; classroom concentration

0 引言

听障生作为特殊教育群体, 主要通过视觉来接收几乎所有的信息, 在课堂教学中, 如果听障生的人脸姿态和凝视方向都不面向教室前方的教师和黑板等信息源, 那么他们就无法通过其他途径去接收信息, 即可以判定其注意力不集中, 因此, 通过监测头部姿态实现课堂专注度计算具有较高的可行性, 这对于提高教育质量、优

化教学策略具有重要意义。

国内外对课堂专注度评估已有不少研究。文献[1]总结了基于深度学习的多模态情感识别最新进展, 涵盖发展里程碑、数据集、深度学习模型原理、特征提取与信息融合方法。文献[2]以课堂监控为传感器, 提出融合人脸与步态识别的多模态生物特征融合检测方法, 实现课堂情况的监测。文献[3]使用头部姿态估计模型获取用户的头部欧拉角, 将头部欧拉角通过计算转换

收稿日期: 2025-10-24; 修回日期: 2025-12-01。

基金项目: 浙江省自然科学基金项目(LTGG23F020003)。

作者简介: 卓 然(1986-), 男, 硕士, 副教授。

通讯作者: 高 飞(1974-), 男, 博士, 教授。

引用格式: 卓 然, 高 飞. 基于头部姿态估计的听障生课堂专注度监测方法及系统[J]. 计算机测量与控制, 2026, 34(6): 49-54, 61.

为专注度评分,但其采用了简单的权重计算,没有考虑到时间的影响。文献[4]根据头部姿势和眼睛纵横比检测学生的参与度,采用二分类法评估学生的专注状态,该方法无法归一化专注度,存在较大误差。文献[5]和文献[6]等均统计抬头率来评估专注度,该方法头部特征过于单一、鲁棒性较差。文献[7]提出了基于面部情感特征的专注度评估方法。文献[8]提出了基于学生对教师动作面部反应相似系数的专注度评估方法,并构建了基于 SVM 分类的专注度加权计算模型。文献[9]通过眼动仪分析了眼神注视和课堂教学质量的关系,但并未深入探讨眼神凝视与学生的专注度之间的直接联系。文献[10]提出了基于眼睛凝视方向的用户专注度计算方法。文献[11]融合凝视方向、头部姿势和面部表情等多模态数据训练姿态识别模型用以推理学生的专注度,该方法过于复杂,多模态数据难以采集。

虽然已有不少课堂专注度评估的研究,但仍存在难以归一化、多模态数据难以采集、模型过于复杂等问题,为此,针对特殊教育场景下听障生的视觉交互特点,提出基于头部姿态估计的欧拉角计算方法及基于余弦函数及采样计算的归一化课堂专注度评估方法,设计并实现了听障生课堂专注度监测系统,实验证明,该方法及系统简单有效。

1 听障生课堂专注度评估

1.1 关键技术与方法实现概述

本系统围绕听障生课堂专注度实时、非侵入式检测的核心需求,重点解决了如下关键技术问题:

1) 轻量化头部姿态估计。针对特殊教育课堂算力资源有限、需兼顾实时性与准确性的特点,系统采用基于人脸关键点集合投影的欧拉角估计算法。该方法仅依赖 4 个稳定面部关键点(双眼、鼻尖、嘴中点),快速估算 Pitch、Yaw、Roll 三个方向的姿态角,避免了复杂三维重建或深度神经网络推理带来的高计算开销,显著提升了处理效率。

2) 归一化专注度建模方法。考虑到听障生主要依赖视觉注释获取信息,其头部越正对讲台,专注度越高。为此,本方法创新性地引入余弦函数将 3 个欧拉角映射至 $[0, 1]$ 区间,有效实现了姿态角到专注度的非线性平滑转换。该方法不仅物理意义明确,且天然具备对称性和单调性,确保评估结果与人类观察一致。

3) 动态采样与缓冲调度机制。为应对多摄像头并发接入与分析速度不匹配的问题,系统设计了基于循环队列的视频帧缓冲区,并采用轮询采样策略(如每 5 s 抽取一帧)。当缓冲区满时自动丢弃旧帧,防止内存溢出;同时支持按教师规模动态调整采样频率,平衡精度

与系统负载。

4) 鲁棒的人脸检测与 ID 绑定机制。在实际课堂环境中,存在光照变化、遮挡、后排学生图片分辨率低等挑战。系统集成经过优化的 DSFNet 模型^[12],具备较强的遮挡鲁棒性,并通过摄像头 ID 与课堂实体关联,确保每帧分析结果能准确回溯至对应班级、课程及学生群体,支撑后续统计分析。

上述技术共同构成了本系统低侵扰、高实时、易部署的核心优势,使其特别适用于特殊教育场景下的常态化教学监测。

1.2 专注度评估框架

听障生课堂专注度评估框架如图 1 所示,主要由如下模块构成:视频序列获取、人脸检测、头部欧拉角计算、单人姿态量化估计、单帧采样估计、课堂整体专注度计算。在视频序列获取中,根据教室摄像头的 IP 地址、用户名、密码等信息,基于 RTSP 协议获取每一帧图像。在人脸检测和欧拉角计算模块,主要根据算法获取每张人脸的关键点信息,包括眼睛、鼻子等图像坐标,然后计算头部欧拉角,包括 Pitch、Roll 和 Yaw 三个方向角度。采用余弦函数分别将 3 个欧拉角归一化到 $[0, 1.0]$ 的范围,进而取其平均值,即可得到单人姿态的量化估计值。在此基础上,通过计算该帧单人姿态的平均值即可得到每帧采样估计结果,进而得到课堂整体专注度。

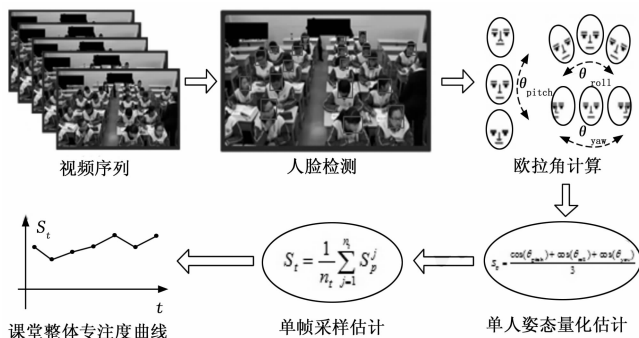


图 1 听障生课堂专注度评估框架

1.3 头部姿态估计

考虑到课堂专注度监测并不需要特别精确的数值,其归一化数值只需要与专注度基本成正相关即可,即数值越大,说明越专注,数值越小,专注度也越低。为提高计算效率,仅考虑通过头部姿态估计的欧拉角计算得到单人姿态量化估计。具体而言,借鉴文献[13],系统采用透视投影几何的头部姿态估计算法。该模型将三维人脸刚性结构投影到二维图片平面,通过已知的 3D 人脸基准点与其在图像中检测到的 2D 对应关键点之间的对应关系,求解从世界坐标系到相机坐标系的旋转和平移变换问题。在此基础上,提取旋转矩阵并转换为欧

拉角 (Pitch、Yaw、Roll), 分别表示头部在俯仰、偏航和滚轮 3 个自由度上的姿态。

为兼顾计算效率与估计精度, 选取 4 个具有高稳定性和语义意义的人脸关键点作为 2D 观测点: 1) P_1 : 左眼中心; 2) P_2 : 右眼中心; 3) P_3 : 鼻尖; 4) P_4 : 嘴部中间。对应的 3D 模型点采用标准平均人脸模型中的预定义坐标。这些关键点在正面视角下分布合理, 能有效约束 3 个旋转自由度, 且在实际课堂视频中即使存在轻微遮挡或光照变换, 仍具备较高的检测鲁棒性。

在获得 2D~3D 点对后, 系统调用 EPnP 求解位姿, 进而将旋转矩阵 R 分解为欧拉角 θ_{roll} 、 θ_{yaw} 和 θ_{pitch} 。为简化表述并适应轻量化部署需求, 本文进一步采用几何近似方法, 直接利用关键点间的相对位置关系估算角度, 如图 2 所示。根据这 4 个关键点在图像平面上的投影坐标 $P_1(x_1, y_1)$ 、 $P_2(x_2, y_2)$ 、 $P_3(x_3, y_3)$ 和 $P_4(x_4, y_4)$ 可按式 (1)~(3) 快速估算 3 个方向的欧拉角:

$$\theta_{pitch} = \arctan \frac{z_1 - z_2}{|y_1 - y_2|}, (y_1 \neq y_2, x_3 \neq x_4) \quad (1)$$

$$\theta_{roll} = \arctan \frac{y_3 - y_4}{|x_3 - x_4|}, (x_3 \neq x_4) \quad (2)$$

$$\theta_{yaw} = \arctan \frac{z_4 - z_3}{|x_3 - x_4|}, (x_3 \neq x_4) \quad (3)$$

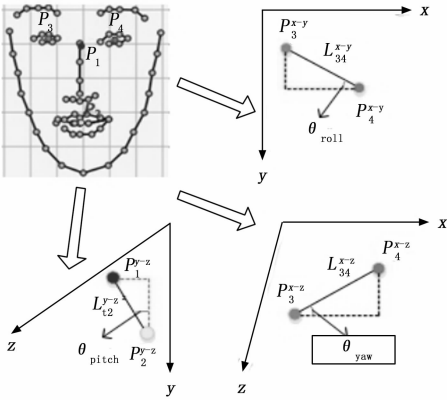


图 2 头部姿态估计

上述简化策略虽牺牲部分几何精度, 但在听障生课堂这一特定场景下 (学生基本面向讲台、姿态变换范围有限), 仍能保证姿态估计结果与真实专注状态高度一致, 同时显著降低计算开销, 满足实时性要求。

1.4 课堂专注度评估

由于听障生课堂教学主要依靠手语, 因此, 越专注的学生, 应该双眼紧盯讲台, 即头部姿态越正, 则 θ_{roll} 、 θ_{yaw} 和 θ_{pitch} 越接近于 0° ; 而越不专注的学生, 双眼视线严重偏离讲台, 即头部姿态越歪, 则 θ_{roll} 、 θ_{yaw} 和 θ_{pitch} 越接近于 90° 。为此, 为便于归一化度量课堂专注度, 引入余弦函数进行专注度计算, 如式 (4) 所示:

$$S_p = \frac{\cos(\theta_{pitch}) + \cos(\theta_{roll}) + \cos(\theta_{yaw})}{3} \quad (4)$$

其中: S_p 表示某个学生的专注度, 很明显, $S_p \in [0, 1.0]$, 越接近 1, 表示越专注, 越接近 0, 表示越不专注。

对于整个课堂的专注度评估, 采用一定时间间隔进行单帧采样分析。由于人脸检测准确率并不能保证每次采样结果都一样, 因此, 对每帧采样进行平均, 得到单次采样的专注度评估结果, 如式 (5) 所示, 然后, 可根据式 (6) 计算得到整堂课的专注度评估结果 S :

$$S_t = \frac{1}{n_t} \sum S_p \quad (5)$$

$$S = \frac{1}{n} \sum S_t \quad (6)$$

其中: n_t 表示第 t 帧采样检测到的学生数量, S_t 表示第 t 帧专注度评估值, n 表示整堂课的采样数量。

2 听障生课堂专注度监测系统

2.1 功能架构

听障生课堂专注度监测系统的功能架构如图 3 所示, 主要包括系统配置、数据管理及查询统计等 3 大主要模块, 具体功能下面详述。

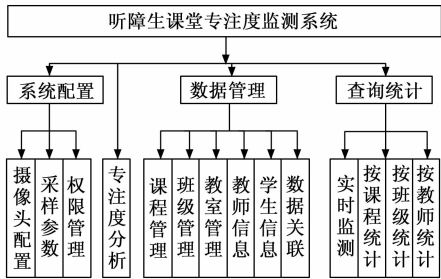


图 3 系统功能架构

为保障系统高效、稳定运行, 各功能模块通过标准化的数据接口与统一的数据库进行松耦合交互。系统整体采用“配置驱动—数据绑定—分析执行—结果反馈”的闭环运行机制, 其核心数据流如下:

1) 系统配置→数据管理: 摄像头与课堂实体的绑定。在系统配置模块中完成摄像头 (IP 地址、用户名、密码、端口等) 录入后, 管理员需通知“数据关联”子功能将该摄像头与具体的课堂实体建立一对一映射关系。该映射通过数据库中的 `camera_classroom_link` 表实现, 字段包括 `camera_id` 和 `class_session_id`。而 `class_session` (课堂会话) 实体本身已关联课程、班级、教师、教室等信息, 因此, 一旦摄像头被绑定至某堂课, 系统即可自动推导出完整的上下文元数据。

2) 数据管理→专注度分析: 动态任务调度依据。专注度分析模块启动时, 首先从数据库的 `class_session` 表中查询当前处于进行中状态的课堂记录, 并通过

外键关联获取对应的摄像头配置参数 (IP、认证信息等) 及学生名单。这些信息被封装为任务描述符 $\langle \text{session_id}, \text{camera_config}, \text{student_list} \rangle$, 供采样线程和分析线程使用。此过程确保分析任务始终与最新教学安排同步, 支持多课堂并发处理。

3) 专注度分析→数据管理: 结果持久化。分析线程每完成一帧图像的处理, 即生成结构化记录, 包括: session_id 、timestamp、检测到的学生 ID 列表、每位学生的欧拉角与个体专注度评分、当前帧整体专注度值等。该结果通过 ORM 接口写入 attention_record 表, 与所属课堂、学生实体自动关联, 便于后续追溯与聚合统计。

4) 数据管理↔查询统计: 按需数据服务。查询统计模块不直接访问原始视频或分析算法, 而是通过预定义的 SQL 视图或 RESTFUL API 从数据库中提取聚合数据。实时监测功能定时轮询 attention_record 表中最近 30 s 内对应 session_id 的记录, 计算华东平均专注度并推送至前端; 按班级统计功能则通过 JOIN class_session , attention_record , student , class 等多表关联, 按天/周/课程维度聚合平均专注度; 所有查询均基于索引优化 (如 $\text{session_id} + \text{timestamp}$ 联合索引), 确保相应延迟低于 500 ms。

整个系统采用事件驱动加轮询混合机制, 课堂开始/结束事件由教务系统或人工标记触发, 激活/终止对应分析任务; 而在任务执行期间, 采样与分析线程以固定间隔轮询摄像头流, 形成稳定的处理流水线。

通过上述设计, 系统实现了配置灵活、数据一致、分析解耦、查询高效的运行模式, 既满足特殊教育课堂对实时性的要求, 又保障了历史数据的可追溯性与统计分析的准确性。

2.2 模块功能

2.2.1 系统配置

系统配置模块主要包括摄像头配置、采样参数、权限管理等功能。摄像头配置包括摄像头添加、删除、修改、查询及参数设置等功能; 参数设置包括摄像头 IP 地址、用户名、密码等通讯信息; 采样参数主要用于设置图像采样间隔, 以适配专注度分析及相应算力; 权限管理则主要用于维护系统用户的权限、用户名、密码等信息。

2.2.2 数据管理

数据管理主要包括课堂、课程、班级、教室、教师、学生、专注度等实体信息维护, 其实体关系如图 4 所示, 其中, 课堂是整个数据库的牵引, 每堂课对应一个摄像头, 而一个摄像头可以对应多堂课; 每堂课归属于一门课程, 每门课程包含多堂课; 每课堂由一位教师进行授课, 而每位教师可教授多堂课; 每堂课有多位学生在上课, 而每位学生可以聆听多堂课; 每堂课的采样

分析结果明细保存在专注度实体中。很显然, 根据课堂信息, 可以找到相应的摄像头、学生、班级、教师、课程、专注度及教室等信息。

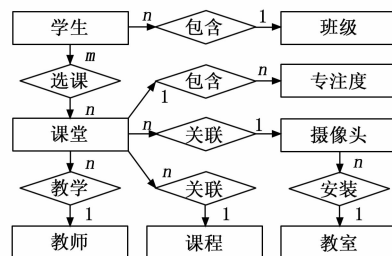


图 4 系统实体关系图

2.2.3 专注度分析

专注度分析模块主要采用至少两个线程执行相应的逻辑。系统开启后, 在后台启动采样和分析两个线程。采样线程首先从数据库中读取正在教学的课堂信息, 然后根据实体关系, 可查询到配置的摄像头及其 IP 地址、用户名、密码、ID 等参数信息, 然后采用轮询方式, 连接摄像头, 根据每隔多少秒采样一次等采样策略, 获取到图像帧, 并与 ID 等构建图像数据结构, 加入缓冲区。分析线程启动后, 采用循环方式, 从缓冲区逐帧读取图像数据, 根据第 1 节的算法, 分别进行头部检测、头部姿态估计及单帧专注度分析, 并根据缓冲区每帧图像对应的 ID, 找到对应的教室及课堂, 将结果存入数据库中。整个专注度分析模块的核心流程如图 5 所示, 其中, c_i 表示第 i 个摄像头及其相关配置, p_i 表示当前获取到当前摄像头 c_i 的视频帧, ID_i 表示摄像头 c_i 对应的 ID。

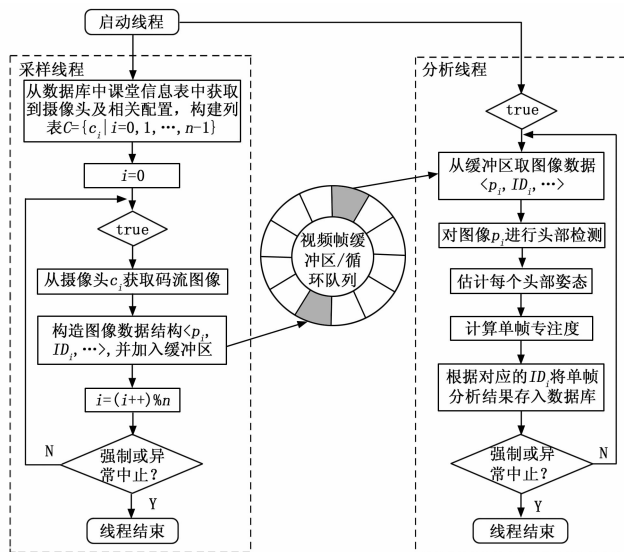


图 5 专注度分析模块的核心流程

考虑到取码流与分析速度并不完全匹配, 为防止内存溢出, 设计了视频帧缓冲区, 具体而言, 由于图像数

据通常较大,为避免因为重复开拓大内存带来的计算开销,设计了循环队列作为缓冲区,并采取一定的策略防止队列溢出,例如,若队列已满,则直接丢弃采样图像。此外,根据摄像头数量、算力配置等,可设计多个采样和分析线程,或者采用分布式架构,采用多台算力设备进行计算,确保系统与算力相适配。

2.2.4 查询统计

查询统计主要包括实时监测、按课堂统计、按班级统计及按教师统计等功能。实时监测功能可实时点击任意正在教学的课堂,查看其实时采样的专注度情况,按课堂、班级或教师等统计主要查询历史课堂的专注度情况,进而为反哺课堂教学提供参考。

2.3 系统监测实例与结果分析

为更直观地说明系统在实际教学场景中的运行结果,以下为某特殊教育学院一堂手语语文课为例进行说明。

该课堂共有18名听障学生,教室前方安装有一台支持RTSP协议的高清网络摄像头。系统配置采样间隔为5s,摄像头参数已在系统配置模块中完成录入。课程开始后,系统自动启动采样线程与分析线程。

在某一采样时刻,如上课第12分钟,摄像头捕获到当前帧图像。系统通过人脸检测算法识别出16名学生的面部区域,其中有两名学生因低头记笔记导致面部被遮挡未被检出。随后,利用DSFNet模型提取每张人脸的关键点,利用式(1)~(3)计算出每位学生的头部欧拉角(Pitch、Yaw、Roll)。例如,学生A的3个角度分别为 $\theta_{roll}=8^\circ$ 、 $\theta_{yaw}=-5^\circ$ 、 $\theta_{pitch}=2^\circ$;而学生B的3个角度为 $\theta_{roll}=42^\circ$ 、 $\theta_{yaw}=38^\circ$ 、 $\theta_{pitch}=15^\circ$ 。将学生A和B的3个角度分别代入式4计算其个体专注度,学生A的专注度为0.995,学生B的专注度为0.832,因此该帧的整体专注度为0.87。该结果显示在系统前端页面,教师可通过实时监测功能查看当前班级专注状态,若连续多帧 S_t 低于预设阈值,触发报警提示,辅助教师调整教学节奏、更换互动方式。

3 实验验证

3.1 模型评估方法

基于头部姿态估计的听障生课堂专注度监测系统已在相关特殊教育学院进行试用。为确保所提出的基于余弦函数与采样策略的归一化专注度评估模型具有良好的准确性、稳定性和场景适应性,本文设计了多维度的验证方案。

在基准对比方面,本文将所提方法与人工观察评估结果进行系统性对照。对于人工方法的评估,由两位具备特殊教育经验的教师独立制定学生专注度观察表,每5分钟记录每位学生的典型行为作为评估标准,评估范

围为0~1。典型行为分为如下4级,每满足一级,专注度增加0.25:1)涵盖眼神注视:对学生的眼神接触,是否注视教师、黑板或学习材料;2)身体姿态:观察学生的身体姿态,其坐姿是否端正、面向讲台,还是频繁转头、趴桌、玩文具等;3)课堂参与:学生是否主动回答问题、举手发言、有无参与小组讨论;4)任务完成:学生是否按时完成课堂练习、笔记记录的完成程度。该量表参考了教育心理学中常用的课堂参与度观察工具,并在校听障生课堂中经过多轮试用优化,具备较高的内部一致性。

在实验设计上,构建了一个包含10个真实课堂视频的数据集,覆盖不同课程类型(语文、数学、手语表达)、不同时段(上午/下午)、不同班级规模(12~22人)以及不同光照与座位布局条件,以检验模型在多样教学场景下的鲁棒性。每个视频时长10分钟,共采集约600帧有效样本。

在量化指标方面,除计算模型输出与人工评分之间的平均绝对误差(MAE)和平均相对误差(MRE)外,还引入皮尔逊相关系数衡量二者线性相关程度。此外,为考察采样策略对结果的影响,分别在5、15和30s三种采样间隔下重复实验,分析时间粒度对评估稳定性的作用。

3.2 对比实验

表1展示了在不同采样间隔(5、15、30s)下,针对10个视频样本,采用人工方法与本文提出的方法对课堂专注度进行评估的结果对比。通过比较两种方法的评估值及其偏差,可以系统地分析本文方法的准确性、稳定性及适用性。

整体性能可以通过计算平均绝对误差(MAE)如式(7)、平均相对误差(MRE)如式(8)获得。平均绝对误差为0.018,表明本文方法在大多数情况下能够很好地逼近人工评估结果;平均相对误差为2.361%,表明本文方法在整体上具有较高的评估精度:

MAE = 1/N * sum_{i=1}^N |y_i^d - y_i^m| (7)

MRE = 1/N * sum_{i=1}^N |(y_i^d - y_i^m) / y_i^d| * 100% (8)

其中: y^d 为人工评估结果, y^m 为本文方法评估结果。

为了分析采样间隔对结果的影响,将数量按采样间隔分组,分别计算各组的平均偏差。5s采样间隔的偏差均值为1.834%,15s采样间隔的偏差均值为2.232%,30s采样间隔的偏差均值为3.017%,这表明随着采样间隔增加,偏差呈现显著上升趋势。这可能是因为较长的采样间隔导致行为信息丢失,尤其是短暂的注意力波动无法被捕捉,从而影响评估精度。

表 1 与人工方法对比实验

视频序号	采样间隔/s	专注度评估/%		偏差/%
		人工方法	本文方法	
1	5	75	74	1.15
	15	77	77	0.71
	30	72	74	2.79
2	5	73	72	1.02
	15	73	72	2.43
	30	76	79	3.07
3	5	75	73	2.39
	15	80	78	2.44
	30	73	76	3.89
4	5	78	78	0.49
	15	79	78	0.61
	30	72	75	3.81
5	5	84	85	1.13
	15	87	86	0.33
	30	84	83	1.19
6	5	78	78	0.04
	15	76	74	2.89
	30	79	80	1.08
7	5	62	63	1.49
	15	65	64	1.39
	30	63	64	2.28
8	5	75	71	4.97
	15	74	77	4.61
	30	75	78	4.67
9	5	75	72	3.73
	15	79	77	2.95
	30	74	72	2.68
10	5	77	79	1.93
	15	77	74	3.96
	30	73	77	4.71

为了分析人工方法与本文方法的线性相关性，采取皮尔逊相关系数计算，如式（9）所示。计算皮尔逊相关系数为 0.943，表明人工方法与本文方法之间存在的极强的正线性相关关系，尽管个别样本偏差较大，但整体模式高度一致：

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

(9)

其中： x_i 为人工方法值， y_i 为本文方法值。

4 结束语

针对现代特殊教育课堂情境中听障生信息交互存在的问题，给出了课堂专注度评估框架，提出了基于头部姿态估计的听障生课堂专注度评估方法，开发了听障生课堂专注度监测系统，并通过实验进行了验证。由于专注度并无客观真实的样本，且其本身是一个主观评估价值，这一稳定在对应数量级上的评估仍具有较好的有效

性和实用性。虽然本文方法及系统具有一定的有效性，但无法检测特定表情，且在复杂场景下如后排座同学的头部姿态估计准确性不足的问题，将来的研究工作可结合微表情检测，这有助于提高特殊教育课堂教学质量的精细化评估。

参考文献：

[1] ZHANG S, YANG Y, CHEN C, et al. Deep learning—based multimodal emotion recognition from audio, visual, and text modalities: a systematic review of recent advancements and future prospects [J]. Expert Systems with Applications, 2024, 237 (C): 121692.

[2] ZHANG T, DAHLAN H A B, XIE Z, et al. Multimodal biometric fusion sentiment analysis of teachers and students based on classroom surveillance video streaming [J]. Applied Mathematics and Nonlinear Sciences, 2024, 9 (1): 1–20.

[3] XU X, TENG X. Classroom attention analysis based on multiple Euler angles constraint and head pose estimation [C] //MultiMedia Modeling: 26th International Conference, Daejeon, South Korea, Proceedings, Part I, Springer, 2020: 1–8.

[4] UCAR M U, ÖZDEMİR E. Recognizing students and detecting student engagement with real-time image processing [J]. Electronics, 2022, 11 (9): 1500.

[5] ZHAO W, JIA S, XUE Q, et al. Calculation method of classroom head-up rate based on head pose estimation [C] // IEEE 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC), 2022: 1–10.

[6] DAI Y, ZHAO H, ZHANG X. Evaluation method of teaching effect based on visual calculation in classroom environment [C] // IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2021: 10–18.

[7] SUMER Ö, GOLDBERG P, D'MELLO S, et al. Multimodal engagement analysis from facial videos in the classroom [J]. IEEE Transactions on Affective Computing, 2021, 14 (2): 1012–1027.

[8] KAWAMURA R, MURASE K. Concentration estimation in e-learning based on learner's facial reaction to teacher's action [C] // Proceedings of the 25th International Conference on Intelligent User Interfaces Companion, 2020: 103–104.

[9] YANG J, LIU C, ZHANG Y, et al. The teacher's eye gaze in university classrooms: evidence from a field study [J]. Innovations in Education and Teaching International, 2023, 60 (1): 4–14.

（下转第 61 页）