基于单幅图像 PnP 头部姿态估计的学习注意力可视化分析

陈平[1]，皇甫大鹏[1]，骆祖莹[2]，李东兴[2]
（1. 北京师范大学信息网络中心，北京 100875；2. 北京师范大学信息科学与技术学院，北京 100875）

摘 要：提出了一种基于单幅图像头部姿态估计的学生注意力可视化分析方法，采用随机级联回归树进行人脸特征点定位，引入了一个统计测量获得的刚性模型作为 3D 人脸近似，实现基于 PnP (perspective-n-point)映射的单幅图像头部姿态估计，最后将学生视线投射到教师授课的视频图像上，实现学生学习注意力的可视化分析。实验结果表明：对于 Biwi 标准库，该方法可以将头部姿态估计角度平均误差降低到 4.88°；方法具有粗颗粒度的计算并行性，使用 4 线程并行计算可以获得 2.37 倍的加速效果；实现了 3 种典型学习状态（专注、关注、漠视）的注意力可视化分析。

关键词：头部姿态；可视化；注意力分析；课堂教学量化
中图分类号：TP391.7
文献标识码：A
doi: 10.11959/j.issn.1000−436x.2018205

Visualization analysis of learning attention based on single-image PnP head posture estimation

CHEN Ping[1], HUANGFU Dapeng[1], LUO Zuying[2], LI Dongxing[2]
1.Center of Information & Network, Beijing Normal University, Beijing 100875, China
2.College of Information Science and Technology, Beijing Normal University, Beijing 100875, China

Abstract: Owing to the fact that the head-mounted eye tracker is unsuitable to be widely used in the large-scale classroom evaluation under expenditure limitation, a novel method was proposed for student learning attention visualization analysis based on single-image PnP head posture estimation. The method applicate the stochastic cascade regression tree technique to locate facial feature points. With the feature points, the PnP (perspective-n-point) mapping technique is used to estimate the head posture based on a 3D rigid facial model obtained through statistical measurement. The method finally projects the gaze point on the frame image of the teaching video, which realizes the visualization of student learning attention. Experiments demonstrate the following advantages of the method. 1) The method limits the average head-posture estimation errors under 4.88°with Biwi database. 2) Thanks to the coarse-grained computing parallelism of the method, the work achieves 2.37X speedup with four threads. 3) the work has implemented student learning attention visualization analyses for three typical learning cases including engagement, attention, disregard.

Key words: head posture, visualization, attention analysis, classroom evaluation

1 引言

课堂教学效果评价是教学质量评价与提升的关键[1]。受限于信息技术的发展水平，目前课堂评价缺乏高采样率（每秒 1-2 次）全过程动态的量化评价手段。而对于课堂教学是否激发学生好奇心、是否引发学生思考、是否学习专注等重要评价指标，则必须通过高采样率全过程动态地量化评价来进行分析[1]。

头戴式眼动仪[2]可以用于课堂评价的学生学习

收稿日期：2018−09−08
通信作者：骆祖莹，luozy@bnu.edu.cn
基金项目：国家自然科学基金资助项目（No.61274033, No.61271198）
Foundation Item: The National Natural Science Foundation of China (No.61274033, No.61271198)

注意力可视化分析[3]。但是头戴式眼动仪价格昂贵[2]，且一台眼动仪只能用于一名学生的注意力分析，因此头戴式眼动仪难于满足课堂评价对教室中 30~50 名学生注意力分析的需求。

为了低成本代价对课堂里众多学生的学习注意力进行可视化分析，本文提出了一种基于两支高清摄像头的学生学习注意力可视化分析方法。如图 1 所示，本文方法利用安装在黑板上方的前摄像头(F)对教室内学生的头部姿态进行估计，获取头部坐标信息的 X/Y/Z 3 个维度和表示头部旋转信息的偏航角(Yaw)、俯仰角(Pitch)、旋转角(Roll) 3 个维度等 6 个维度数据[4]，再利用几何变换技术、将学生双眼中间点处的人脸法线投射到后摄像头(B)拍摄的教师授课场景图像上，以实现学生学习注意力的可视化分析。本文工作主要如下。

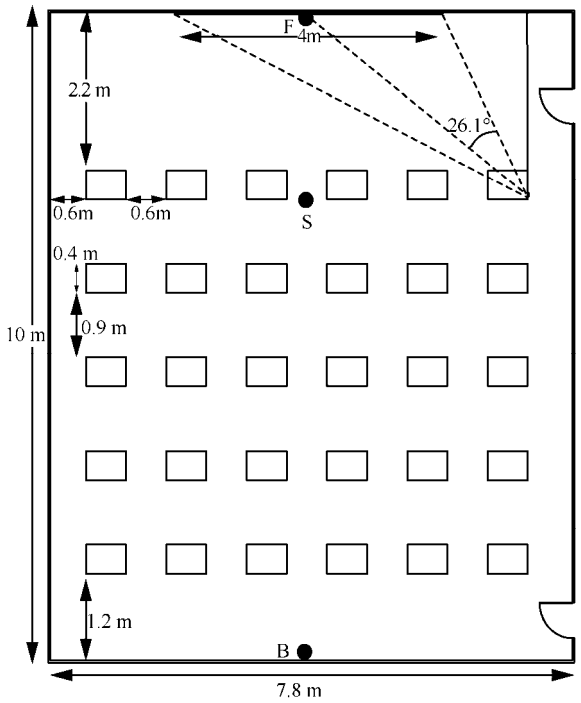


图 1 标准教室的布局及其前后 2 个摄像机的安装位置

1) 提出了一种基于单幅图像的头部姿态估计方法。采用简洁方法对摄像头参数进行标定[5]，再采用随机级联回归树[6]方法检测人脸的 19 个特征点，并用监督下降算法（SDM）[7]来提高特征点定位的速度及顽健性，最后采用一个统计测量的 3D 刚性模型对 PnP 问题[8-9]进行初始化，实现基于 19 个特征点的头部姿态估计。

2) 基于数学推导、提出了一种几何变换方法，利用头部姿态估计参数、将学生鼻尖点处的人脸法线投射到后摄像头拍摄的教师授课场景图像上。

3) 用于课堂评价的学生注意力可视化并行分析系统，4 个线程可以获得 2.37 倍加速效果。

4) 实现了 3 种典型学习状态（专注、关注、漠视）的学生注意力可视化分析。

2 研究背景

2.1 本研究基于的标准教室

在图 1 所示的标准教室中，教师在 4 m 左右长度的黑板前、对着教室里 30 名左右的学生进行授课。在本文研究中，在 F 点安装一台高清摄像头，称为前摄像头，主要对学生的听课注意力进行分析；在 B 点也安装一台高清摄像头，称为后摄像头，主要对教师的授课行为进行分析。

2.2 课堂教学评价对于学习注意力分析要素

通过视线与表情的组会，可以对学生的学习状态进行评价，具体有如下几种：1) 专注是学生将视线盯着教师，以期与教师实现眼神互动交流；2) 思考是指学生将目光放在教师或黑板区域，同时出现皱眉的表情；3) 困惑是指学生出现皱眉的表情，但不时将视线从黑板区域挪开；4) 好奇是学生将视线盯着教师，同时出现惊奇的表情。

2.3 基于单摄像头的低成本注意力分析方法

单摄像头方案使用屏幕自带或外接的单摄像头对头部姿态和眼球姿态进行估计，以实现固定式的廉价注意力分析。几乎不需要经费投入，且可以通过综合检测眼球姿态与头部姿态的方法来计算人眼注视方向、以确定被注视的屏幕区域[10]。

3 基于头部姿态估计的学生注意力可视化

3.1 基于头部姿态估计的学生注意力可视化方法

综合眼动仪和单摄像头注意力分析系统的优点，本文提出一种基于单幅图像头部姿态估计的学生注意力可视化分析方法，构建了一套相应的学生注意力可视化分析系统。如图 1 所示，本文方法先利用安装在黑板上方中间位置的前摄像头对学生听课情况进行录像，再采用图 2 所示方法对学生头部姿态进行估计，最后利用数学推导、将学生视线投射到后摄像头录制的教师授课视频上。

如图 2 所示，本文方法主要由以下的 6 步操作组成。

1) 数据（视频帧）获取：通过微软 1080p 的 LifeCam 摄像头采集课堂教学视频，并离散出视

频帧。

2) 摄像机标定：为了提高头部姿态识别的精度，需要采用方便精确的标定方法对摄像机参数进行标定。

3) 人脸检测 :使用于仕琪教授公开的人脸检测器[11]从视频帧中检测人脸。

4) 人脸特征点检测 :本文使用随机级联回归树获得 19 个人脸特征点的坐标信息，用于提供 PnP 求解中的二维信息。

5) 头部姿态估计 :基于一个统计测量获取的标准人脸模型[7]，通过求解 PnP 获得 2D/3D 之间的映射关系，输出头部姿态的旋转与平移矩阵。

6) 学生视点定位 :根据头部姿态的旋转平移矩阵信息，再利用空间坐标的转换关系将学生的视点投射到后摄像头拍摄的教师授课视频中，以实现学生学习注意力的可视化显示。

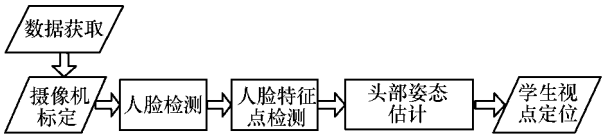


图 2　基于单幅图像头部姿态估计的学生注意力可视化分析的工作原理与算法流程

由于采用现成的处理软件进行视频帧获取与人脸检测[11]，所以下面内容将不对这两个步操作进行论述，仅对本文方法用到的摄像头标定、人脸特征点定位、头部姿态估计、学生视点定位等四步操作所用到的具体方法依次进行详细论述。

3.2　摄像头内部参数标定

用式(1)表述二维和三维之间的映射关系。

$$\boldsymbol{m}=\lambda\boldsymbol{C}[\boldsymbol{R}\ \ \boldsymbol{t}]\boldsymbol{M} \tag{1}$$

其中，λ 为比例因子，$\boldsymbol{m}$ 表示图像坐标系中特征点的坐标矩阵，$\boldsymbol{C}$ 表示摄像机的内参数矩阵，$\boldsymbol{R}$ 表示旋转矩阵，$\boldsymbol{t}$ 表示平移矩阵，$\boldsymbol{M}$ 表示目标坐标系中人脸特征点的三维坐标矩阵。式(1)可以进一步变换为如下形式。

$$\begin{bmatrix}u\\v\\1\end{bmatrix}=\lambda\begin{bmatrix}f_x&0&c_x\\0&f_y&c_y\\0&0&1\end{bmatrix}\begin{bmatrix}R_{11}&R_{12}&R_{13}&t_1\\R_{21}&R_{22}&R_{23}&t_2\\R_{31}&R_{32}&R_{33}&t_3\end{bmatrix}\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix} \tag{2}$$

其中，(u,v)是图像中的特征点的坐标，可以通过相应的人脸特征点检测算法得到；(X,Y,Z)是在目标坐标系中人脸特征点的三维坐标，可以直接从一个基于统计测量模型获得的三维标准脸来获取该点坐标；f_x、f_y、f_z、c_x、c_y 分别为摄像头的内部参数，可以通过对具体摄像头进行标定得到。

摄像机/摄像头一般会给出内部参数，参数的精确值可以通过对摄像头进行标定获得。文献[5]给出了一种工作原理简单且廉价的标定方法，可以省去几万或几十万造价的标定板的额外费用，其标定精度也基本满足本文工作的需求。

定义单应矩阵 $\boldsymbol{H}=\boldsymbol{C}[\boldsymbol{R}\ \ \boldsymbol{t}]$，可以将式(1)简化为

$$\boldsymbol{m}=\lambda\boldsymbol{H}\boldsymbol{M} \tag{3}$$

为了通过图像中的特征点和三维空间中对应的特征点的映射关系来求出 $\boldsymbol{H}$，可以将式(3)写成如下向量形式。

$$\begin{bmatrix}u\\v\\1\end{bmatrix}=\lambda\boldsymbol{C}(\boldsymbol{r}_1\ \ \boldsymbol{r}_2\ \ \boldsymbol{r}_3\ \ \boldsymbol{t})\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix} \tag{4}$$

其中，$\boldsymbol{r}_1$、$\boldsymbol{r}_2$、$\boldsymbol{r}_3$ 分别为矩阵 $\boldsymbol{R}$ 的列向量。因为本文标定选用的是平面标定板，所以 $Z=0$，式(4)可化简为

$$\begin{bmatrix}u\\v\\1\end{bmatrix}=\lambda\boldsymbol{C}(\boldsymbol{r}_1\ \ \boldsymbol{r}_2\ \ \boldsymbol{t})\begin{bmatrix}X\\Y\\1\end{bmatrix} \tag{5}$$

则矩阵 $\boldsymbol{H}$ 可以进一步表示为

$$\boldsymbol{H}=[h_1\ \ h_2\ \ h_3]=\lambda C[\boldsymbol{r}_1\ \ \boldsymbol{r}_2\ \ \boldsymbol{t}] \tag{6}$$

其中，$\boldsymbol{h}_1$、$\boldsymbol{h}_2$、$\boldsymbol{h}_3$ 为 $\boldsymbol{H}$ 的列向量。鉴于旋转向量在构造中是相互正交的，即具有旋转向量点积为 0 和向量长度相等这 2 个约束条件[5]。

$$\begin{aligned}&\boldsymbol{r}_1^{\mathrm{T}}\boldsymbol{r}_2=0\\&\|\boldsymbol{r}_1\|=\|\boldsymbol{r}_2\|\end{aligned} \tag{7}$$

根据式(7)将式(6)变换为

$$\begin{aligned}&h_1^{\mathrm{T}}\boldsymbol{C}^{-\mathrm{T}}\boldsymbol{C}^{-1}h_2=0\\&h_1^{\mathrm{T}}\boldsymbol{C}^{-\mathrm{T}}\boldsymbol{C}^{-1}h_1=h_2^{\mathrm{T}}\boldsymbol{C}^{-\mathrm{T}}\boldsymbol{C}^{-1}h_2\end{aligned} \tag{8}$$

通过给定两张或两张以上的图片就可以将矩阵 $\boldsymbol{C}$ 求出，摄像机的标定工作进行完毕。

文献[5]给出了标定法，只用不同视场的两张 3×3(4 个对应点)的棋盘图像就能得到相机内部参数，考虑噪声和数值稳定性，本文在标定过程中使用的标定板是在 A4 纸上打印的 7×10 的棋盘图像，如图 3 所示。实际拍摄了 15 张不同角度的图片用

于标定，实验通过寻找棋盘格图像中的角点(如图 4 所示)，并将角点作为参照点对相机进行标定。

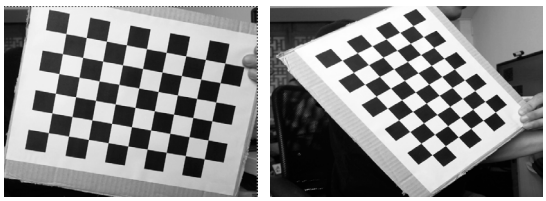


图 3 不同视角的标定板图像

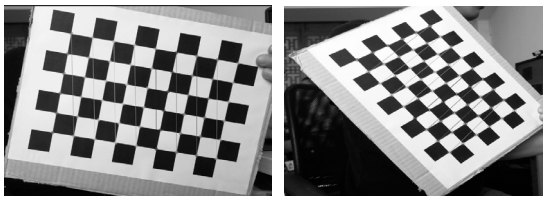


图 4 自动寻找角点

3.3 基于随机级联回归树的人脸特征点定位

随机级联回归树(R-CR-C)[6]是传统级联回归的改进方法。它是通过将级联结构并行设计，采用尺寸、形状自适应的数据更新方法和局部特征提取来提高人脸特征点检测的准确性。

3.3.1 级联回归

级联回归方法的目的是找到一个形状模型的更新方法，使更新的形状更接近真实形状。可以用式(9)对其进行形式化。

$$U: f(I, s_0) \to \delta_s,$$
$$\text{s.t.} \quad \left\| s_0 + \delta_s - \hat{s} \right\|_2^2 = 0 \tag{9}$$

其中，$\boldsymbol{U}$ 代表要训练的形状更新器，$\boldsymbol{I}$ 代表给定的人脸图像，s_0 代表初始化的形状，$f(\boldsymbol{I}, s_0)$是与形状特征有关的映射函数，δ_s 是形状要更新的量，$\hat{s}$ 是真实的形状。

在传统的级联回归方法[6,12-13]中,形状的更新是通过一个由 D 个弱分类器串联形成的强分类器来实现的，如式(10)所示。

$$R = r_1 \cdots r_D \tag{10}$$

其中，$r_d=\{\boldsymbol{A}_d,\boldsymbol{b}_d\}$ $(d=1\cdots D)$，$\boldsymbol{A}_d$是投影矩阵，$\boldsymbol{b}_d$是第 d 个回归器的偏移量。$\boldsymbol{A}_d$、$\boldsymbol{b}_d$都是通过标定好的人脸图像训练学习得到。假设现在已经训练好了一个强回归器 $\boldsymbol{R}$，在检测阶段，运用第一个弱回归器更新当前的形状 s'_0得到新的形状 s'_1，然后将形状 s'_1作为第二个弱回归器的输入，如此下去，直到得到最终的形状估计 s'_D。具体而言，可以用式(11)来计算第 d 个形状 s'_d。

$$s'_d = s'_{d-1} + A_d \cdot f(I', s'_{d-1}) + b_d \tag{11}$$

其中，s'_{d-1}是回归器已经得到的第 $d-1$ 个形状。

3.3.2 自适应随机级联回归树

相对传统的级联回归自适应随机级联回归树改变了级联回归的结构设计，并且对给定的输入图像进行参数的自适应更新。

随机级联回归树定义为由宽度 W 的级联回归线组成，每个级联回归线的深度为 D 组成的回归树，图 5 给出了一棵 W=3 的回归树。

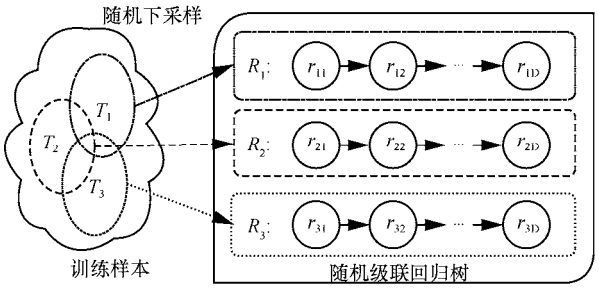


图 5 随机级联回归树

给定 N 张已标定的图片作为数据集 $\boldsymbol{T}=\{\boldsymbol{I}_1,\cdots,\boldsymbol{I}_N\}$，通过随机下采样生成 W 个子训练样本集 $\{\boldsymbol{T}_1,\cdots,\boldsymbol{T}_W\}$。整个回归树可以表示为 $\boldsymbol{U}=\{\boldsymbol{R}_1,\boldsymbol{R}_2,\cdots,\boldsymbol{R}_W\}$，第 w 个子训练集用于第 w 个级联回归线的训练, 第 w 个级联回归线包含 D 个的弱回归器可以表示为 $\boldsymbol{R}=\boldsymbol{r}_{w,1}\ldots\boldsymbol{r}_{w,D}$。

训练弱回归器需要获得初始的形状特征和初始形状与真实形状之间的差异。形状特征表示为所有人脸标记点在固定领域内的局部特征描述子的串联。由于所给图像的大小不同，如果特征领域的大小固定，则会造成计算得到的特征相差很大，具体如图 6 所示，所以必须采用式(12)给出的自适应方案、来调节不同图片大小所产生的特征邻域的大小变化。

$$S_p(d) = \frac{S_f}{K(1+e^{d-D})} \tag{12}$$

其中，$S_p(d)$为第 d 个弱回归器中的特征点邻域大小，K 为固定的缩放值，S_f为通过之前更新的人脸大小 s_{d-1} 估计出来的值，S_f 可以用瞳孔之间的距离、两外眼角之间的中点与两外嘴角之间的中点的距离最大值来表示。

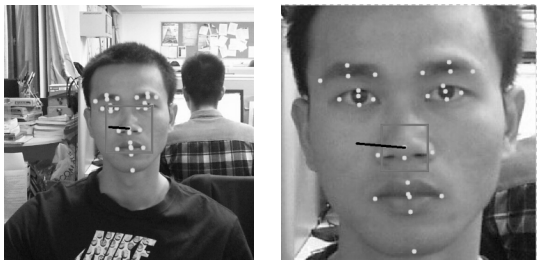


图 6　自适应的面积与形状更新

初始形状和真实形状的差异也和人脸大小有很大关系，如在不同人脸大小的情况下，当在形状初始化时都将鼻尖特征点定在左脸颊，形状更新的大小相差会很大，如图 6 中两黑线所示，所以对于第 w 条级联回归线，可以用式(13)来替代式(11)，以进行形状自适应更新。

$$s'_{w,d}=s'_{w,d-1}+S_f(s'_{w,d-1})(A_{w,d}f(I',s'_{w,d-1})+b_{w,d}) \quad (13)$$

其中，$s'_{w,d}$ 表示第 w 条级联回归线中第 d 个弱分类器得到的估计形状，$s'_{w,d-1}$ 表示第 w 条级联回归线中第 d–1 个弱分类器得到的估计形状，$\boldsymbol{A}_{w,d}$、$\boldsymbol{b}_{w,d}$ 分别表示第 w 条级联回归线中第 d 个弱分类器中的投影矩阵和平移量，$S_f(s'_{w,d-1})$大小与 $s'_{w,d-1}$ 有关。

3.4　基于 PnP 问题求解的头部姿态估计

3.4.1　PnP 问题

1989 年 Horaud 等[14]给出了位姿估计的 PnP 问题定义，求解 PnP 问题就是求解式(1)中的旋转矩阵 $\boldsymbol{R}$ 和平移矩阵 $\boldsymbol{t}$，而用于求解 $\boldsymbol{R}$ 和 $\boldsymbol{t}$ 的已知参量则包括：通过标定获得的摄像机内参数矩阵 $\boldsymbol{C}$、通过随机级联回归树检测得到的图像坐标系中特征点坐标矩阵 $\boldsymbol{m}$、通过一个已知统计模型标定得到的目标坐标系中人脸特征点的三维坐标矩阵 $\boldsymbol{M}$。本文用 Lepetit 等[15]在 2009 年提出的 Epnp 算法进行 PnP 求解。

3.4.2　用于头部姿态估计的 PnP 问题欧拉角求解

为更直观的表示头部姿态，需进一步计算头部在三维空间中的位置信息和旋转信息。式(1)中矩阵 $\boldsymbol{t}$ 中的 3 个元素直接表示头部姿态中的 X/Y/Z 位置信息，而头部姿态的旋转信息可以通过欧拉角来表示，要得到欧拉角，需要通过旋转矩阵求得。本文采用 RQ 分解得到 Givens 旋转从而得到欧拉角，一个三维 Givens 旋转是绕 3 个坐标轴中的一个轴进行旋转，3 个 Givens 旋转可以分别表示为

$$\boldsymbol{R}_x(\psi)=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\psi & -\sin\psi\\ 0 & \sin\psi & \cos\psi\end{bmatrix} \quad (14)$$

$$\boldsymbol{R}_y(\theta)=\begin{bmatrix}\cos\theta & 0 & \sin\theta\\ 0 & 1 & 0\\ -\sin\theta & 0 & \cos\theta\end{bmatrix} \quad (15)$$

$$\boldsymbol{R}_z(\varphi)=\begin{bmatrix}\cos\varphi & -\sin\varphi & 0\\ \sin\varphi & \cos\varphi & 0\\ 0 & 0 & 1\end{bmatrix} \quad (16)$$

其中，Ψ、θ、Φ 分别表示绕 X、Y、Z 轴旋转的角度。

RQ 算法将旋转矩阵 $\boldsymbol{R}$ 分解为 $\boldsymbol{R}_x(\Psi)$、$\boldsymbol{R}_y(\theta)$、$\boldsymbol{R}_z(\Phi)$这 3 个矩阵的转置与一个上三角矩阵 $\boldsymbol{A}$ 的乘积，即 $\boldsymbol{R}=\boldsymbol{A}\boldsymbol{R}_z^{\mathrm{T}}\boldsymbol{R}_y^{\mathrm{T}}\boldsymbol{R}_x^{\mathrm{T}}$。得到 3 个 Givens 矩阵后，就可以通过代数方法很简单地求解出 3 个旋转角 Ψ、θ、Φ，即欧拉角的值。

3.5　基于头部姿态的学生注意力可视化

基于头部姿态的学生注意力可视化就是将双眼中间点的面部法线投射到后摄像头拍摄的教师授课视频上，如图 7 所示，必须先计算出坐标系 $O_SX_SY_SZ_S$ 中的射线 O_SZ_S 在平面 $O_FX_FY_F$ 上的交点 I，再计算出点 I 在坐标系 $O_BX_BY_BZ_B$ 内的映射点 I'坐标。

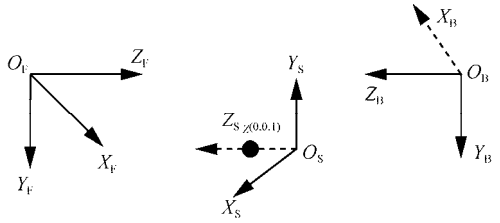


图 7　用于学生注意力可视化的 3 个局部坐标系

前摄像头 F 的局部坐标系 $O_FX_FY_FZ_F$ (其中，坐标轴 X_F 垂直朝向学生的方向为正方向)、被试同学 S 头部以双眼中间点为原点的局部坐标系 $O_SX_SY_SZ_S$ (约定当 Y_S 轴垂直地面，Z_S 轴垂直黑板时的方向为头部初始方向，也即头部未旋转时的方向)、以及后摄像头 B 的局部坐标系 $O_BX_BY_BZ_B$ (X_B 垂直朝向学生的方向为正方向)。

当学生 S 头部旋转平移后，即坐标系 $O_SX_SY_SZ_S$ 相对于 $O_FX_FY_FZ_F$ 坐标系发生了旋转平移，可以获得如下所示的头部旋转矩阵 $\boldsymbol{R}$ 和平移矩阵 $\boldsymbol{t}$

$$\boldsymbol{R}=\begin{bmatrix}R_{11} & R_{12} & R_{13}\\ R_{21} & R_{22} & R_{23}\\ R_{31} & R_{32} & R_{33}\end{bmatrix} \quad (17)$$

$$\boldsymbol{t}=\begin{bmatrix}t_1\\ t_2\\ t_3\end{bmatrix} \quad (18)$$

并假设在 $O_SX_SY_SZ_S$ 坐标系中的有一点 z，其坐标为(0，0，1)，具体如图 7 所示。根据式 (19)所给出的映射关系可以将旋转后的 z 点坐标和原点坐标 O_S，在 $O_FX_FY_FZ_F$ 坐标系下表示出来，分别表示为 $O_S(t_1，t_2，t_3)$和 $z(R_{13}+t_1，R_{23}+t_2，R_{33}+t_3)$。

$$\begin{bmatrix} X_F \\ Y_F \\ Z_F \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & t_1 \\ R_{21} & R_{22} & R_{23} & t_2 \\ R_{31} & R_{32} & R_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_S \\ Y_S \\ Z_S \\ 1 \end{bmatrix} \tag{19}$$

最后，通过相似三角形对应边成比例得到学生的视点在 $O_FX_FY_FZ_F$ 坐标系下的表示，具体计算过程如式(20)和式(21)所示。

$$\begin{cases} \dfrac{t_3-(R_{33}+t_3)}{t_3} = \dfrac{(R_{13}+t_1)-t_1}{\Delta x} \\ \dfrac{t_3-(R_{33}+t_3)}{t_3} = \dfrac{t_2-(R_{23}+t_2)}{\Delta y} \end{cases} \tag{20}$$

$$\begin{cases} x = t_1+\Delta x = t_1-\dfrac{R_{13}}{R_{33}}t_3 \\ y = t_2-\Delta y = t_2-\dfrac{R_{23}}{R_{33}}t_3 \end{cases} \tag{21}$$

以上变换只是将学生视点 I 的位置在坐标系 $O_FX_FY_FZ_F$ 内表示出来，但是为了对视点 I 进行可视化，还需要借助摄像头 B 拍摄的图像来显示。如下图 8 所示，该图既可以表示在坐标系 $O_BX_BY_BZ_B$ 的实际空间中视点的位置信息，也可表示在摄像头 B 拍摄的图像坐标系中视点的位置信息，而 I 视点在图像中的位置信息就是本文需要求解的值。

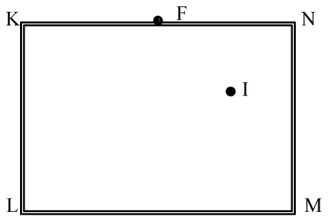


图 8 视点 I 的位置信息，点 F 为前摄像头，点 K/L/M/N 分别为黑板平面的 4 个角

本文通过测量可以获得图 8 中的点 K/L/M/N 物理坐标(x_K,y_K)、(x_L,y_L)、(x_M,y_M)、(x_N,y_N)，同时通过手工标定可以获得点 K/L/M/N 在摄像头 B 拍摄的图像坐标系中的映射点 K’/L’/M’/N’坐标$(x_{K'},y_{K'})$、$(x_{L'},y_{L'})$、$(x_{M'},y_{M'})$、$(x_{N'},y_{N'})$，视点 I(x_I,y_I)和其在图像坐标系中的映射点 I’ $(x_{I'},y_{I'})$之间必然存在如式(22)线性插值关系。

$$\begin{cases} \dfrac{x_K-x_I}{x_K-x_N} = \dfrac{x_I'-x_K'}{x_N'-x_K'} \\ \dfrac{y_I-y_K}{y_L-y_K} = \dfrac{y_I'-y_K'}{y_F'-y_K'} \end{cases} \tag{22}$$

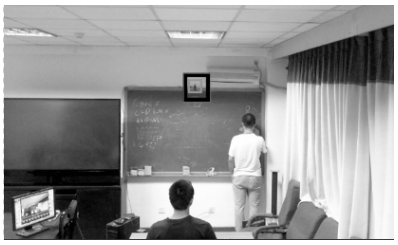
对式(22)进行变换后，可以得到如式(23)所示的 I’ $(x_{I'},y_{I'})$求解方程。

$$\begin{cases} x_I' = \dfrac{x_K-x_I}{x_K-x_N}(x_N'-x_K')+x_K' \\ y_I' = \dfrac{y_I-y_K}{y_L-y_K}(y_F'-y_K')+y_K' \end{cases} \tag{23}$$

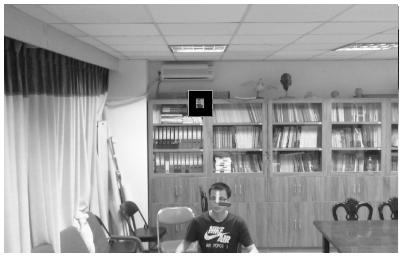
4 实验结果

4.1 学生学习注意力可视化方法的实验设置及其可视化效果

如图 9 所示，本文将一间会议室改造为一口小教室。在图 9(a)中，被试 S 坐在黑板前方 3 m 处，看着正在书写版书的教师 T，黑板正上方安装了前摄像头 F（图中黑色正方形处），对 S 的学习状况进行监控。在图 9(b)中，后摄像头 B（图中黑色正方形处）安装在 S 的后方，对 T 的授课状况进行监控。为了保证测试精度，在安装过程中，需对 F 和 B 摄像头进行了标定。



(a) 可视化的视线投射点及其前摄像头



(b) 可视化的头部姿态及其后摄像头

图 9 本文方法的实验设置与显示效果

如图 9(b)所示，基于 F 拍摄的视频帧图像，利用本文提出的单幅图像 PnP 头部姿态估计方法、计

算出 S 头部姿态的三维角度信息和三维坐标轴（共 6 个维度信息），其中，为了便于显示，6 维信息被标在鼻尖点。利用 3.5 节推导出的基于头部姿态的学生注意力可视化方法、可以计算出 S 目光注视点的物理位置，并标注到 B 拍摄的视频帧图像上，图 9(a)中的就是 S 目光的注视点。

4.2　本文方法的头部姿态估计精度测试

用 Biwi Kinect 数据库[16]对基于单幅图像的 PnP 头部姿态估计方法进行精度测试。Biwi Kinect 数据库给出了 20 个人(14 个男性，6 个女性)头部转动不同方向的深度数据和 RGB 图像，共 24 段视频帧序列(少数人录制了两次)。所有图像中的头部位置和旋转角度都进行了标定，平移和旋转的标定误差分别为 1 mm 和 1°，旋转角的范围为：俯仰角(pitch)±60°、偏航角(yaw)±75°，旋转角(roll)±50°。由于学生学习注意力分析研究主要关注学生是否将视线投向前方的黑板区域，如图 1 所示，本文只关心特定范围内的学生头部姿态，所以本文实验的头部旋转角范围缩小为：俯仰角(pitch)±30°、偏航角(yaw)±45°，旋转角(roll)±50°。

采用本文提出的头部姿态估计方法（R-CR-C + EPnP）、对 Biwi 库所有符合注意力分析需求的图片进行头部姿态旋转角计算，以获取俯仰角(pitch)、偏航角(yaw)、旋转角(roll)的估计值，将估计值与标注值之间差值的绝对值作为估计值的误差；对所有图片的头部姿态旋转角估计值的误差进行平均、可以获得本文方法的平均误差。如表 1 所示，本文方法所获得的俯仰角(pitch)、偏航角(yaw)、旋转角(roll)估计值的平均误差分别为 7.11°、5.02°、2.51°，这 3 个旋转参量总的平均误差为 4.88°, 明显小于已有的头部姿态估计方法[16-19]（5.91°~25.21°），表明本文提出的头部姿态估计方法（R-CR-C + EPnP）具有较高的测试精度。

为了更直观、更全面地评估本文提出的头部姿态估计方法的精度，本文使用对 Biwi 库中第一段视频帧序列共 387 张图片进行头部姿态 6 个维度参量估计，并计算出每张图片估计值对于标定值的偏差, 所有计算结果如图 10 所示。如图 10(a)~图 10(c)所示，本文方法对于头部姿态在 X/Y/Z 方向的位移偏差均在可接受范围内，x 轴和 y 轴方向位移估算值最大误差分别小于 40 mm 和 50 mm, z 轴方向位移估算值最大误差小于 230 mm。如图 10(d)~图 10 (f)所示，本文方法对于头部姿态旋转角估计较准，基本上能够和标定值曲线相吻合，偏差较小。

表 1　与现有头部姿态估计方法的平均误差比较

方法	Pitch/°	Yaw/°	Roll/°	Mean/°
random forests[16]	8.5	9.2	8.0	8.6
CLM[17]	18.30	28.30	28.49	25.21
CLM-Z[18]	12.03	14.80	23.26	16.69
CLM with GAVAM[18]	5.10	6.29	11.29	7.56
RF-TR-D[19]	5.15	7.8	4.8	5.91
R-CR-C + EPnP	7.11	5.02	2.51	4.88

4.3　本文方法的注意力可视化分析效率测试

对如图 9 所示的前/后摄像头所拍 1080P 高清视频进行单人注意力可视化分析，视频时长为 2 min。同时为了对本文注意力可视化分析方法的并行加速性能进行测试，采用 i5-4570(4 核)CPU 的 1~4 个物理线程对两段视频进行串/并行处理, 实验系统采用 32 GB 内存和 TitanX 显卡（12 GB 显存），以保证内存和显卡不成为硬件系统的性能瓶颈。

先对两段时长为 2 min 的视频进行帧图像读取、人脸检测、人脸特征点检测、基于 PnP 的头部姿态估计、注意力可视化等 5 步操作，分别获取各步的运行时间和整个算法总的运行时间。本文先用单线程来获取串行计算的运行时间；以此为基准，再分别使用 2~4 个线程获取并行计算的运行时间，以观察并行加速的效果。从表 2 中列出的所有数据可以获得如下结论。

1) 当使用一个线程进行学生学习注意力可视化分析串行处理时，对于 1080P 单人脸视频的处理速度很慢, 处理一帧图像需要 738.22 ms, 即一秒钟只能处理 1.39 帧。

2) 对学生学习注意力可视化分析串行处理的各步运行时间进行比较，可以发现人脸检测和人脸特征点检测这 2 步最耗时, 分别为 541.48 ms 和 122.39 ms,

表 2　本文方法的多线程平均加速效果及其各步操作的平均耗时

线程数目	读取帧/ms	人脸检测/ms	人脸特征点检测/ms	基于PnP头部姿态估计/ms	注意力可视化分析/m	总耗时/m	帧率(FPS)	加速比(X)
1	5.00	541.48	122.39	28.37	13.38	738.22	1.39	1.00
2	5.11	284.84	109.08	25.88	11.88	460.74	2.26	1.63
3	5.09	191.77	101.37	25.51	11.39	357.47	2.94	2.12
4	5.27	153.35	102.09	26.05	14.41	322.55	3.31	2.38

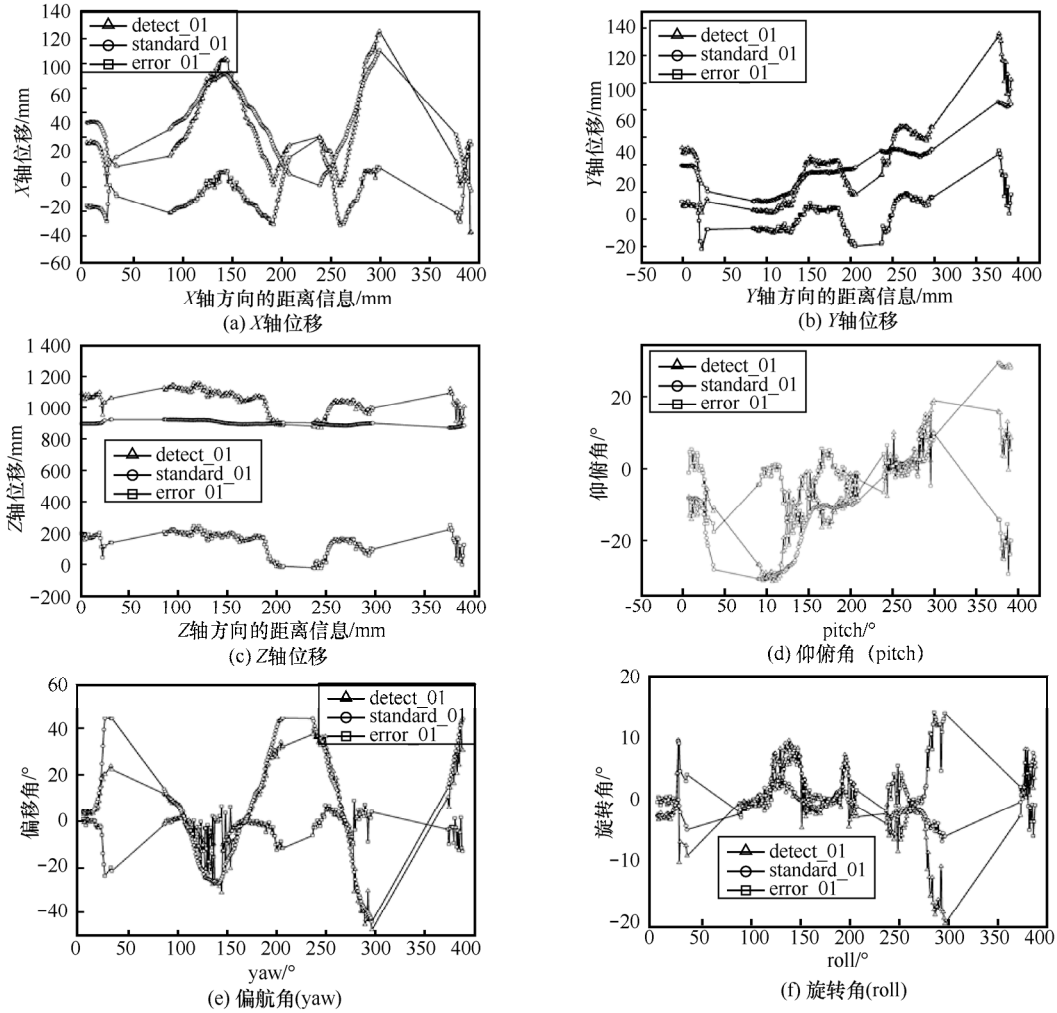


图 10 头部姿态 6 个维度参量估计结果的精度分析，detect 代表检测结果，standard 代表标准库的数据，error 代表偏差大小

占总运行时间的 72.37%和 16.36%（共 88.73%），所以提高注意力可视化分析速度的关键在于如何降低人脸检测和人脸特征点检测这 2 步的耗时。

3) 使用 2~4 个线程进行注意力可视化分析并行计算，发现并行计算可以有效降低人脸检测的耗时，但对于其他 4 步（读取帧、人脸特征点检测、头部姿态估计、注意力可视化）则加速效果不明显。鉴于人脸检测操作占整个算法耗时的主要部分（72.37%），使用 2~4 个线程分别取得 1.63 倍、2.12 倍、2.38 倍的加速效果，这表明本文算法具有很好地并行加速性能。

以上分析可知, 1) 对于 1080P 视频中的单人注意力分析都非常耗时，即使采用装有目前最好显卡（TitanX）的单台计算机、进行多线程并行计算，只获得 3.31 帧/秒的处理速度，难于实现单人注意力分析的实时处理。2) 为了在今后实际应用中、使用 4 KB 高清摄像机对课堂教学中 30~50 名学生进行实时注意力分析，必须采用具有众多节点机的计算集群进行大规模并行处理，本文方法使用单幅图像进行 PnP 头部姿态估计，将完全适合进行众多节点机的并行处理。

4.4 典型学习场景的可视化分析

在一个典型场景对课堂教学过程中 3 种典型的学生学习状态（专注、关注、漠视）进行可视化分析。在实验中教师与学生的姿态尽量保持不变，为了便于显示可视化的效果，本文取后摄像头拍摄的其中一帧教师授课图像作为学生视点的投射背景，而将学生视点（黑框中）序列投射到教师的授课背景图像上，具体可视化效果如图 11 所示。

如图 11(a)所示，学生处于专注的学习状态，目光聚焦于教师，基于头部姿态的学生视点（黑框中）密集分布在教师的头部附近。如图 11(b)所示，学生

处于关注的黑板的一个区域，基于头部姿态的学生视点（黑框中）密集分布在该区域。如图 11(c)所示，学生处于漠视的学习状态，从学生的头部姿态可见其正在看着黑板外（黑板右上角外边）的一个区域，基于头部姿态的学生视点（黑框中）密集分布在该区域。

从图 11 的 3 个可视化结果可见：本文方法较为精确地给出学生的视点分布，能够对学生学习注意力进行较为精确地分析，从原理上证明了本文方法可以用于课堂教学的学生学习注意力分析。

(a)专注：学生目光聚焦教师、与教师进行眼神交流

(b)关注：学生目光看着黑板内区域

(c) 漠视：学生目光游离出黑板区域

图 11　3 种典型学生学习状态的注意力可视化分析

5　结束语

基于计算机视觉、计算机图形学、高性能并行计算技术，本文提出了一种基于单幅图像头部姿态识别的学生注意力可视化分析方法。基于已有的 Biwi 库对本文方法的头部姿态估计精度进行了验证，还对本文方法的注意力可视化分析结果进行验证，都获得了满意的结果，从原理上表明了本文方法能够用于课堂教学中的学生学习注意力分析。此外，本文进行的学生注意力可视化并行分析实验表明：本文方法具有较好的并行加速效果。

本文实验系统仅能够对单人进行注意力可视化分析，为了对课堂内 30~50 名学生进行注意力可视化实时分析，下一步需开展基于大规模计算集群与 4 KB 高清视频的头部姿态估计并行算法研究。

参考文献：

[1] SCOTT V.Clicking in the classroom: using a student response system in an elementary classroom[J].New Horizons for Learning, 2014, 11(1):21-23

[2] WYDER S, CATTIN P C. Eye tracker accuracy: quantitative evaluation of the invisible eye center location[P].International journal of computer assisted radiology and surgery, 2018,13(10): 1651-1660.

[3] PORTA M, RICOTTI S, PEREZ C J.Emotional e-learning through eye tracking[C]//2012 IEEE Global Engineering Education Conference (EDUCON), 2012: 1-6.

[4] ERIK M, MOHAN M.Head pose estimation in computer vision: a survey[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(4):607-626

[5] ZHANG Z.A flexible new technique for camera calibration[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11):1330-1334

[6] FENG Z H, HUBER P, KITTLER J, et al.Random cascaded-regression copse for robust facial landmark detection[J].IEEE Signal Processing Letters, 2015, 1(22):76-80

[7] XIONG X, TORRE F.Supervised descent method and its applications to face alignment[C]//The IEEE Conference on Computer Vision and Pattern Recognition, 2013: 532-539.

[8] FISCHLER M A, BOLLES R C.Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J].Communications of the ACM, 1981, 24(6):381-395

[9] HORAUD R, CONIO B, LEBOULLEUX O, et al.An analytic solution for the perspective 4-point problem[C]// Computer Vision and Pattern Recognition, 1989: 500-507.

[10] VALENTI R, SEBE N, GEVERS T.Combining head pose and eye location information for gaze estimation[J].Image Processing IEEE Transactions on, 2012, 21(2):802-815

[11] WU S Y, YU S Q, CHEN W S, et al. An Implement of Image Statiching based on Binocular Cameras//Proceedings of 2011 Third Chinese Conference on Intelligent Visual Surveillance, 2011: 96-99。

[12] DOLLÁR P, WELINDER P, PERONA P.Cascaded pose regression[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010: 1078-1085.

[13] BURGOS A X, PERONA P, DOLLÁR P.Robust face landmark estimation under occlusion[C]//the IEEE International Conference on Computer Vision. 2013: 1513-1520.

[14] XU D, TAN M, JIANG Z, et al.A shape constraint based visual positioning method for a humanoid robot[J].Robotica, 2006, 24(04): 429-431.

[15] LEPETIT V, MORENO N F, FUA P. Epnp: an accurate $o(n)$ solution to the PnP problem[J].International Journal of Computer Vision, 2009, 81(2):155-166

[16] FANELLI G, WEISE T, GALL J, et al. Real time head pose estimation from consumer depth cameras[M]. Pattern Recognition. Springer Berlin Heidelberg, 2011: 101-110.

[17] SARAGIH J M, LUCEY S, COHN J F. Deformable model fitting by regularized landmark mean-shift[J].International Journal of Computer Vision, 2011, 91(2):200-215

[18] BALTRUAITIS T, ROBINSON P, MORENCY L P.3D constrained local model for rigid and non-rigid facial tracking[C]//Computer Vision and Pattern Recognition (CVPR), 2012: 2610-2617.

[19] KAYMAK S, PATRAS I. Multimodal random forest based tensor regression[J].IET Computer Vision, 2014, 8(6):650-657

[作者简介]

陈平（1974–），男，四川达州人，博士，北京师范大学高级工程师，主要研究方向为主要研究方向为数据挖掘、模式识别。

皇甫大鹏（1982–），男，江苏徐州人，北京师范大学工程师，主要研究方向为数据挖掘、大数据、无线网络。

骆祖莹（1968–）,男，江苏连云港人，北京师范大学教授，主要研究方向为课堂教学自动评价、情感计算。

李东兴（1984–），男，山东无棣人，北京师范大学博士生，主要研究方向为计算机视觉、自然语言处理和课堂教学自动评价。