

DOI: 10.19650/j.cnki.cjsi.J2108085

基于机器学习的头部自由视线追踪方法 及其在电动病床端的应用*

胡佳辉¹, 陆永华¹, 张进海², 杨昊铮¹, 刘 韬¹

(1. 南京航空航天大学机电学院 南京 210016; 2. 成都飞机工业(集团)有限责任公司 成都 610073)

摘要:头部自由的三维视线追踪具有重要的意义。针对传统的视线追踪方法存在的精度低、设备复杂和穿戴受限等问题,提出了一种基于机器学习的单目头部自由的三维视线追踪技术,构建出了两个轻量、高精度、实时的视线追踪模型,分别可进行注视点估计和注视方向估计。注视点估计模型首先利用 Dlib 进行人脸特征点定位得到眼睛图像进而进行 PnP 解算得到头部姿态信息,将两种信息与部分特征点坐标共同输入到多通道卷积神经网络中,并最终估计得到人眼注视点。注视方向估计网络是注视点估计网络的简化版本。将本文所提出的视线追踪技术与电动病床相结合,构建出一套基于眼球驱动的电动病床系统,该系统允许患者利用眼睛控制电动病床的运作。实验结果表明:所提出的注视点模型在 MPIIGaze 数据集上的误差为 4.1 cm,注视方向估计网络在 ColumbiaGaze 数据集上的误差为 7.2°,两模型的精度分别比 iTracker 和 UlinFT 提高了 6.8% 和 2.7%。所构建的眼球驱动电动病床系统有效地提高了患者的生活水平,满足了患者的诸多需求。

关键词: 视线追踪; 眼球驱动; 机器学习; 智能病床

中图分类号: TP183 TH772 **文献标识码:** A **国家标准学科分类代码:** 520.2040

Machine learning-based free-head gaze tracking method and its application on the electric sickbed

Hu Jiahui¹, Lu Yonghua¹, Zhang Jinhai², Yang Haozheng¹, Liu Tao¹

(1. College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; 2. Chengdu Aircraft Industry Co., Ltd., Chengdu 610073, China)

Abstract: The free-head 3D gaze tracking is of great significance. The traditional eye tracking methods have problems of low accuracy, complex equipment and limited wearing. To address these issues, a monocular head free 3D eye tracking technology based on machine learning is proposed. Two lightweight, high-precision and real-time eye tracking models are formulated, which can estimate the gaze point and gaze direction, respectively. For the gaze point estimation model, Dlib is used to locate the facial feature points to get the eye image. Then, PNP is used to get the head pose. Two kinds of information and part of the feature point coordinates are taken as the input into the multi-channel convolutional neural network. Finally, the gaze point is estimated. The gaze direction estimation network is a simplified version of gaze point estimation network. The proposed eye tracking technology is combined with the electric sickbed to establish a set of electric sickbed system based on eyeball drive, which allows patients to use their eyes to control the operation of the electric sickbed. Experimental results show that the error of the proposed gaze point estimate model on MPIIGaze dataset is 4.1 cm. The error of gaze direction estimation network in ColumbiaGaze data set is 7.2°, and the accuracy of the two models is 6.8% and 2.7% higher than those of iTracker and UlinFT, respectively. The system of eye driven electric sickbed can improve the living standard of patients and meet the needs of patients.

Keywords: eye tracking; eyeball drive; machine learning; intelligent sickbed

收稿日期: 2021-06-13 Received Date: 2021-06-13

* 基金项目: 国家自然科学基金(51975293)、航空科学基金(2019ZD052010)项目资助

0 引言

人类 80% 的感官信息是由眼睛接收的。注视是人类行为的重要表示, 注视方向和注视行为的改变是视觉注意力的良好指标, 并且还与人内心思想或心理状态有关。因此, 在智能系统中, 跟踪人类注视方向的能力是必不可少的。长期以来人们一直认为视线跟踪器可以作为传统计算机输入设备的一个很好的替代品, 它已经在许多领域得到应用, 包括人机交互, 虚拟现实, 社交分析或医疗保健等。

在最近几年中, 已有多家供应商提供了专用视线追踪设备, 且部分设备具有良好的实时性和准确性。但许多高性能设备非常昂贵, 且存在设备复杂、眼镜不适配、穿戴不便等问题, 因此, 有大量工作提出了基于计算机视觉的准确视线追踪技术。这些工作的思路之一是基于外观的方法 (appearance-based methods), 该方法使用眼睛图像以及头部姿态信息来预测人的注视方向且受到了广泛的关注。

一般来说, 基于外观的三维视线追踪可以被表述为一种回归问题, 该问题可通过多种方法得以解决, 其中, 利用神经网络的方法被证明是有效的^[1]。该方法将眼睛图像和头部信息输入到深度卷积神经网络中并在网络的最后一层编码三维凝视向量。因此, 有了足够的训练数据, 就可以训练网络的参数。然而, 即使利用深度神经网络 (deep neural network, DNN) 做回归分析, 在真正测试时其准确性通常会被限制在 $6^\circ \sim 10^\circ$ 左右, 且具有很高的个体间方差, 训练过的模型对于一个新的对象或者在一个新的环境中可能会完全失败。这是由许多因素导致的, 包括训练和标定数据稀少、人眼外观以及人眼解剖构造的差异、头部姿态的引入使模型复杂化^[2]。另外, 头部运动的不受限对于视线追踪的推广至关重要, 通过固定头部来提高预测精度的视线跟踪器往往在现实中的应用面十分狭窄。

随着机器学习和深度学习的发展, 通过建立端到端的视线追踪模型成为近年来的研究热点。Mora 等^[3]提出一种三维视线数据采集方法, 用以获取高精度的视线方向数据。Wood 等^[4]提出一种使用计算机图形学的理论来合成眼睛样本的方法, 为视线追踪模型提供虚拟的数据。然而, 将这些方法尽管能够获得较高的视线追踪精度, 但模型的泛化能力依旧不容乐观, 且复杂的数据采集过程限制了视线追踪的应用。Krafka 等^[5]提出一种注视点估计网络, 但对算力要求较高。Park 等^[6]提出一种基于小样本学习的个性化方法, 该方法在一种新颖的网络框架 FAZE 中利用不到 9 个标定样本即可实现较好的个体适应性。然而, 该方法所采用的编解码网络对算力

要求极高, 在移动端或普通嵌入式主板上根本无法得到应用。潘新星等^[7]提出一种通过卷积目标检测恢复目标区域深度信息的 3D 眼球追踪实现方法, 但主要用于穿戴设备和 VR 领域。迟健男等^[8]利用双相机实现了视线追踪, 但该方法无法在移动端应用。Zhang 等^[9]提出一种基于注意力机制的全脸视线估计方法。网络的输入是人脸图像, 并采用端到端的学习策略, 直接学习出相机坐标系下的视线方向。因此可以认为该方法隐式估计了头部姿态。该方法的优点是简单直接, 但内部机制不明, 并未利用头部姿态的良好解析公式, 因此其可解释性较差。针对该问题, Zhang 等^[10]又提出了一种基于特征拼接的方法, 该方法将头部姿态信息与提取出的眼睛特征拼接, 用以学习相机坐标系下的视线。针对头部自由的问题, Krafka 等^[5]提出了一种利用人脸网格来表征人脸空间位置的方法, 但该方法仍旧不允许用户进行头部姿态的变化。

从研究现状可以看出, 虽然部分研究成果能够利用机器学习的方法达到较高的视线追踪精度, 但都对算力有较高的要求且输入数据的形式较为复杂。为解决上述问题, 本文提出了两个全新的端到端卷积神经网络, 分别对注视点和注视方向进行高精度的估计。模型具有轻量、高精度和高实时性等特点。实验结果表明: 所提出的注视点模型在 MPIIFaceGaze 数据集上的误差为 4.1 cm, 注视方向估计网络在 ColumbiaGaze 数据集上的误差为 7.2° 。此外, 本文还提出了一种基于眼球驱动的电动病床系统, 采用本文所提出的注视点估计网络作为视线跟踪器, 该系统具有控制精度高、响应速度快的特点。

1 基于外观的视线追踪方法介绍

1.1 视线追踪问题与数学模型

基于外观的视线追踪问题一般是利用纯视觉的方法对人眼注视方向进行估计。注视点和注视方向由眼球运动和头部姿态组成, 如图 1 所示。图 1 分别讨论了注视点估计问题和注视方向估计问题。

视线方向 \mathbf{g} 可在头部坐标系下和相机坐标系下分别由 \mathbf{g}_h 和 \mathbf{g}_c 来表示。图 1 中, ϕ_g 表示视线方向和头部坐标系中水平面之间的夹角, θ_g 表示视线方向和头部坐标系中垂直平面之间的夹角。 R_{hc} 表示头部坐标系到相机坐标系的变换关系。因此, 视线在头部坐标系中的运动可以表示为:

$$\mathbf{g}_h = [-\cos(\phi_g) \sin(\theta_g), \sin(\phi_g), -\cos(\phi_g) \cos(\theta_g)]^T \quad (1)$$

头部坐标系和相机坐标系的变换关系与头部姿势和头部位置有关, 这里可将其表示为:

$$R_{hc} = f(h_a, h_b) \quad (2)$$

其中, h_a 表示头部姿态, h_b 表示头部位置。一般来说, 视线追踪的最终目标是在相机坐标系中获得人的视线方向, 因此必须将头部坐标系中的视线方向旋转转换为相机坐标系中的视线方向:

$$g_c = Rg_h \quad (3)$$

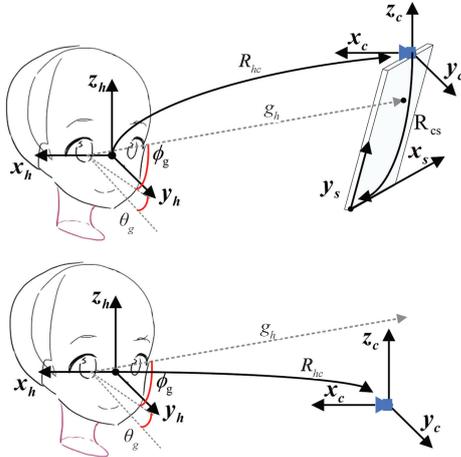


图 1 视线方向模型示意图

Fig. 1 Illustration of the gaze direction model

对于注视点估计问题, 估计结果应该是注视方向与屏幕在屏幕坐标系中的交点。在实际应用场景中, 相机坐标系和屏幕坐标系的位置一般是固定的, 转换关系可以用 R_{cs} 表示。因此, 注视点估计问题的数学模型可以表示为:

$$pos = p(R_{cs}R_{hc}g_h) \quad (4)$$

其中, pos 表示屏幕上的坐标点, 函数 $p()$ 表示三维注视方向到二维注视点的变换。

1.2 基于机器学习的视线追踪方法

在基于机器学习的视线追踪问题中, 基于外观的三维视线追踪可以被表述为一种回归^[11]:

$$f: x_i \in \mathbf{R}^p \rightarrow g_c \in \mathbf{R}^3 \quad (5)$$

式中: x_i 表示特征集, 包括图像特征和头部姿态估计, g_c 表示三维空间中的注视方向。基于外观的二维视线追踪可表述为另一种回归:

$$f: x_i \in \mathbf{R}^p \rightarrow pos \in \mathbf{R}^2 \quad (6)$$

对于回归问题, 可以利用卷积神经网络表示输入和输出之间的映射关系, 并在相应的数据集上进行学习, 并最终求解出一个较优的 f , 如图 2 所示。利用机器学习进行视线追踪的难点主要集中在模型构建方法、数据处理方式、个性化标定方法这 3 点。诸多基于机器学习的视线追踪研究成果通过限制头部运动来提高视线估计精度。本文提出了一种新型视线追踪网络架构, 该网络实现了基于单目相机的头部自由的三维视线追踪, 具体形式将在第 2 章中进行介绍。

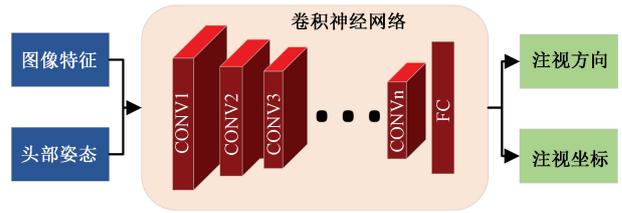


图 2 基于机器学习的视线追踪模型

Fig. 2 The eye tracking model based on machine learning

2 三维头部自由的视线追踪模型

本文提出了一种三维头部自由的注视点估计卷积神经网络模型, 相较于现有的视线追踪模型^[1,5-6,9,12-14], 其具有更高的精度、更低的算力要求和更快的运算速度, 且支持头部在一定的范围内自由地转动。模型架构如图 3 所示。

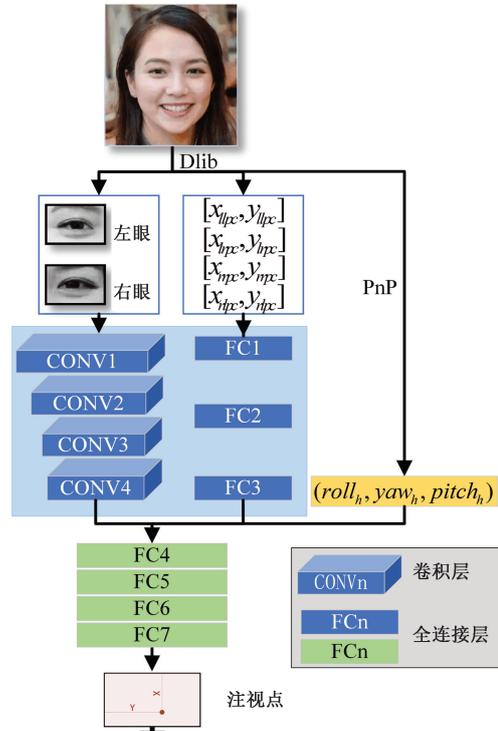


图 3 注视点估计网络示意图

Fig. 3 Illustration of the point estimation network

模型的输入为: 左右眼图像, 左眼的左右眼角坐标点 $[x_{llec}, y_{llec}]$ 、 $[x_{lrec}, y_{lrec}]$, 右眼的左右眼角坐标点 $[x_{rlec}, y_{rlec}]$ 、 $[x_{rrec}, y_{rrec}]$ 以及头部姿态 ($yaw_h, pitch_h$)。

该网络首先通过 Dlib 机器学习包进行人脸坐标定位和头部姿态估计, 在人脸坐标点的基础上, 截取出左右眼图像, 降低左右眼图像分辨率至 250×250 并转换为单通道灰度图像, 而后将两张灰度图叠加得到一张 $250 \times 250 \times 2$ 的图像, 将该双通道图像输入到卷积网络支

路中。在另一条支路中,将人脸坐标点定位得到的左右眼的左右眼角坐标输入至3层全卷积网络。对坐标点定位结果进行 PnP 解算得头部姿态,并将头部姿态和两条神经网络支路的运算结果汇合输入至全连接层 FC3 中,最终得到人眼注视点 $[x, y]$ 的估计。

为减少过拟合发生的概率,该模型并未将人脸图像加入到输入通道中,使模型只依赖于眼睛和眼睛的位置特征。与人脸图像相比,眼睛图像表现出更少的可变性,这使得使用较少的数据可以更好地进行拟合并改进泛化。同时考虑左右眼能够在一定程度上提高预测精度^[14],且为减少模型任务的复杂度,该模型将左右眼灰度图像相叠加构成一张双通道图像输入卷积神经网络中,而不是为左右眼图像单独构建权重共享的卷积神经网络,实验结果表明,该方法提高了模型预测精度且减少了模型大小。

该网络利用两眼左右眼角的坐标来表征头部与相机之间的距离,因此可以通过简单的标定来解决不同个体的眼睛大小差异性,但因为大部分个体之间的眼睛大小差异不大,即使将眼睛形状的差异考虑在内,模型也能够未标定的情况下进行中等精度的视线估计。

该网络的卷积层具体尺寸如下:CONV1($5 \times 5 \times 12$), CONV2($3 \times 3 \times 24$), CONV3($3 \times 3 \times 48$), CONV4($3 \times 3 \times 64$)。全连接层的尺寸如下:FC1(32), FC2(64), FC3(128), FC4(512), FC5(128), FC6(64), FC7(2)。

图3所示网络的最终预测结果为视线与显示器屏幕的交汇点,因此特别考虑了头部位置。但对于纯视线方向估计问题,本文对原有模型进行的简化,得到了如图4所示的视线方向估计模型,该模型的最终预测结果是视线在相机坐标系下的方向,该模型的预测结果虽然不能直接用于眼球驱动,但在用户头部位置固定的场景中,该模型相较于上一模型能够得到更高精度的视线估计结果。

模型直接将卷积结果和经过全连接层 FC1 作用后的头部姿态信息相融合,因此可将 FC1、FC2、FC3 认作是利用全连接网络来拟合头部坐标系到相机坐标系的旋转变换 R 。

该网络的卷积层具体尺寸如下:CONV1($5 \times 5 \times 12$), CONV2($3 \times 3 \times 24$), CONV3($3 \times 3 \times 48$), CONV4($3 \times 3 \times 64$)。全连接层的尺寸如下:FC1(256), FC2(128), FC3(64), FC4(2)。

3 基于眼球驱动的电动病床系统设计

3.1 病床结构

该基于眼球驱动的电动病床结构如图5所示,除了普通病床的主体结构以外,该病床还包括坐姿调整装置、

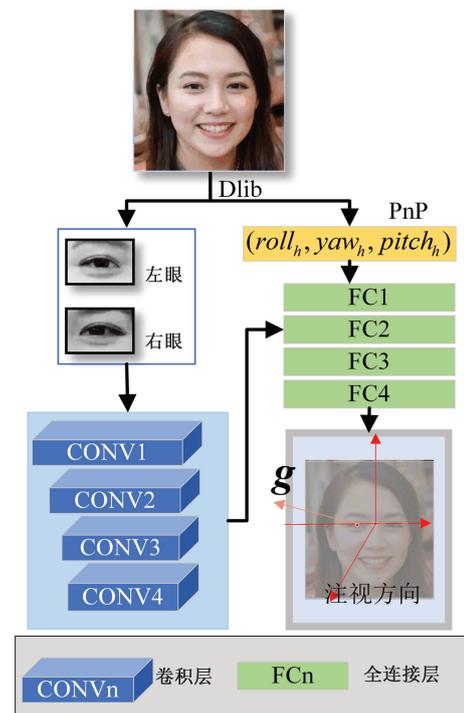


图4 注视方向估计网络示意图

Fig. 4 Illustration of the gaze direction estimation network

左右侧翻装置、排泄装置、腿姿调节装置、病床行走装置、控制单元、相机、小型显示器以及电源。

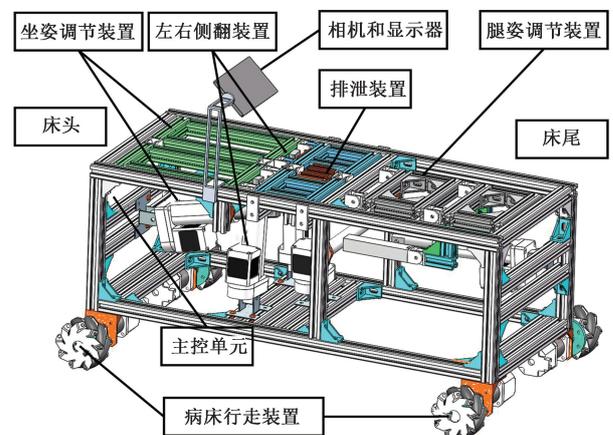


图5 电动病床三维结构图

Fig. 5 3D structure of the electric sickbed

电动病床的各个关节都由步进电机驱动,除了病床行走装置以外,该病床配备5套小型步进电机系统来驱动电动病床进行坐姿调整、左右侧翻、辅助排泄以及腿姿调节,这4组动作又间接通过眼球来驱动。

该系统使得一些重症患者和残疾人能够通过眼睛来调节病床的关节角度从而达到调整卧躺姿态和辅助排泄的效果。4组动作的运动效果如图6所示。

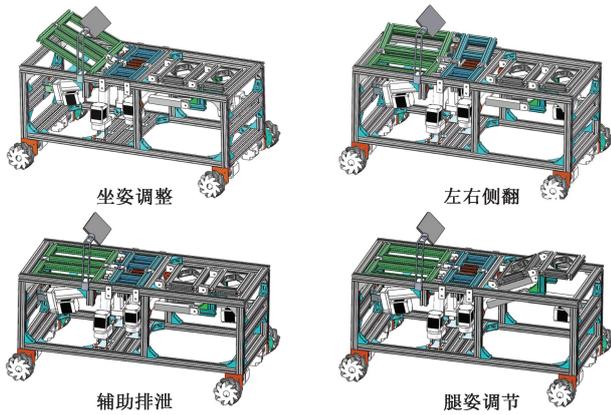


图 6 调整动作示意图

Fig. 6 Illustration of adjustment action

3.2 病床控制与驱动系统

眼球驱动是进行视线追踪的目的之一, 本文将所提出的注视点估计网络作为眼球驱动电动病床系统的视线追踪器。基于眼球驱动的电动病床系统能够将用户的眼动信息转换为电动病床的控制信息。相机采集用户头部图像, 将图像输入到主控单元进行视线追踪从而计算得到用户在前方显示器上的注视点, 而后进行注视信息判决从而确定用户的选择。将用户选择结果传输给下位机, 下位机生成步进电机的控制信号, 最终达到对电动病床进行控制的效果。具体控制流程如图 7 所示。

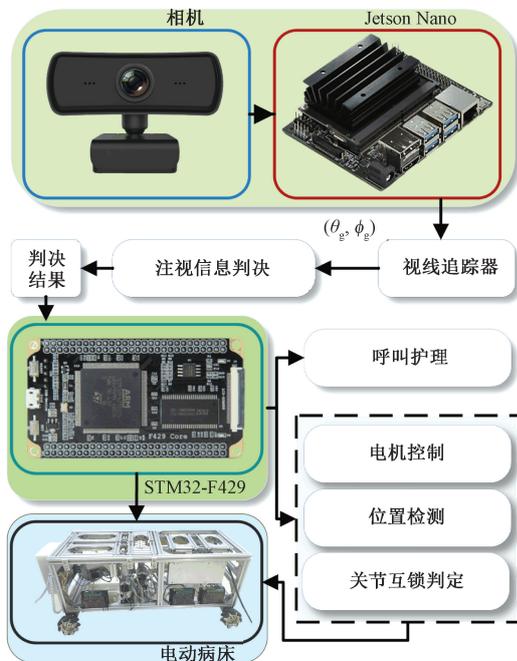


图 7 基于视线的控制流程图

Fig. 7 Control flow chart based on gaze

该系统采用 NVIDIA Jetson Nano 小型嵌入式主板作为主控单元, 运行 Linux 操作系统进行视线估计和注视信息判决, 下位机为 STM32-F429。

3.3 软件部分

视线判决是将视线信息转换为用户选择信息的过程。在本系统中, 用户通过注视显示器显示的控制选择按钮 3 s 即触发控制, 通过闭眼 5 s 即触发锁屏, 通过注视解除锁屏按钮 3 s 即触发解除锁屏。

模型的训练和部署都基于 PyTorch 机器学习库, 利用 Qt 开发图形用户界面应用程序。考虑到在后续开发过程中上位机会加入其他拓展功能, 本文利用神经网络加速框架 TensorRT 提高网络的运算效率。上位机软件框架如图 8 所示。

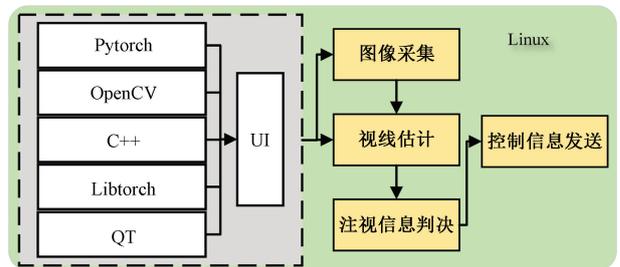


图 8 软件系统架构

Fig. 8 Software system architecture

4 试验与分析

本文分别对注视点估计网络、注视方向估计网络以及病床实体控制进行试验分析。注视点估计网络和注视方向估计网络采用不同的数据集和训练方法。

4.1 数据集设置

本文利用 ColumbiaGaze 数据集^[15]对注视方向估计网络进行训练, 利用 MPIIGaze 数据集^[10]对注视点估计网络进行训练。两个数据集都由不同的照明条件、凝视目标和头部运动设置组成。

ColumbiaGaze 使用相机阵列来获取多视图样本。该数据集包含 5 880 张不同凝视方向和头部姿势的图像对应于 56 个参与者。对于每个参与者有 5 个头部姿势和 21 个不同的凝视方向, 这使得该数据集比其他公开可用的凝视数据集提供了更多的图像和固定凝视目标。该数据集采集的对象种族多样, 且其中 21 人佩戴眼镜。

而 MPIIGaze 数据集是通过单一相机采集不同姿态和注视点的人脸图像构建的, 更符合真实场景。数据集包含 213 659 张图片, 这些图片是在 3 个多月内从 15 名参与者的日常笔记本电脑使用过程中收集的。在外观和照明方面, 该数据集比现有的数据集变化更大。

实验过程中使用这两个数据集(分别为56名和15名受试者)的所有可用数据。数据集设置概况如图9所示。

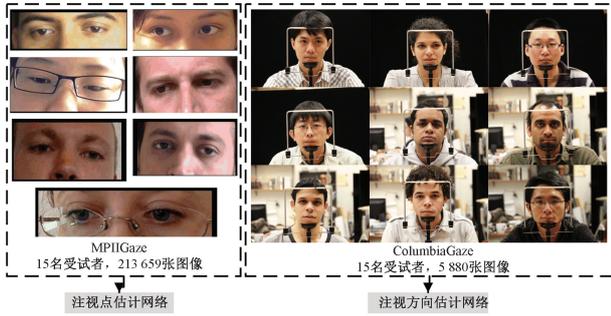


图9 数据集设置示意图

Fig. 9 Illustration of data set setting

本文在训练过程中对眼睛的边界框和眼角的地标位置进行抖动,利用眼睛区域的动态裁剪能够使模型对光线、摄像机运动、噪声或面部和眼睛坐标检测错误的微小变化更鲁棒。MPIIGaze数据集自带有眼角坐标等信息,因此本文直接对现有的眼角坐标进行位置抖动,模型也不需要利用Dlib进行特征点定位和头部姿态解算。而ColumbiaGaze并未对眼部位置进行标记,因此需先利用Dlib进行特征点定位后再加入位置抖动。

4.2 模型训练

为比较网络性能,本文对文献[16]所提出的网络进行训练用于和本文提出的注视点估计网络的比较,对文献[13]所提出的网络进行训练用于和本文提出的注视方向估计网络的比较。对两种注视点估计网络进行训练时,将数据批处理大小设为128,学习速率设为0.001,权重衰减速率设为0.0001,优化器选用Adam,训练迭代次数为100K,损失函数使用L2。对本文所提出的注视方向估计网络进行训练时,将数据批处理大小设为32,训练迭代次数为6K,损失函数选用均方误差(MSE Loss),其他超参数设置与注视点估计网络相同。对文献[12]所提出的网络进行训练时,将数据批处理大小设为16,训练迭代次数为32K,其他超参数设置与文献[13]相同。

4.3 模型精度对比与分析

在开放数据集上训练得到模型后,利用新成员不同数量的标定样本进行模型微调训练,得到如图10(a)、(b)所示的训练结果,横坐标表示再次训练采用的标定样本数或参考样本数,纵坐标表示平均预测误差。图10(a)所示为模型平均预测误差随标定样本数的变化情况,平均预测误差用平均欧氏距离表示。将从文献[15]基于iTracker衍生出的3个模型与本文提出的模型进行比较。图10(b)所示为模型平均预测误差随参考样本数

的变化情况,平均预测误差用平均L1距离表示。将从文献[13]衍生出的5个模型与本文提出的模型进行比较。

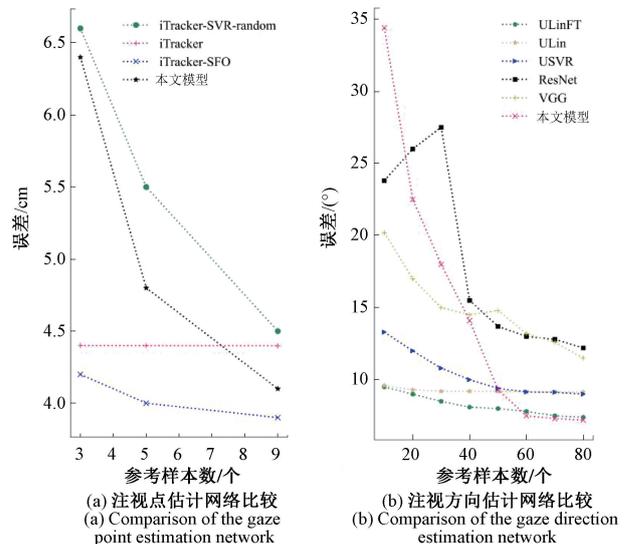


图10 不同标定样本数量下的预测精度曲线图

Fig. 10 Prediction accuracy curves of different calibration samples

观察图10(a)的曲线可以发现,本文所提出的注视点估计网络相较于其他3个网络误差下降得更快,说明该网络比与其他3个网络相比提取用户特征信息能力更强。造成该现象的原因可能在于该网络并未将人脸图像和对应特征的特征空间这些高维度的图像信息作为网络的输入,而是利用人眼位置和直接计算得到的头部姿态作为位置和姿态信息。另外,采用9个标定样本后,本文提出的模型能够保持最低的预测误差。

观察图10(b)的曲线可以发现,本文所提出的网络相较于其他5个网络在只有10个参考样本时,预测误差最高。本文认为造成该现象的原因在于本文提出的模型结构较为简单,在完成预训练后针对特定用户进行视线预测时反而会更依赖参考样本的数量。但随着参考样本数增加,模型的预测精度不断提高,在参考样本数达到80时,本文提出的模型预测误差达到几个网络预测误差的最低值。

在注视方向估计网络的试验过程中,本文还加入了混淆矩阵用于进一步显示模型的预测性能,如图11所示。混淆矩阵的数据来源于模型在ColumbiaGaze数据集上的预测结果,将方向处于垂直平面 $-10^{\circ} \sim 10^{\circ}$ 以及水平平面 $-5^{\circ} \sim 5^{\circ}$ 范围内的视线规定为中,方向在垂直平面内为 10° ,水平平面内为 15° 的视线规定为右上,以此类推。

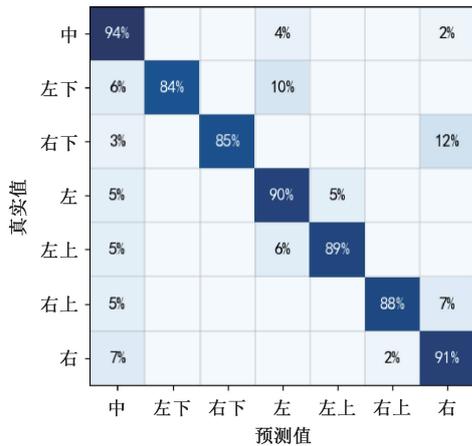


图 11 注视方向估计网络预测结果混淆矩阵
Fig. 11 Confusion matrix of gaze direction estimation network prediction results

观察图 11 可发现,左下和右下的预测准确性较低,本文认为,造成该现象的原因在于当注视方向偏向于下方时,人眼开合程度会降低,造成部分角膜被遮挡进而影响模型的预测效果。最终的试验结果如表 1 所示,本文将注视点估计网络在 9 个标定样本基础上的平均预测误差作为估计误差,将注视方向估计网络在 80 个参考样本基础上的平均预测误差作为估计误差。

表 1 模型精度对比表

Table 1 Comparison table of model accuracy

作用	模型名称	估计误差	描述
注视点估计网络	iTracker-SVR	4.5 cm	iTracker+SVR+随机位置
	-random		
	iTracker	4.4 cm	iTracker 网络的改进版本
	iTracker-SFO	3.9 cm	iTracker+小样本学习
	本文模型	4.1 cm	轻量网络+头部自由
注视方向估计网络	ULinFT	7.4°	无监督线性适应模型微调
	ULin	9.1°	无监督线性适应
	USVR	9.0°	无监督 SVR 适应
	ResNet	12.2°	残差网络的修改版
	VGG	11.5°	VGG 网络的修改版
	本文模型	7.2°	轻量网络+头部自由

观察表 1 可以发现,本文提出的注视点估计网络具有较高的估计精度,但相较于 iTracker-SFO 网络,其估计精度稍低。本文认为造成该现象的原因在于 iTracker-SFO 使用了先进的小样本学习理论进行用户标定,能够大幅提高标定效果,但其标定过程计算复杂,时长可达数

分钟。另外,本文提出的注视方向估计网络在 ColumbiaGaze 数据集上相较于其他网络的预测效果最佳,预测精度比采用了复杂的无监督学习和线性适应的 ULinFT 网络高 0.2°,具有一定先进性。

4.4 模型大小和运算速度对比

训练后得到的模型保存为 .pt 文件后直接在 Jetson Nano 上进行部署,且在不加载 TensorRT 加速框架的情况下对各个网络的运算速度进行统计。几种注视点估计网络和注视方向估计网络的模型大小和运算速度如表 2 所示。

表 2 模型大小和运算速度对比表

Table 2 Comparison table of model size and operation speed

作用	模型名称	模型大小/MB	运算速度/ms
注视点估计网络	SAGE ^[16]	9.6	69.5
	iTracker ^[5]	26.7	267.2
	本文模型	3.6	36.5
注视方向估计网络	Faze ^[6]	67.8	342.2
	GazeNet ^[10]	90.0	443.3
	本文模型	1.6	22.5

4.5 病床响应速度与控制效果分析

基于前文所述的视线追踪和眼球驱动方法,开发出基于眼球驱动的电动病床硬件设备实物模型,如图 12 所示,该系统采用本文提出的注视点估计网络。

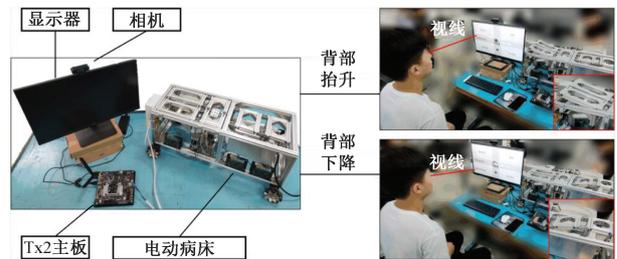


图 12 基于眼球驱动的电动病床实物模型

Fig. 12 Physical model of electric sickbed based on eye driving

为体现该基于眼球驱动电动病床的控制效果和效率,本文对各个控制方法通过眼球驱动响应时间进行统计。在用户前方适当位置的显示器上会显示 6 个选择按钮,将系统计算得到用户开始注视对应的选择按钮作为计时起点,相应的电机开始动作作为计时终点(锁屏和解锁屏则将界面切换完成作为计时终点),进行 20 次测试,起点和终点之间的平均时间间隔作为响应时间。测试结果如表 3 所示。

观察表 3 可发现,系统对于各个指令的响应延时都能保持在 0.5 s 以内,解锁屏和锁屏的响应延时更是能保

表 3 系统响应速度统计表

Table 3 Statistical table of system response speed

动作方式	响应时间	动作方式	响应时间
坐姿控制模式	3.21	腿姿控制模式	3.33
抬高背部	3.13	抬高腿部	3.23
降低背部	3.45	降低腿部	3.49
侧翻控制模式	3.27	排泄控制模式	3.26
左侧翻	3.31	开启排泄门	3.19
右侧翻	3.49	关闭排泄门	3.45
呼叫护理	3.12	解锁屏/锁屏	3.16/5.19

持在 0.2 s 以内,这使得系统能够高效、及时地将用户的注视信息转换为驱动信息并做出响应。

5 结 论

本文针对基于单目相机的头部自由视线追踪问题,提出了两个视线追踪网络,能够分别对注视点和注视方向进行估计,并构建出了基于眼球驱动的电动病床系统。

1) 提出了新型注视点估计网络,创新性地将双通道左右眼图像、左右眼的左右眼角坐标以及通过 Dlib 计算得到的头部姿态作为网络的输入来估计注视点,实现了基于单目相机的头部自由注视点估计,模型具有高精度和轻量的特点。网络模型在 MPIIGaze 数据集上的误差为 4.1 cm。

2) 提出了新型注视方向估计网络,该网络在本文提出的注视点估计网络的基础上简化而来,网络同样具有高精度和轻量的特点。网络模型在 ColumbiaGaze 数据集上的误差为 7.2°。

3) 提出了一套基于眼球驱动的电动病床控制方法并构建了相应的实物模型,采用本文所提出的注视点估计网络作为视线追踪器。该系统具有控制精度高、响应速度快等特点,该系统使得重症患者或残疾人能够通过眼球来控制电动病床,方便了病人的生活起居。

参考文献

- [1] ALBERTO K, MORA F, MONAY F, et al. EYEDIAP[P]. Eye Tracking Research and Applications, 2014.
- [2] 赵凯琳,靳小龙,王元卓.小样本学习研究综述[J].软件学报,2021,32(2):349-369.
- ZHAO K L, JIN X L, WANG Y ZH. A review of few-shot learning [J]. Journal of Software, 2021, 32(2): 349-369.
- [3] MORA K, ODOBEZ H J M. Geometric generative gaze estimation (G3E) for remote RGB-D cameras[C]. IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2014.
- [4] WOOD E, BALTRUŠAITIS T, MORENCY L. Learning an appearance-based gaze estimator from one million synthesised images [P]. Eye Tracking Research & Applications, 2016.
- [5] KRAFKA K, KHOSLA A, KELLNHOFER P, et al. Eye tracking for everyone [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, DOI: 10.1109/CVPR. 2016. 239.
- [6] PARK S, MELLO S D, MOLCHANOV P, et al. Few-shot adaptive gaze estimation [J]. Proceedings of the IEEE International Conference on Computer Vision 2019, October, 2019: 9367-9376.
- [7] 潘新星,汪辉,陈灵,等.基于卷积目标检测的3D眼球追踪系统深度估计[J].仪器仪表学报,2018,39(10):241-248.
- PAN X X, WANG H, CHEN L, et al. Depth estimation of 3D eye tracking system based on convolution target detection[J]. Chinese Journal of Scientific Instrument, 2018, 39(10):241-248.
- [8] 迟健男,张闯,王翠娟,等.双摄像机双光源视线追踪系统标定方法研究[J].仪器仪表学报,2011,32(4):883-891.
- CHI J N, ZHANG CH, WANG C J, et al. Calibration for gaze tracking system of binocular vision [J]. Chinese Journal of Scientific Instrument, 2011,32(4):883-891.
- [9] ZHANG X C, SUGANO Y, FRITZ M, et al. It's written all over your face: Full-face appearance-based gaze estimation [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2017.
- [10] ZHANG X C, SUGANO Y, FRITZ M, et al. MPIIGaze: Real-world dataset and deep appearance-based gaze estimation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,41(1):162-175.

- [11] SUGANO Y, MATSUSHITA Y, SATO Y. Learning-by-synthesis for appearance-based 3d gaze estimation [C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014: 1821-1828.
- [12] YU Y, ODOBEZ J M. Unsupervised representation learning for gaze estimation [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.
- [13] XIONG Y Y, KIM H J, SINGH V. Mixed effects neural networks (MeNets) with applications to gaze estimation [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019.
- [14] HUANG Q, VEERARAGHAVAN A, SABHARWAL A. TabletGaze: Dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets [J]. Machine Vision and Applications, 2017, 28 (5): 445-461.
- [15] SMITH B A, YIN Q, FEINER S K, et al. Gaze locking: Passive eye contact detection for human object interaction [J]. UIST 2013-Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, 2013: 271-280.
- [16] HE J, PHAM K, VALLIAPPAN N, et al. On-device few-shot personalization for real-time gaze estimation [C]. 2019 International Conference on Computer Vision Workshop, ICCVW 2019, IEEE, 2019: 1149-1158.

作者简介



胡佳辉, 2020 年于南京航空航天大学就读硕士研究生, 主要研究方向为视线追踪和眼球驱动。

E-mail: hujiahui@nuaa.edu.cn

Hu Jiahui is a master student at Nanjing University of Aeronautics and Astronautics since 2020. His main research interests include eye-tracking and eye-driving.



陆永华 (通信作者), 2005 年于南京航空航天大学获得博士学位, 现为南京航空航天大学教授, 主要研究方向为智能检测与控制、测试系统、机器人。

E-mail: nuaa_lyh@nuaa.edu.cn

Lu Yonghua (Corresponding author) received his Ph. D. degree from Nanjing University of Aeronautics and Astronautics in 2005. He is currently a professor at Nanjing University of Aeronautics and Astronautics. His main research interests include intelligent measurement and control, measurement system and robotics.



张进海, 2013 年于成都大学获得学士学位, 现为成都飞机工业(集团)有限责任公司工程师, 主要研究方向为机械设计、飞机装配工艺。

E-mail: zjh_hay@163.com

Zhang Jinhai received his B.Sc. degree from Chengdu University in 2013. He is currently an engineer at Chengdu Aircraft Industrial (Group) Co., Ltd. His main research interests include mechanical design and aircraft assembly.