

眼动跟踪研究进展与展望

苟超¹ 卓莹¹ 王康² 王飞跃³

摘要 眼动跟踪是指自动检测瞳孔中心位置或者识别三维视线方向及注视点的过程, 被广泛应用于人机交互、智能驾驶、人因工程等. 由于不同场景下的光照变化、个体眼球生理构造差异、遮挡、头部姿态多样等原因, 眼动跟踪的研究目前仍然是一个具有挑战性的热点问题. 针对眼动跟踪领域, 首先概述眼动跟踪研究内容, 然后分别论述近年来瞳孔中心检测及视线估计领域的国内外研究进展, 综述目前眼动跟踪主要数据集、评价指标及研究成果, 接着介绍眼动跟踪在人机交互、智能驾驶等领域的应用, 最后对眼动跟踪领域的未来发展趋势进行展望.

关键词 眼动跟踪, 人眼检测, 视线估计, 注意力分析

引用格式 苟超, 卓莹, 王康, 王飞跃. 眼动跟踪研究进展与展望. 自动化学报, 2022, 48(5): 1173–1192

DOI 10.16383/j.aas.c210514

Research Advances and Prospects of Eye Tracking

GOU Chao¹ ZHUO Ying¹ WANG Kang² WANG Fei-Yue³

Abstract Eye tracking is a process of automatically detecting the location of pupil or recognizing the gaze direction and gaze point, and is widely used in human-computer interaction, intelligent driving, ergonomics, and so on. Eye tracking is still a challenging and hot topic due to the changes of illumination, the differences of individual eyeball physiological structure, occlusion, and the diversity of head pose. In this paper, we focus on the study of eye tracking. Firstly, we give an introduction of eye tracking, followed by discussing the research progress of pupil detection and gaze estimation in recent years. Then, we review the datasets and evaluation metrics with results, followed by introducing the application of eye tracking in human-computer interaction, intelligent driving and other fields. Finally, the future trends of eye tracking are discussed.

Key words Eye tracking, eye detection, gaze estimation, attention analysis

Citation Gou Chao, Zhuo Ying, Wang Kang, Wang Fei-Yue. Research advances and prospects of eye tracking. *Acta Automatica Sinica*, 2022, 48(5): 1173–1192

人类获取信息很大程度上依赖于人眼感知的视觉信息, 人眼被称为心灵的窗户, 在人的自然交互中有着重要的意义, 它能感知周围环境、表现人的注意力、传达个人的情感等. 因此, 眼动跟踪被广泛研究与应用. 眼动跟踪任务及应用示例如图 1 所示. 眼动跟踪一般是指自动检测人眼瞳孔相对位置或估计视线方向的过程. 图 1(a) 为眼动追踪任务示例.

眼动能描述瞳孔的动态变化, 能直观地反映人的注视点和注视时间, 对于描述人类视觉感知过程与特点、重现个体基本认知加工过程有重要意义, 被广泛应用于人机交互感知、智能驾驶座舱、情感计算、意图分析等领域^[1-2], 图 1(b) 为眼动追踪应用示例. 由于环境光照的多变、个体眼球生理特性的差异、头部姿态的动态变化等, 眼动跟踪仍然是一个具有挑战的热点研究问题. 眼动跟踪的研究可以追溯到 20 世纪初, 研究人员采用了接触式的眼动跟踪方法, 通过在眼睛周围放置成对电极的眼电图分析, 或者基于线圈嵌入式的巩膜搜索方法来获取并分析眼动信息; 在 20 世纪 50 年代有了第 1 部记录基于飞行员瞳孔跟踪及视线分析的控制飞机操作应用系统的视频资料; 在 60 年代头戴式眼动跟踪得到发展; 在 90 年代视线估计方面开始使用计算机进行计算处理控制. 近年来, 随着计算机视觉、图形学、人工智能等新兴技术的发展以及应用需求的增加, 非接触式的基于视频图像的眼动跟踪引起了学术界和工业界大量深入研究^[1, 3-7].

本文根据多年眼动跟踪领域的研究基础与积

收稿日期 2021-06-08 录用日期 2021-09-17
Manuscript received June 8, 2021; accepted September 17, 2021

国家自然科学基金(61806198), 广州市重点研发计划(202007050002), 深圳科技计划项目(RCBS20200714114920272) 资助

Supported by National Natural Science Foundation of China (61806198), Key Research and Development Program of Guangzhou (202007050002), Shenzhen Science and Technology Program (RCBS20200714114920272)

本文责任编辑 黄华

Recommended by Associate Editor HUANG Hua

1. 中山大学智能工程学院 深圳 518107 中国 2. 英伟达 圣克拉拉 95051 美国 3. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室 北京 100190 中国

1. School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen 518107, China 2. Nvidia, Santa Clara 95051, USA 3. The State Key Laboratory for Management and Control of Complex System, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

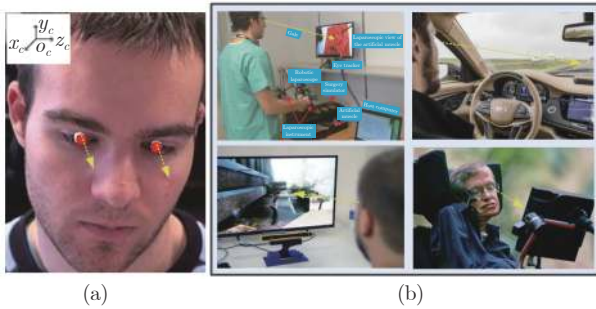


图 1 眼动跟踪人任务及应用示例

Fig.1 Examples of eye tracking and corresponding applications

累, 详细综述该领域的研究进展, 并对未来发展趋势进行展望. 本文结构如下: 第 1 节整体概述眼动跟踪技术的发展. 第 2 ~ 3 节分别阐述人眼瞳孔检测、视线方向与注视点的估计方法研究进展, 重点阐述基于图像的眼动跟踪研究进展及相关的数据集. 第 4 节综述相关数据集及性能比较方法. 第 5 节介绍相关应用. 第 6 节对眼动跟踪的未来发展趋势进行展望.

1 发展概述

眼动跟踪是自动实现眼球瞳孔中心及注视点定位的过程. 人类获取的外部信息主要通过人眼感知的视觉信息, 眼动能直观地反映人的注视点和注视时间, 对于描述人类视觉感知过程与特点、探究个体基本认知过程有重要意义. 常见的眼动行为主要有注视、扫视和平滑追踪有 3 种形式. 眼动可以通过时间和空间两个维度来进行描述, 常见指标有平均注视时间、注视次数、注视顺序、平均扫视幅度、眼跳次数、扫视持续时间、扫视方向等. 早期的眼动跟踪技术应用于心理学领域, 这一时期常用方法有直接观察法、后象法和机械记录法等. 直接观察法和后象法只能粗略地描述眼动, 机械记录法需要角膜与测量装置接触, 作为一种接触式的眼动跟踪方法, 容易造成眼部不适. 由于精度不足和调试难度大等问题, 这些方法逐渐被淘汰. 目前眼动跟踪方法可以分为探查线圈记录法、红外线法、电流记录法和视频记录法 (Video oculography, VOG) 4 类.

1.1 探查线圈记录法

探查线圈记录法最早由 Robinson^[8] 于 20 世纪 60 年代提出, 该方法利用带有线圈的隐形眼镜在电磁场中运动, 通过电磁感应原理产生的信号来测量眼球水平及垂直方向的运动, 从而实现眼睛位置检测. 眼动有时伴随着头部运动, 传统的探查线圈记

录法利用亥姆霍兹线圈构建磁场, 但无法直接捕捉头部运动^[2]. Eibenberger 等^[9] 利用 Merritt 线圈产生稳定磁场, 并利用 MATLAB 进行仿真, 实现了无需头戴设备的检测, 并成功降低了测量误差. 此外还提出了一种新的硬件系统和校准方法, 可以将记录模拟信号的老式线圈系统改装为数字记录系统, 同时指出不同的运算放大器可能有助于探查线圈记录法的性能提升^[10]. 探查线圈记录法测量准确、采样率高, 但无法长时间佩戴, 成本也相对较高.

1.2 红外线法

红外线法需要在眼部附近安装红外光敏管, 根据瞳孔、巩膜、角膜等不同光学界面反射的图像来测量眼动. 常用方法包括瞳孔-角膜反射法、虹膜-角膜反射法等. 这种测量方法目前已被广泛地研究和应用, 许多学者提出了光线检测或瞳孔检测的算法提升测量准确率. 文献 [11] 提出了快速鲁棒椭圆检测算法, 利用径向对称性搜索瞳孔位置, 并利用两个细化步骤来适应某些图片中发生椭圆形变的瞳孔外观, 增强了模型的泛化能力. Mestre 等^[12] 利用 12 次角膜反光获得高清瞳孔和角膜图像, 同时使用了一个标准化因子校正瞳孔中心反射点向量. 标准化因子的引入提升了垂直眼动跟踪 43.2% 的准确率, 同时减少了设备对于光源数量的要求. 其他具有代表性的相关算法和技术有 Santini 等^[13] 制作的开源的 EyeRecToo 软件、Else^[14]、星光镜和霍夫变换结合^[15] 等. 红外线法是一种非侵入式的眼动跟踪方法, 不会造成眼部不适且精度较高.

1.3 电流记录法

电流记录法是一种成本较低、操作性强的眼动跟踪方法, 在闭上眼睛的情况下也可以正常工作^[16]. 眼球发生转动时会产生生物电现象, 具体是指角膜和视网膜新陈代谢率的不同导致产生电位差, 电位变化可以通过传感器测量眼部周围皮肤获得. 如图 2 所示, 苏黎世联邦理工学院研究的一种电流记录法眼动跟踪设备^[17], 通过不同的电极片分别记录垂直眼动和水平眼动. 由于设备问题, 如果电位差过小或电极接触不良, 电流记录法无法保证测量效果, 所以电流记录法并不适用于日常眼动跟踪应用, 一般应用于特定实验室环境下.

1.4 视频记录法

视频记录法主要利用视频摄像机记录的眼球运动过程, 通过计算机自动分析处理视频图像. 摄像机可以采集人眼图像或者眼球反射的红外线图像, 然后通过计算机自动处理图像数据, 实现非侵入式

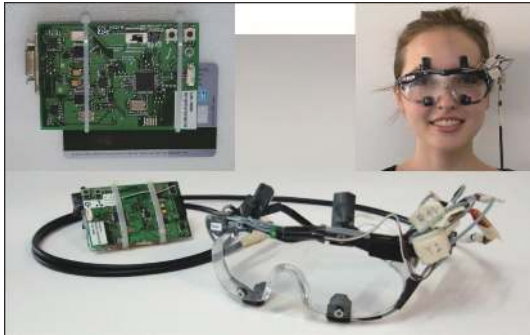


图2 苏黎世联邦理工学院的电流记录法眼动仪^[17]
Fig.2 An eye tracker based on electrooculography
from eidgenössische technische hochschule^[17]

的人眼瞳孔中心检测或者视线注视点估计。从摄像机安装方式上, 视频记录法可以分为头盔式、遥测式和头部固定式, 相机数量上可以使用单相机或多相机。单相机方法通常使用红外光源引起暗瞳效应, 此时瞳孔较暗, 虹膜较亮, 容易分割瞳孔区域从而实现眼动追踪。多相机方法拥有更大的视野, 通常使用一台相机拍摄眼睛图像, 其他相机用于定位头部位置或进行校准^[2]。Khamis 等^[18] 研发 EyeScout 系统, 利用摄像头识别人体运动信息, 并利用 Pursuits 方法检测眼动和屏幕中物体运动轨迹的相关性, 进而判断是否需要调整眼动追踪装置的位置以适应人体运动。Kong 等^[19] 将 VOG 技术应用于人耳前庭功能检测, 使用红外摄像机和红外发光二极管获得视线水平移动数据, 利用形态学运算和轮廓检测获取瞳孔位置。Brousseau 等^[20] 提出一种融合了三维视线估计的混合模型, 训练了多个卷积神经网络用于眼睛区域定位和特征提取, 在传统位置估计网络的基础上增加了质量中心结构, 实验结果表明眼动跟踪的准确率有显著提升。

随着计算机视觉、模式识别、图形学、人工智能等技术的快速发展, 基于视频图像分析的非接触式眼动跟踪受到广泛关注, 大量相关研究获得开展, 取得了突破性进展及应用, 本文将重点论述基于计算机视觉及机器学习的眼动跟踪研究进展, 分别从眼动跟踪包含的瞳孔检测及视线估计两个任务进行详细阐述。

2 瞳孔检测

瞳孔检测, 有时也称为人眼(中心)检测, 其目的是在给定的人脸图像中自动完成瞳孔位置的定位。准确的人眼瞳孔检测是视线估计、注意力分析及人机交互等的基础。人眼瞳孔检测领域已经开展了数十年的研究。基于不同的特征将相关方法总结为基于形状结构、基于外观纹理和基于上下文信息三类。

2.1 基于形状结构的瞳孔检测方法

人眼的特殊生理结构构成, 尤其是虹膜的椭圆形几何结构可以帮助实现瞳孔中心的检测。在早期的研究中, Yuille 等^[21] 设计了一个参数化的形变模型来描述人眼瞳孔的几何形状, 通过拟合 11 个参数来覆盖可能的形变空间, 从而完成基于形状信息的瞳孔检测。Hansen 等^[22] 基于虹膜的椭圆形状, 提出一种似然模型来捕捉瞳孔周边信息, 利用期望最大化和随机抽样一致算法来拟合图像中的椭圆结构, 从而实现瞳孔检测。Valenti 等^[23] 首先检测圆形曲线边缘, 然后提取相等强度的等照度线曲率, 并设计合适的基于投票的方法来实现准确的瞳孔定位。他们在等照度线方法的基础上添加尺度不变性, 最后用每个区域关键点的尺度不变特征变换向量与数据库进行匹配确定眼睛位置^[24]。还有一些研究基于脸部结构信息来完成人眼中心定位, 脸部结构信息主要包括人脸五官构成, 即人眼、鼻子等的结构分布, 这些先验分布可以快速帮助完成人脸中的人眼区域定位。Tan 等^[25] 提出一种增强的画像结构^[26] 来建模描述人脸几何分布关系, 并通过构建左眼、右眼及鼻的位置关系来实现人脸中人眼中心的定位。Wang 等^[27] 提出一种基于全卷积网络的瞳孔定位与虹膜分割方法, 并利用瞳孔中心、虹膜区域以及其内外边界的形状结构信息来同时实现瞳孔定位与虹膜分割, 部分结果如图 3 所示, 其中图 3(a) 为原图, 图 3(b) 为真实标签, 图 3(c) 为模型检测结果, 图 3(d) 为基于瞳孔、虹膜及其边界的形状结构进行后处理的最终结果。Gowroju 等^[28] 提出一种基于 U-Net 的编码解码瞳孔分割方法来实现瞳孔的准确定位。

基于形状结构的瞳孔检测方法主要出现在早期的眼动跟踪研究中, 其主要观点是利用先验知识, 基于人眼在视频图像中呈现的特定椭圆或者圆形形状及几何关联关系的特性来对人眼瞳孔进行定位。由于其过多依赖于固有先验几何知识, 无法较好应对复杂多变的环境, 且由于视频图像角度变化导致形状拉伸及畸变, 传统基于形状结构的方法普遍存在鲁棒性不足的问题。

2.2 基于外观纹理的瞳孔检测方法

基于形状的瞳孔检测方法无法应对有较大头部姿态的场景, 而基于外观纹理特征的方法对头部姿态等更为鲁棒, 其核心是通过设计一些特征提取方法来描述瞳孔的边缘、颜色等特征。Araujo 等^[29] 提出一种基于相关性滤波的内积检测子来实现人眼瞳孔中心检测。Wu 等^[30] 提出一种基于深度学习的特

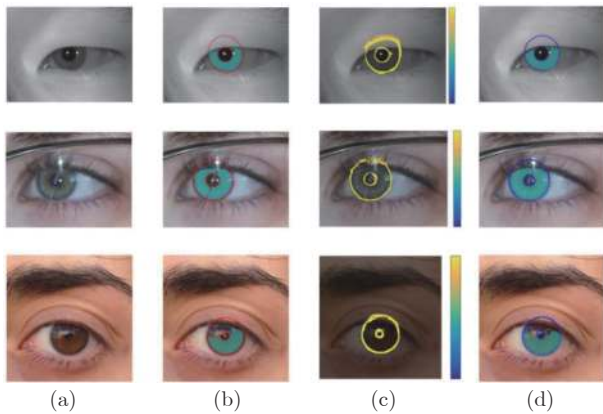


图 3 基于 IrisParseNet 的瞳孔检测与虹膜分割结果示例图^[27]

Fig.3 Some localization and segmentation results based on IrisParseNet^[27]

征提取方法来实现无环境限制的人眼中心检测. Fuhl 等^[31]提出训练双级深度卷积网络来实现人眼检测,并建立一个瞳孔中心的概率分布来实现更为精确的定位,大规模实验数据上的测试效果表明了该方法的有效性.文献^[32]提出通过图像增强、Gabor 滤波、聚类分析和相似性度量的相邻区域比较提取眼部区域,实现精确的人眼检测.文献^[33]提出基于 R-CNN 和 YOLO、能同时实现虹膜及眼周区域检测的方法,通过深度特征来表征眼部外观纹理信息,从而实现准确定位. Xia 等^[34]提出一种基于全连接网络的方法来实现瞳孔检测.将人眼瞳孔定位看作是一种特殊的语义分割任务,以人眼瞳孔为中心的高斯热度图为分割目标,利用全连接网络实现人眼外观纹理特征表征学习,从而实现准确有效的人眼瞳孔检测,部分实验结果如图 4 所示,其中图 4(a) 是最终检测结果,图 4(b) 为左瞳孔热度图,图 4(c) 为右瞳孔热度图. Hsu 等^[35]提出了两种人眼瞳孔检测的方法,一种是基于脸部几何转换的方法,另一种是基于生成对抗网络的图像转换方法.其中生成对抗方法利用真实人眼的外观纹理特征分布,实现眼部区域的正面归一化,解决了较大头部姿态情况下的人眼瞳孔中心检测问题,并在 GAZE 等开源数据集上取得了不错的效果.

基于外观纹理特征的人眼瞳孔检测方法主要基于数据统计分布,效果依赖于海量的标注数据.随着深度学习等技术的快速发展,基于深度学习及外观纹理特征的人眼瞳孔检测方法^[30, 33-34]受到越来越多的关注,其有效性和鲁棒性也推动了眼动跟踪相关的实际应用.而另一方面,随着数据量的增加,不同个体之间的差异也随之增加,数据分布更为多样,

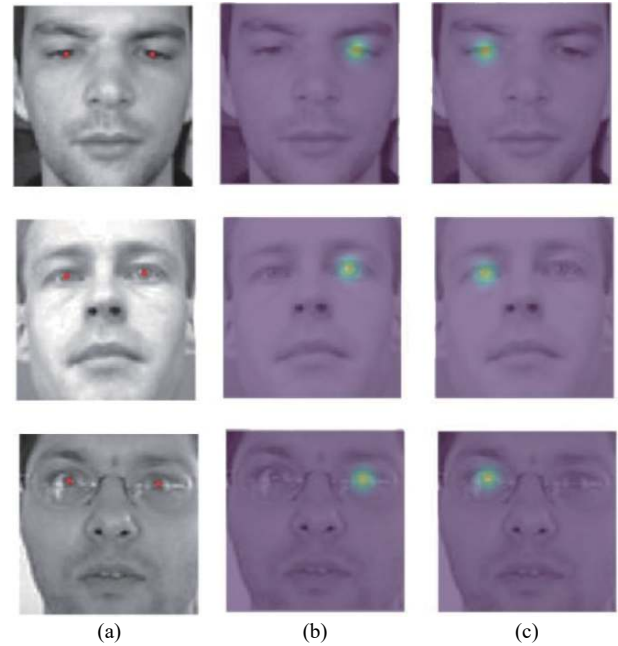


图 4 基于 FCN 的瞳孔检测结果示例图^[34]

Fig.4 Some pupil localization results based on FCN^[34]

使得基于纹理特征的检测能力下降.同时,海量数据需要大量的人力来进行手动标注,如何利用少量有限样本设计出更具有鲁棒性的有效模型,是基于外观纹理特征的人眼瞳孔检测需要解决的主要问题.

2.3 基于上下文信息的瞳孔检测方法

基于上下文信息的方法主要利用眼部区域及其上下文人脸结构与纹理信息来实现人眼瞳孔的准确定位.文献^[36]利用关键点特征,基于由粗到细的思想,在监督下降方法^[37]的基础上提出多尺度非线性特征映射来实现人眼瞳孔精确检测.启发于脸部关键点检测方法, Gou 等^[38]提出基于仿真图像来学习级联线性回归模型,为获取眼部区域的上下文信息,他们首先基于已有脸部关键点检测方法提取眼部大致区域,然后初始化人眼位置并利用级联回归迭代优化,最后实现人眼中心的准确定位,实验结果表明该方法比基于真实图像学习的模型更加鲁棒有效. Gou 等^[39]进一步深入研究,考虑闭眼状态下的眼部全局外观纹理对人眼瞳孔定位的影响,并组合仿真人眼和真实人眼图像为训练对象,从而进一步优化人眼瞳孔检测结果;在级联回归的基础上,引入人眼状态(睁眼/闭眼)的概率来动态调整上下文特征的权重,具体流程如图 5 所示. Xiao 等^[40]提出利用人脸关键点来提取眼部区域,然后利用虹膜区域的圆形拟合来实现准确的瞳孔定位.为使得人眼检测模型更为鲁棒, Gou 等^[41]提出基于平行视觉

系统^[42-45]的瞳孔检测方法, 具体流程如图 6 所示, 核心思想是通过平行图像步骤, 引入生成对抗学习^[46-47], 实现虚实互动的图片生成与优化, 从而为计算实验的级联模型学习提供多样的且带标注的训练样本, 最终实现准确有效的瞳孔检测. Choi 等^[48]基于人脸上下文信息实现瞳孔检测. 提出先检测人脸, 然后基于眼角等关键点提取眼部区域, 最后设计一个分割深度网络来实现准确的瞳孔中心检测. Lee 等^[49]进一步开展基于人脸上下文信息的瞳孔检测与跟踪. 提出先检测人脸, 并通过设计的基于自注意力生成对抗网络去除了戴眼镜的影响, 然后根据人脸关键点提取眼部区域, 最后基于上下文信息通过深度神经网络实现瞳孔中心检测. 该方法在 GI4E 数据集上归一化误差小于 0.05 的情况下, 检测准确率达到 99.84%.

综上所述, 基于上下文信息的瞳孔检测方法较好地利用人眼及人脸结构上下文信息, 在人眼中心检测中应用较为广泛. 但是基于上下文信息的瞳孔检测方法对遮挡等情况敏感, 而基于形状结构、特征学习可以部分解决遮挡等问题. 由此, 构建将上下文信息和特征学习融合的模型是该领域研究的一个重要研究方向.

3 视线估计

视线估计是眼动跟踪的重要任务之一, 目的是估计视线的三维方向向量或者二维注视点位置. 视线估计方法可以分为基于模型和基于特征学习两类.

3.1 基于模型的视线估计方法

基于模型的视线估计主要是通过指根据人眼生理结构所构建的眼球模型, 并通过人脸关键点、瞳孔中心及深度信息、眼角位置等的几何关系来估计人眼的视线方向或者注视点^[50]. 如图 7 的三维眼球模型所示, 人眼的视线方向实际上是视轴方向, 可由角膜曲率中心与中央凹 (位于眼球后方的视网膜中敏锐度最高的部位) 来共同确定; 人眼的光轴方向则会通过瞳孔中心、角膜曲率中心和眼球中心.

人眼的光轴与视轴之间存在着被称为卡帕角的角度差. 卡帕角的大小约为 5° , 其准确值因人而异且唯一.

根据估计原理的不同, 基于模型的视线估计方法可以进一步分为基于二维映射和基于三维几何两类方法.

3.1.1 基于二维映射的视线估计方法

基于二维映射的方法主要指构建注视点或者视线方向与用户标定或者检测的二维视线参数的映射模型来实现视线估计, 二维视线参数包括眼角点、瞳孔位置、眼球中心、角膜曲率中心等. 实际应用时首先计算得到二维视线参数, 然后输入映射模型得到视线方向或者注视点. 该类方法主要包括基于角膜反射的方法、交比值法和单应性归一化法三类.

1) 基于角膜反射的方法将眼角点或者角膜反射点 (普尔钦斑) 作为眼球运动的基准点, 通过分析基准点与瞳孔中心位置的矢量与视线向量的映射关系来实现视线估计. 一般而言, 眼角点难以检测, 而角膜反射点可以通过额外的光源辅助确定^[51].

基于角膜反射的方法一般要求多个摄像头, 以及一个甚至多个的外部近红外线光源来辅助确定眼球半径、角膜曲率半径等不随视线变化而变化的、因人而异的眼睛特征^[52]. Hennessey 等^[53]基于 Goss 等^[54]提出的眼睛模型, 在经过校准后, 实现了摄像头视野范围内的、允许头部自由运动的视线估计. Liu 等^[55]认为虹膜受光线折射的影响较小且更容易被分割出来, 故通过虹膜相关的特征来求得虹膜中心位置, 然后使用虹膜中心而非瞳孔中心来与角膜曲率中心共同确定人眼的光轴方向, 最后结合卡帕角变换矩阵与角膜曲率中心来求得人眼的视线方向. Zhou 等^[56]提出了一个基于梯度特性的虹膜中心定位方法, 并将视线方向近似为从虹膜中心到注视点的方向, 最后以双眼估计得到的视线方向的平均值作为最终的视线估计结果.

基于角膜反射的方法主要是对注视点的估计. 该方法模型简单、易于理解; 且使用的近红外线光源不易受到其他光源的影响, 故该法对光照条件的稳健性良好. 但由于基于角膜反射的方法中假定的

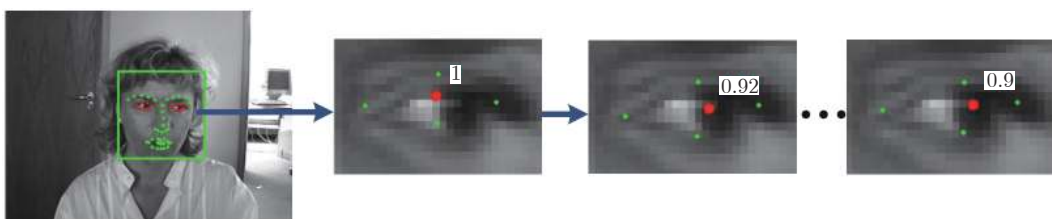


图 5 基于级联回归的瞳孔检测及状态估计流程^[39]

Fig. 5 The framework of cascade regression for simultaneous pupil detection and eye state estimation^[39]

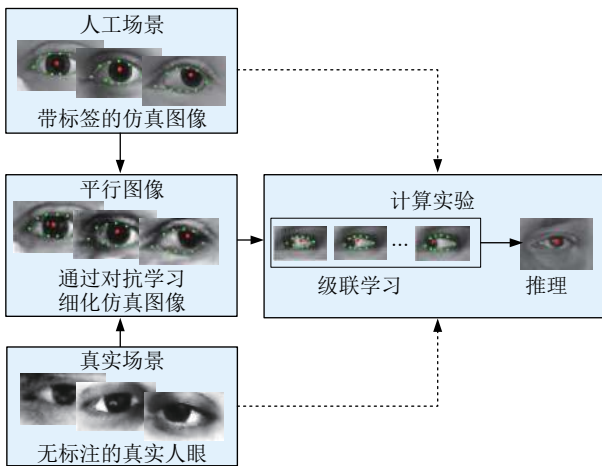


图 6 基于平行视觉的瞳孔检测方法^[41]

Fig.6 The framework of pupil detection based on parallel vision^[41]

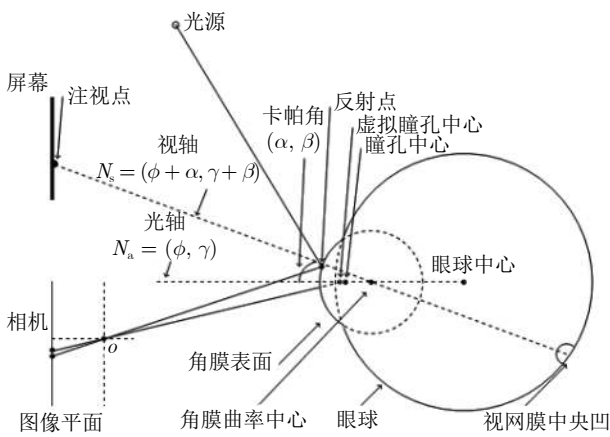


图 7 三维眼球模型及视线估计^[50]

Fig.7 3D eyeball model and gaze estimation^[50]

球形角膜形状与真实角膜形状存在差异, 故其视线估计的精确度有限; 且为求解角膜反射点在照相机成像平面上的位置等, 该法对图像处理的要求很高. 另外, 基于角膜反射的方法需要进行校准, 对实验对象的头部运动范围也有所限制. 具体的校准复杂程度以及实验对象的头部运动范围与系统配置相关. 根据文献 [51], 若仅使用一台摄像机和一个光源, 仅当头部完全静止时才能估计注视点; 若使用一台摄像机和多个光源, 可在完成多点校准过程后自由移动头部来估计注视点; 若使用多台摄像机和多个光源, 可经过简单的单点校准程序后对注视点进行估计.

2) 交比法主要是利用交比在射影变换下的不变性来进行视线估计. 其中的交比射影不变性是指空间中一条直线经空间几何变换后保持了直线上对应点列的交比不变. 交比法主要是对二维屏幕上注

视点的估计. 如图 8 所示, 交比法涉及两个射影变换: 一是从屏幕到虚拟正切平面的投影; 二是从虚拟正切平面到照相机成像平面的投影.

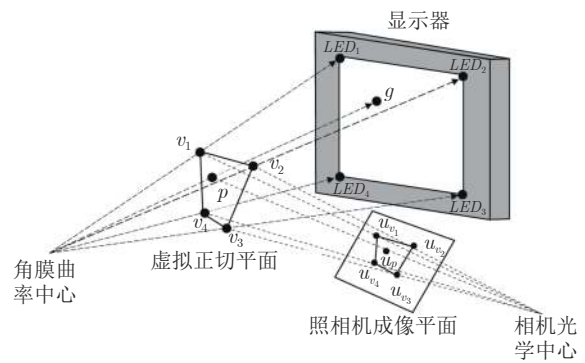
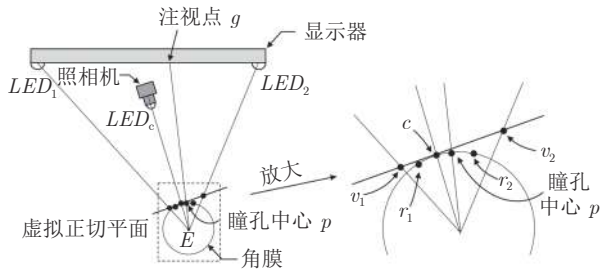


图 8 交比法示意图^[57]

Fig.8 Gaze estimation based on cross-ratio^[57]

理论上^[58], 在屏幕的四个角上分别放置近红外线光源 LED_i ($i = 1, 2, 3, 4$)、并发出闪光; 当人眼在注视屏幕时, 用照相机拍摄, 即可在照相机成像面上获得 LED_i 闪光经人眼角膜反射后的投影点 u_{v_i} ($i = 1, 2, 3, 4$); 还可获得人眼的瞳孔在照相机成像面上的投影点 u_p . 通过 4 个闪光投影点 u_{v_i} ($i = 1, 2, 3, 4$) 和瞳孔投影点 u_p 在照相机成像面上的位置, 即可求得这些点的交比. 根据交比在射影变换下的不变性, 在已知近红外线光源位置的情况下, 可根据瞳孔投影点 u_p 在照相机成像面上的位置以及交比, 求得与近红外线光源同在屏幕上的凝视点 g 的位置. 但实际上, 角膜表面的反射点和瞳孔中心并不在同一虚拟平面上. 为解决该问题, Yoo 等^[57] 在照相机的中心处加设一个近红外线光源, 设光源点 LED_c 和眼球中心的连线与人眼的角膜表面交于点 c , 并以点 c 为切点作角膜表面的虚拟正切平面, 如图 9 所示. Cheng 等^[59] 提出以瞳孔中心为切点作角膜表面的动态虚拟正切平面. 该虚拟正切平面会因注视点的不同而不同, 角膜反射点在动态虚拟正切平面上的具体位置由动态参数矩阵来确定.

交比法的另一个主要误差来源是, 该方法进行视线估计的依据是人眼瞳孔所在的光轴而非代表真实视线的视轴. 为补偿该误差, Coutinho 等^[60] 首先提出对屏幕上的注视点加以固定的二维偏移量来校正视线估计结果. 该二维偏移量会随着头部姿态的改变而改变, 因此固定的二维偏移量不足以准确校正视线估计结果. Coutinho 等^[61] 为此进一步提出, 通过一个比例因子来动态调整该二维偏移量, 该比例因子可由相关的深度变化信息来估计求得. 还有一些研究使用高斯回归^[62]、多项式回归^[63] 等方法对

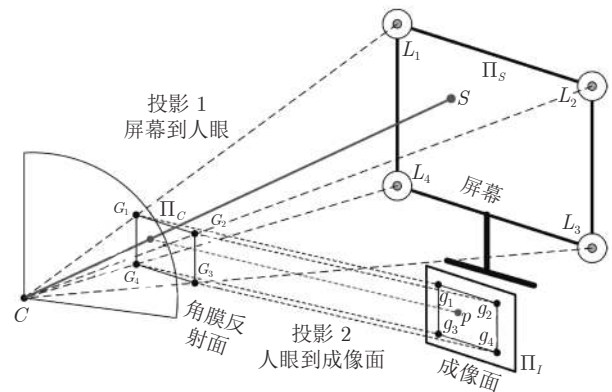
图 9 虚拟切平面示意图^[57]Fig.9 Virtual tangent plane^[57]

误差进行建模,以校正基于交比法的视线估计结果. Arar 等^[64] 则提出使用正则化最小二乘回归对误差进行建模,并通过自适应性融合双眼的注视点信息来提高视线估计的准确性和稳健性.

由于交比法要求角膜表面的反射点都能投射在照相机成像面上,故对实验对象的头部运动范围有所限制.为提高交比法对头部运动范围的稳健性,Coutinho 等^[65] 提出交比特征平面化的方法,用弱透视相机模型和由角膜中心和虹膜平面组成的眼睛模型来补偿头部运动造成的误差. Arar 等^[66] 提出一个基于交比法的多相机视线估计系统.通过多个独立照相机以及对各个相机求得的注视点估计结果的自适应性融合,该系统允许在实验对象头部自由运动条件下进行准确的视线估计.交比法的系统配置要求高,至少需要一台照相机以及 4 个外部光源.鉴于多数的液晶显示器都包含偏振滤光片,使显示器发出的光是偏振的,故可使用偏振摄像头来获取人眼在注视液晶显示器时角膜表面反射的偏振图像,Sasaki 等^[67] 提出使用角膜反射的线性偏振度图像和线性偏振角图像来求得显示器经人眼角膜反射后在照相机成像面上的区域范围,再结合瞳孔中心或角膜边缘的成像位置来进行视线估计,而无需额外的外部光源.

3) 单应性归一化法是对屏幕上注视点的估计,主要利用单应性变换进行视线估计.如图 10 所示,设外部光源在人眼角膜表面反射的各点所在的虚拟平面为角膜反射平面 Π_C ,并设该平面经单应性归一化后的平面为归一化平面 Π_N .首先将照相机成像面 Π_I 上的瞳孔中心位置通过单应性矩阵 H_I^N 投影映射到角膜反射面的归一化平面 Π_N 上.然后通过单应性矩阵 H_N^S 将瞳孔中心在归一化平面 Π_N 的位置投影映射到屏幕 Π_S 上,以求得注视点位置.

交比法与单应性归一法对屏幕注视点的估计都基于二维平面几何关系,但不同于交比法使用交比这一平面投射不变量,单应性归一法直接利用投影变换关系.因此,单应性归一法不仅可以使瞳孔

图 10 单应性归一法示意图^[5]Fig.10 Gaze estimation based on homography^[5]

中心作为投影对象,还可以使用瞳孔边缘、角膜边缘等眼睛特征,因为单应性变换允许点、直线和圆锥曲线的投影变换;单应性归一法对外部光源的配置要求更为灵活,只需要检测到 4 个稳定的闪光点即可.另外,使照相机成像面变换到归一化平面上的单应性矩阵 H_I^N 归一化了头部姿态对视线估计的影响,使得单应性归一法对实验对象的头部姿态更为稳健.

单应性归一法基于瞳孔中心与角膜反射点共面的假设,通过单应性矩阵 H_N^S 将归一化平面 Π_N 上的瞳孔中心位置投影到屏幕 Π_S 上的注视点位置.上述假设,以及基于瞳孔中心所在的光轴而非视轴来进行视线估计,是单应性归一法的主要误差来源.为此,Kanai 等^[68] 使用径向基函数对注视点位置进行校正.

为提高单应性归一法的使用范围,Ma 等^[69] 提出将单应性变换矩阵 H_I^N 分别替换为相似变换和仿射变换以适用于只具有 2 个闪光点和 3 个闪光点的视线估计场景. Shin 等^[70] 结合双目约束,在单个红外线光源的场景实现基于单应性归一法的视线估计. Luo 等^[71] 使用眼睛坐标系、照相机成像平面坐标系和屏幕坐标系,通过几何关系求得对应的映射关系,从而提出单相机、单光源场景下基于单应性变换的视线估计方法. Huang 等^[72] 提出一种基于自适应单应性矩阵的头部运动校正方法,通过模拟头部变化来预测头部处于新位置时单应性的变化,以校准头部运动对视线估计的影响,从而提高眼动追踪的鲁棒性.

虽然在单应性归一法中,单应性矩阵 H_I^N 归一化了头部姿态对视线估计的影响,但该方法无法在头部自由运动的情况下进行视线估计;类似交比法地,为使外部光源能在角膜上反射,单应性归一法对实验对象的头部位置也有所限制.此外,单应性

归一法需要进行标定来求得单应性矩阵 H_N^S , 大大地增加了该法的实际应用难度.

3.1.2 基于三维几何的视线估计方法

基于三维几何的方法主要通过拟合三维眼球模型来确定眼球中心、半径等眼睛特征, 再结合各特征之间的几何关系进行视线估计.

Yamazoe 等^[73] 假设眼球为球体、虹膜是沿眼球表面移动的平面圆, 并根据投影关系确定眼球中心与人脸关键点和眼球直径之间的几何关系, 最后由虹膜中心和眼球中心求得视线方向. Chen 等^[74] 通过添加两个眼角的中点和眼球中心来扩展原先由六个人脸关键点组成的传统三维人脸模型, 再据此模型经校准后进行视线估计. Vincent 等^[75] 对二维的眼部区域关键点进行三角划分以获得眼睛网格, 并求解二维瞳孔中心所在的三角网格的重心, 然后在该三角形的重心坐标系中计算瞳孔中心的三维位置, 再结合眼球中心的三维位置求得视线方向. Sesma 等^[76] 基于瞳孔中心及内外眼角的几何位置关系提出瞳孔中心-眼角矢量特征, 然后使用多项式函数进行视线估计. Sun 等^[77] 使用参数化的虹膜模型来求取虹膜中心位置, 然后结合眼球中心的三维位置确定视线方向. Wood 等^[78] 使用主成分分析法提取眼睛的形状和纹理信息, 提出了一个新的三维可变形眼睛模型, 然后通过综合分析拟合出与给定图像尽量相似的图像, 最后从拟合参数中提取出视线方向. 如图 11 所示, Wang 等^[79] 提出的三维变形眼-脸模型对人脸关键点和眼球中心位置之间的几何关系进行了编码, 由此可获得眼球中心和瞳孔中心的三维位置. 图 11 中 CCS 代表相机坐标系, HCS 代表头部坐标系, ECS 代表眼球坐标系. Park 等^[80] 沿用文献 [81] 中将眼球看作是相交的小圆和大圆的假设, 将使用卷积神经网络求得的眼部区域关键点拟合到三维眼睛模型以进行视线估计. Yiu 等^[82] 使用深度全卷积神经网络对眼部图像进行

图像分割以实现瞳孔定位、虹膜检测等, 然后沿用文献 [83] 中将一系列时序眼睛图像拟合三维瞳孔运动模型的算法, 以实现视线估计.

不同于基于二维映射的视线估计方法通过光源确定瞳孔或者虹膜中心、并标定得到眼睛参数, 基于三维几何的视线估计方法更多通过数据统计分析或者机器学习得到眼球参数, 最后利用三维眼球模型的结构先验来实现视线估计或者注视点检测. 基于三维几何的方法综合利用了数据与先验的特性, 目前正在受到越来越多的关注.

3.2 基于特征学习的视线估计方法

随着摄像头传感器、计算机视觉、图像处理、机器学习技术的快速发展, 目前大部分工作主要基于特征学习来实现视线估计. 基于特征学习的方法通过机器学习模型实现高维特征与低维视线之间的映射, 效果依赖于数据的多样性与学习模型的有效性. 根据特征种类的不同, 相关方法可以进一步分为基于几何特征的方法和基于外观纹理的方法.

3.2.1 基于几何特征的视线估计方法

基于几何特征的方法是指通过机器学习模型来实现瞳孔中心位置、人眼或者脸部关键点位置、结构形状等几何特征与视线方向向量的直接映射.

Huang 等^[84] 选取头部姿态以及眼部区域关键点之间的 6 个距离值作为几何特征, 通过支持向量机和随机森林回归模型学习这些特征与视线之间的映射关系来实现视线估计. Zhang 等^[85] 使用瞳孔中心和内眼角来定义瞳孔-眼角距离比率, 然后使用高斯回归模型将该距离比率映射到视线的水平方向. 文献 [86] 基于主动外观模型方法定位眼部区域, 眼部区域的 36 个特征点分别代表眼睛轮廓、虹膜、瞳孔等, 用检测的关键点的坐标几何关系及特征来估计视线方向, 并用支持向量机对视线方向进行准确分类. Chuang 等^[87] 用左右虹膜、鼻子和嘴部等区

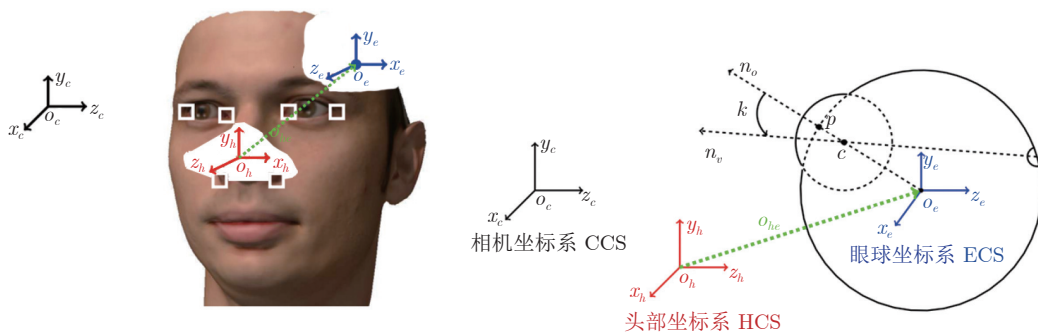


图 11 基于二维关键点及三维眼球模型的视线估计^[79]

Fig. 11 Gaze estimation based on 2D landmarks and 3D eyeball model^[79]

域的 14 个关键点定义了一个新的几何特征, 再据此通过支持向量机检测驾驶员的注视区域. Wang 等^[88]是在检测的瞳孔中心的基础上根据模型估计角膜曲率中心, 得出光轴方向来实现视线估计. 基于瞳孔中心位置等眼睛区域关键点、眼睛图像外观和视线方向存在耦合关系, Cao 等^[89]使用级联回归的方法来同时检测眼睛关键点和估计视线方向. 为了简化视线方向的估计任务, Park 等^[90]首先通过沙漏网络模型在单眼图像检测出瞳孔与眼部两个区域的分割图像, 然后利用视线方向与两个区域形状分布的关联来学习一个深度网络 DenseNet 以实现视线估计.

基于几何特征的方法通过构建眼部几何特征与视线之间的关联关系. 与基于二维映射的视线估计方法不同, 该方法利用图像处理与计算机视觉等方法直接从图像中提取眼睛几何特征, 并通过机器学习实现几何特征与视线方向的模型映射, 减少了个体参数校准环节, 其效果依赖于训练数据与模型的有效性.

3.2.2 基于外观纹理的视线估计方法

基于外观纹理的方法对整个输入图像进行纹理特征提取, 然后学习特征与视线方向之间的映射模型, 以进行视线估计. 早期的基于外观纹理的方法一般是提取人眼图像的手工特征, 例如方向梯度直方图特征或局部二值模式特征等, 再应用最近邻^[91-92]、随机森林回归^[93-96]和支持向量机^[97]等模型来进行视线估计. 随着大量开源数据集的收集和公开, 以卷积神经网络为代表的深度学习模型被大量应用到基于外观纹理的方法中. Yu 等^[98]选取 17 个眼部区域关键点和视线方向的偏航角和俯仰角, 通过提出的约束关键点-视线模型来构建眼部区域关键点位置和视线方向的关系, 最后从模型参数中获取视线方向. Wang 等^[99]认为影响卷积神经网络的视线估计的泛化能力因素是眼睛外观、头部姿态和点估计的过拟合问题, 因此在卷积神经网络框架中引入对抗学习模块来学习对视线方向敏感且同时对眼睛外观和头部姿态变化鲁棒的特征; 并将卷积神经网络扩展到贝叶斯卷积神经网络, 以使用多组参数来改善原先卷积神经网络仅使用一组参数、无法很好地表示复杂情况的点估计过拟合问题. Ali 等^[100]将两只眼睛的图像作为不同通道输入到基于卷积神经网络的视线估计模型中, 并对这两个通道应用不同空间权重以提高视线估计的准确率. 观察到左眼图像、右眼图像和全脸图像在不同场景下对视线估计的结果有着不同的贡献, Zhou 等^[101]提出了具有

自适应性加权策略的视线估计模型, 以依靠更高质量图像来实现更准确的视线估计. Wang 等^[102]提出了从人脸图像提取和以眼睛变形近似的两种方法来获取头部姿态, 再通过 ResNet-34 模型或直接串联的方式将头部姿态特征和视线估计特征结合起来, 从而将头部姿态信息更好地纳入真实场景下的视线估计中. Chen 等^[103]以双眼图像和脸部图像作为输入, 在文献 [104] 提出的神经网络模型基础上, 用扩张卷积代替部分卷积层和最大池化层以保持输入图像的高分辨率, 从而有利于捕捉视线变化时眼睛图像外观的细微变化.

基于特征学习的视线估计方法中, 基于几何特征的方法一般只需要几个校准点、无需大量的训练样本, 就能实现准确率较高的视线估计. 但该方法要求输入图像的分辨率足够高, 以准确提取瞳孔中心、眼角等显著的人眼特征, 且该方法容易受到光照条件等的影响. 另外, 一些关于视线信息的潜在重要特征并未被选取来建立视线估计模型. 基于外观纹理的方法则没有显式地提取特定的人眼结构特征, 而是将整个人脸或人眼图像作为输入并提取高维特征. 该方法操作简单, 除了摄像头、一般无需额外的设备配置; 且对图像分辨率、光照条件等的稳健性良好. 但该方法需要大量的训练数据, 同时所得模型为黑盒模型, 难以解释.

综上所述, 基于模型的视线估计方法需要利用复杂的眼球结构模型, 一般应用于头戴式眼镜或者实验室环境下, 需要进行相应的参数标定. 而基于特征学习的视线估计方法从人眼图像中获取眼睛外观信息或几何特征, 然后通过学习这些特征到人眼视线方向的映射来完成视线方向或者注视点的估计. 随着深度学习、计算机视觉等技术的快速发展, 基于特征学习的视线估计方法取得突破性进展. 相较于基于模型的方法, 基于特征学习的方法鲁棒性更高, 但是其精确度依赖于数据量和模型有效性. 融合利用基于模型先验和基于特征学习两类方法的优势, 是眼动跟踪研究的重点方向之一.

4 数据集及性能比较

基于视频图像分析的眼动跟踪领域已有一定数量的开源公开数据集, 由于眼动跟踪涉及瞳孔中心检测及视线估计两部分, 本节分别对这两个任务的数据集及评价指标进行论述.

4.1 瞳孔中心检测数据集及评价指标

瞳孔中心检测主要数据集见表 1. BioID^[105]和 GI4E^[106]在人眼瞳孔中心检测上使用广泛, 均采集

自实验室环境,利用摄像头捕捉上半身图像,头部姿态范围均较小,其中 BioID 采集 23 名人员的 1521 张图像,分辨率为 384×286 像素,含有闭眼状态;发布于 2013 年的 GI4E 采集了 103 名被试人员的 1236 张图像数据,每个人采集了 12 张不同视线方向的图片,分辨率为 800×600 像素,该数据集还包含了每个测试人员的 1 张正视方向图片. CASIA-Iris^[107] 由中国科学院自动化研究所提出,主要用于虹膜识别,自 2002 年开始 CASIA-Iris 先后被补充完善并发布了 4 个版本,涵盖 70 个国家和地区的超过 3000 名实验对象. 表 1 为 2010 年发布的 CASIA-IrisV4, 包含 54601 张图片. 主要侧重于瞳孔的检测与分割,并偏向于使用头戴式设备进行人眼图像的采集. ExCuSe^[108] 对已有的开源数据集进行了整合,共包含 39001 张图像,用于研究真实环境下瞳孔检测与跟踪,在数据采集过程中,实验对象需要进行驾驶或在超市中寻找物品,瞳孔的位置和环境光会快速变化. Else^[14] 数据集采集了实验对象在驾驶或实验室环境下的眼部图片,共包含 7 个子集,55 712 张分辨率为 384×288 像素的照片,涵盖了眼镜反光、睫毛、眼睑以及环境投影等干扰,对瞳孔检测算法的鲁棒性提出了挑战. LPW^[109] 利用头戴式眼动仪采集了 22 名人员的 66 段视频,共包含 130 856 张图片,覆盖了不同种族、视线、光照的变化,具有一定的挑战性. OpenEDS^[110] 是伦敦大学学院、Facebook 和 Google 共同制作的大规模数据集,使用虚拟现实头戴式设备采集数据,实验在受控光源下进行,采集了分辨率为 400×640 像素的 356 649 张照片,分为 4 个子集,除了对瞳孔进行标注外,还包含眼部关键点、未标记眼部图片、眼动视频帧序列和眼部点云数据. TEyeD^[111] 是目前世界上规模最大的人眼图像开源数据集,图像通过头戴式设备拍摄,采集于乘车、室内活动、室外活动等不同的任务场景,图片数量超过 2 千万,TEyeD 对巩膜、瞳孔、2D/3D 眼部关键点等信息均进行了标注.

对于人眼瞳孔中心检测准确率评价指标,使用广泛的是最大归一化误差^[105],计算公式如下:

$$d_{eye} = \frac{\max(D_{right}, D_{left})}{\|loc_{right} - loc_{left}\|} \quad (1)$$

其中, D_{right} 和 D_{left} 分别是检测的左、右瞳孔中心到真实瞳孔中心的欧氏距离, loc_{left} 和 loc_{right} 分别为标注的真实眼球中心位置. d_{eye} 即为计算得到的最大归一化误差. d_{eye} 的等级代表检测方法准确程度,其值越小代表检测效果越好,其中 $d_{eye} \leq 0.05$ 代表检测到的瞳孔中心位置在瞳孔半径内, $d_{eye} \leq 0.1$ 表示检测到的瞳孔中心位置在虹膜半径区域内. 除了最大归一化误差外,也有相关研究将瞳孔检测视为一个分割任务,通过对比像素级的检测误差或者区域的交并比进行评价.

表 2 和表 3 分别给出了近年来已有方法在 BioID 和 GI4E 数据集上取得的瞳孔中心检测结果. 从表 2 和表 3 可以看出,该领域研究不断取得突破,一些基于手工特征的传统方法^[39, 41, 106, 112] 逐渐被基于深度学习的方法^[34-35, 49] 所超越,其中在 BioID 目前已经取得了 96.7% 的最大归一化误差 $d_{eye} \leq 0.05$,且在 GI4E 上由文献 [100] 更新到 99.8% 的检测率.

综上所述,由于更多研究者的深入研究,以及更多开源人眼标注数据的公开,极大地推动了基于数据驱动的眼动跟踪方法的发展;同时,以深度学习为主的人工智能方法不断取得突破,更为有效的特征表示也极大推动了眼动跟踪领域的快速发展. 除此之外,基于仿真图像与真实图像的融合^[41] 也受到越来越多的关注,结合深度学习的优势,相关研究会进一步推动眼动跟踪的研究,并实现更为广泛的应用.

4.2 视线估计数据集及评价指标

对于视线估计,主要的数据集如表 4 所示. 视线方向和头部姿态的涵盖范围越大,光照条件、遮

表 1 常用瞳孔中心检测数据集
Table 1 Datasets for pupil detection

数据集	发布年份	被试人数	图片/视频数量	图像区域	图像分辨率(像素)
BioID ^[105]	2001	23	图片 1521 张	上半身	384×280
CASIA-Iris ^[107]	2010	≥ 1800	图片 54601 张	人眼、人脸	$320 \times 280, 640 \times 480, 2352 \times 1728$
GI4E ^[106]	2013	103	图片 1236 张	上半身	800×600
ExCuSe ^[108]	2015	未知	图片 39001 张	人眼	$384 \times 288, 620 \times 460$
Else ^[14]	2016	未知	图片 55 712 张	人眼	384×288
LPW ^[109]	2016	22	视频 66 段	人眼	640×480
OpenEDS ^[110]	2019	152	图片 356 649 张	人眼	400×640
TEyeD ^[111]	2021	132	图片 20867 073 张	人眼	$384 \times 288, 320 \times 240, 640 \times 480, 640 \times 360$

表 2 不同方法在 BioID 数据集上的瞳孔中心检测结果对比

Table 2 Comparison of pupil center detection results by different methods on the BioID dataset

方法	$d_{eye} \leq 0.05$ 的检测准确率 (%)	$d_{eye} \leq 0.10$ 的检测准确率 (%)	年份
Aluja等 ^[112]	92.1	98.0	2016
Gou等 ^[30]	91.2	99.4	2017
Choi等 ^[113]	91.1	98.4	2017
Cai等 ^[114]	92.8	—	2018
Levinshtein等 ^[115]	95.3	99.5	2018
Choi等 ^[48]	93.3	96.9	2019
Gou等 ^[41]	92.3	99.1	2019
Xia等 ^[34]	94.4	99.9	2019
Lee等 ^[49]	96.7	99.0	2020

表 3 不同方法在 GI4E 数据集上的瞳孔中心检测结果对比

Table 3 Comparison of pupil center detection results by different methods on the GI4E dataset

方法	$d_{eye} \leq 0.05$ 的检测准确率 (%)	$d_{eye} \leq 0.10$ 的检测准确率 (%)	$d_{eye} \leq 0.25$ 的检测准确率 (%)	年份
Villanueva等 ^[116]	93.9	97.3	98.0	2013
Gou等 ^[30]	94.2	99.1	99.8	2017
Cai等 ^[114]	99.5	—	—	2018
Levinshtein等 ^[115]	99.0	99.9	100	2018
Gou等 ^[41]	98.3	99.8	—	2019
Xia等 ^[34]	99.1	100	100	2019
Lee等 ^[49]	99.8	99.8	100	2020
Hsu等 ^[35]	97.6	99.6	100	2021

挡情况等外观变化越多的视线估计数据集越有利于自然场景下的视线估计。

ColumbiaGaze^[116] 是早期的视线估计数据集, 该数据集要求被试人员观测墙壁上的点, 然后用高

清摄像头 (5184×3456 像素) 记录 58 位被试人员共 5880 张图像的视线方向和头部姿态. 由于每位被试人员的位置是固定不变的且只被记录了 5 种头部姿态以及 21 种不同的视线方向, 所以该数据集中的数据外观变化及数据样本量有限. EYEDIAP^[117] 要求被试人员注视屏幕上的点和悬浮着的三维物体, 然后用 RGB-D 摄像头记录 16 位被试人员共 27 分钟时长的注视数据. RGB-D 摄像头能够获取图像的色彩和深度信息, 但分辨率较低 (640×480 像素). 此前的数据集所收集的视线范围都非常有限, 为了扩大采集到的视线角度范围, UT-multiview^[93] 使用 8 个摄像头来采集 50 位被试人员的 160 个视线方向数据, 采集到了偏航角、俯仰角范围分别为 $\pm 50^\circ$ 、 $\pm 36^\circ$ 的视线方向数据.

为了获取如不同的光照条件等更多的外观变化, 一些数据集通过要求注视移动设备屏幕上的目标点来更高效地获得被试人员在不同场景下的视线数据. GazeCapture^[104] 是麻省理工大学提出的第 1 个大规模视线估计数据集. 该数据集通过众包, 采集并标注了 1474 位被试人员在手机或平板等移动设备屏幕前的、共 2445504 张的视线方向数据, 覆盖了比较多样的背景、光照和姿态变化. 但这些都是用手机/平板等移动设备获得的数据, 所以视线角度范围非常有限. 类似地, MPIIGaze^[118] 采集了 15 位被试人员每天使用笔记本电脑时的图片, 并标注了 6 个人脸关键点、瞳孔中心和视线向量等, 图片的光照条件、人脸外貌有较大变化.

在上述数据集中, 被试人员一般直接面向移动设备的屏幕, 其头部姿态和视线方向都会偏向屏幕所在的地方. 由此得到的头部姿态和视线方向范围非常有限. 为了得到角度范围更大、外观变化更多的视线估计数据集, 一些研究提出合成视线估计数据集^[121]. 由于目前合成的视线估计数据集与真实数

表 4 常用视线估计数据集
Table 4 Datasets for gaze estimation

数据集	人数	图片/视频数量	图像区域	图像分辨率 (像素)	视线角度范围 (偏航角, 俯仰角)	头部姿态范围 (偏航角, 俯仰角)
ColumbiaGaze ^[106]	56	图片 5880 张	全脸	5184×3456	$\pm 15^\circ, \pm 10^\circ$	$\pm 30^\circ, 0^\circ$
EYEDIAP ^[117]	16	视频 94 段	全脸	640×480	$\pm 40^\circ, \pm 30^\circ$	$\pm 40^\circ, \pm 40^\circ$
UT-multiview ^[93]	50	图片 64000 张	全脸	1280×1024	$\pm 50^\circ, \pm 36^\circ$	$\pm 36^\circ, \pm 36^\circ$
GazeCapture ^[104]	1474	图片 2445504 张	全脸	640×480	$\pm 18^\circ, -1.5 \sim +20^\circ$	$\pm 30^\circ, \pm 40^\circ$
MPIIGaze ^[118]	15	图片 213659 张	全脸	未知	$\pm 20^\circ, \pm 20^\circ$	$\pm 25^\circ, -10^\circ \sim +30^\circ$
RT-GENE ^[119]	15	图片 122531 张	全脸	1920×1080	$\pm 40^\circ, \pm 40^\circ$	$\pm 40^\circ, \pm 40^\circ$
Gaze360 ^[120]	238	图片 172000 张	全脸	3382×4096	$\pm 140^\circ, -40^\circ \sim +10^\circ$	$\pm 90^\circ$, 未知
U2Eyes ^[121]	1000	图片 5875000 张	双眼	3840×2160	未知	未知
ETH-Xgaze ^[122]	110	图片 1083492 张	全脸	6000×4000	$\pm 120^\circ, \pm 70^\circ$	$\pm 80^\circ, \pm 80^\circ$

据之间仍然存在较大的差异, 尚未能很好地应用于自然场景下的视线估计, 故继续致力于采集视线方向和头部姿态的涵盖范围更广、外观变化更多、数据量更大的视线估计真实数据集. RT-GENE^[119] 使用深度摄像头追踪被试人员的头部姿态、使用眼镜式眼动仪获取被试人员的视线方向, 采集了 15 位被试人员共 122 531 张的图像数据, 由于该数据集不要求被试人员注视某一目标点, 因此采集到的视线数据更接近自然场景下的数据情况. 但该数据集使用分辨率并不高 (1920×1080 像素) 的深度摄像头追踪被试人员且被试人员距离摄像头较远, 因此图像中的人脸区域分辨率较低; 另外, 还需要使用生成对抗网络对采集得到的图像进行语义修补来去除被试人员佩戴的眼镜式眼动仪. Gaze360^[120] 要求被试人员注视一个移动目标、并用多个摄像头来同时获取多位被试人员的视线方向. 该数据集采集了 238 名 (138 名女性和 100 名男性) 被试人员, 在 5 个室内场景和 2 个室外场景下共 172 000 份视线数据. 受到注视目标的设置以及被试人员对注视目标的遮挡影响, 采集的视线方向俯仰角范围仅为 -40° ~ +10° 且视线方向偏航角数据在其最大值附近的分布较为稀疏. ETH-XGaze^[122] 用 18 个高清摄像头 (6 000×4 000 像素) 来捕捉大范围的头部姿态和视线方向; 同时用 4 个灯箱来设置不同的照明条件. 该数据集记录了 110 位 (47 名女性和 63 名男性) 被试人员, 在 16 种不同照明情况下近 600 个的视线方向, 共采集了 1 083 492 张图像; 采集到的视线方向偏航角、俯仰角范围分别为 ±120°, ±70°.

一般使用估计的视线方向三维向量与真实的视线方向三维向量之间的平均绝对误差作为视线估计准确率的评价指标, 其计算公式如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{\mathbf{g}}_i - \mathbf{g}_i| \quad (2)$$

其中, \mathbf{g} 表示估计的视线方向三维向量, $\hat{\mathbf{g}}$ 表示真实的视线方向三维向量; n 表示测试样本数量, i 表示第 i 个测试样本. 平均绝对误差越小表示对视线方向的估计越准确.

表 5 给出了近年来已有方法在 MPIIGaze 和 EYEDIAP 数据集上取得的视线估计平均绝对误差结果. 从表 5 可以看出, 目前的方法主要基于深度学习模型^[123-125]. 同样由于大量标注数据的采集以及深度模型的快速发展, 视线估计得到了一定突破, 分别在 MPIIGaze 和 EYEDIAP 实现了最好的 4.1° 和 5.3° 的误差. 此外可以看出, 视线估计研究开始处于瓶颈阶段, 提升和改进的程度有限, 需要跳出对深度学习方法的依赖, 从不同的角度, 比如仿

真数据及生成模型等, 来探究视线估计的新方法及理论.

表 5 不同方法在 MPIIGaze 及 EYEDIAP 数据集上的视线估计绝对误差结果对比

方法	MPIIGaze	EYEDIAP	年份
Hierarchical Generative ^[126]	7.5°	15.2°	2018
Dilated-Net ^[127]	4.8°	5.9°	2018
RT-GENE ^[119]	4.3°	5.9°	2019
Faze ^[123]	5.2°	—	2020
FAR-NET ^[124]	4.3°	5.7°	2020
CA-Net ^[125]	4.1°	5.3°	2020

5 眼动追踪的应用

由于眼动跟踪的广泛应用, 越来越多的商业眼动跟踪仪被研发出来. 本文首先列举目前已有的商业化眼动仪, 然后从人机交互、智能驾驶、医疗健康、人因工程四个方面介绍目前眼动跟踪的应用现状.

5.1 商用眼动仪

由于软硬件技术的快速发展及需求的扩增, 目前世界上已有多家公司致力于生产商业眼动仪及相关软件, 主要有 Tobii、SR Research、Seeing Machines、iMotion 等, 表 6 介绍了部分厂商的最新商用眼动仪. Tobii 公司生产的眼动仪在全球被广泛应用, 可以实现现实场景中高效的眼动数据采集, 并提供了眼动眼镜、辅助记录器、红外标记器、Tobii Studio 分析软件等组件, 方便快速配置与使用; 还内置了陀螺仪等传感器来减少头部运动对眼动跟踪的影响. Dikablis Glasses 设计小巧便携, 其独特设计允许佩戴者戴眼睛使用. Smart Eye 眼动仪来自瑞典, 重点应用于智能驾驶场景, 一直致力于汽车工业市场, 拥有沃尔沃、丰田、本田等大量车企客户. 它为用户提供一个远程眼动跟踪系统, 包括视频成像及分析软件, 可实现高精度全帧速率的测量头部和视线. 除了应用于智能驾驶外, Smart Eye 眼动仪还可以用于飞行模拟器、网页设计、阅读研究和心理实验室等多种研究领域. GP3 是一款研究级眼动仪, 支持眼动领域相关学术研究, 支持丰富的应用程序编程接口, 便于开发应用. Looxid-VR 面向虚拟现实应用, 可以在使用者观看虚拟现实画面的同时采集眼动和瞳孔数据, 强大的数据可视化功能使其得到广泛应用. VIVE Pro Eye 内置

表 6 主要眼动仪介绍
Table 6 Introduction to some main eye trackers

眼动仪型号	厂商	类型	特点
Tobii Pro Glasses 3	Tobii	眼镜式	搭载16个红外光源, 配备超广角摄像机, 内置陀螺仪, 具有完整的数据采集、分析、应用程序编程接口功能支持.
EyeLink 1000 Plus	SR Research	遥测式	具有高采样率、低噪声等特点, 允许头部自由运动, 兼容多种第三方数据处理平台, 适用于多种研究人群和场景.
Dikablis Glasses 3	Ergoneers	眼镜式	轻便小巧, 误差范围约 $0.1^\circ \sim 0.3^\circ$, 配备高清摄像机, 配备D-Lab数据分析软件, 可自动分析感兴趣区域.
Smart Eye Pro	Smart Eye	遥测式	可以配置多个摄像头, 自动捕捉面部关键点, 支持视线3D重建, 配备应用程序编程接口与多种第三方数据分析软件.
GP3	Gazepoint	遥测式	误差范围能达到 $0.5^\circ \sim 1^\circ$, 提供开放的标准应用程序编程接口和软件开发工具包, 兼容iMotions的眼动追踪模组.
LooxidVR	Looxid Labs	虚拟现实	可同步采集眼动和瞳孔数据. 支持脑电数据的采集, 配备数据可视化平台, 基于Unity引擎的应用程序编程接口支持定制用户交互界面和特效.
VIVE Pro Eye	HTC和Valve	虚拟现实	可采集眼动数据, 支持可视化. 整套系统融合了顶级的图像、音频、人体工程学硬件设计, 能营造更为真实的虚拟现实体验.

Tobii 眼动跟踪系统, 可以在运行虚拟现实功能的同时采集眼动数据; 还可在使用者注视的区域呈现清晰图像, 并对其他区域进行一定程度的模糊处理, 具有较好的注意力可视化能力. 此外, 还有美国的 Oculus、Magic Leap、微软发布的 HoloLens 等商用虚拟现实智能设备等都集成了眼动跟踪系统, 在交互控制、目标识别、身份验证、健康监测、社交和多人协作等多个领域广泛应用.

5.2 人机交互

人机交互技术赋能人类与机器的交互, 眼动跟踪是实现人机交互的主要方式之一, 相关研究开始于上个世纪 90 年代, 典型任务包括点击、移动屏幕上的物体、在菜单中选择等^[17]. 分析眼动注视情况有助于更快、更准确地判断用户的选择, 优化用户体验. 例如使用手机或者平板电脑的前置摄像头来捕捉人眼及头部的动态信息, 可以分析人脸表情等相关特征, 实现注意力估计^[7, 128-130]. 此外, 基于视线估计的人机交互可以利用眼动信息开发辅助工具来帮助弱势群体表达情感与意识, 提升生活质量. KlaiB 等^[131] 构建的智能家居结合了 Tobii 眼动跟踪设备、Amazon 的 Alexa 智能语音系统以及 Azure 云计算技术. 该智能家具仅通过语音或利用眼动控制屏幕即可完成各项指令, 降低了老人、残疾人等具有特殊需求的人群的使用难度. 眼动也是人机交互界面设计的重要参考, 例如文献 [132-133] 中各项眼动指标常常用于评价航空航天领域人机交互界面的设计效果. 眼动分析也是广告设计和企业营销的重要参考, 大量研究致力于探究如何最大程度吸引消费者的注意力. Lee 等^[134] 利用眼动追踪技术对网络横幅广告的效果和吸引力进行了评价, 发现动画广告并没有比静态广告更具吸引力优势. 文献 [135] 指

出在广告领域, 对于场景感知和视觉搜索的眼动研究处于早期阶段, 眼动跟踪技术将在未来的广告等商业领域起到更大的作用. 图 12 是一种基于眼动的智能展板示例.

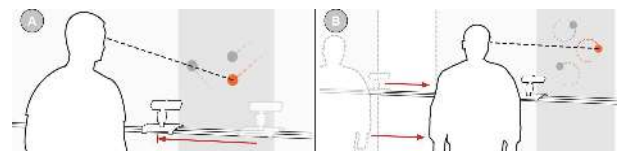


图 12 根据用户眼动自动调整画面的智能展板^[18]
Fig.12 A smart public display using user's eye movement to adjust the content^[18]

5.3 智能驾驶

人是交通系统中最活跃、最不稳定的因素, 绝大多数交通事故的产生都与驾驶员的失误有关, 带来巨大的人身安全威胁和财产损失. 2017 年国家安全监管总局、交通运输部发布的最新研究报告显示, 中国交通事故总量仍居高位, 因此非接触式的眼动跟踪的智能驾驶系统成为了一种有效的解决方案^[55, 136]. 利用眼动信息可以分析驾驶员在具体交通场景中的感知、分析和决策的过程. Rezaei 等^[137] 利用 Haar 特征和卡尔曼滤波构建了一个快速且适应性强的眼动跟踪系统, 可适应驾驶中光照的快速变化. 实际应用中, 营运车辆重量大、操作困难, 对驾驶员的行为进行监控和规范有重要意义. Mandal 等^[138] 提出了一套基于视觉的驾驶员疲劳检测系统, 能简便地部署在公交车等大型车辆上. 该系统融合了人脸识别、眼动追踪等多项技术, 在实际应用中具有较好的鲁棒性. 在自动驾驶应用中, 人机共驾是重要研究课题之一. 眼动研究是人机切换时机的一个重要研究方向. Zeeb 等^[139] 让驾驶员在驾驶时参与

其他任务,以分析驾驶分心情况下驾驶员的反应时间和眼动数据,构建了一个融合模型对驾驶的危险程度进行评价,进一步确定人机驾驶切换时机.如图13所示,Deng等^[140]提出基于驾驶员注视点的显著性图来对驾驶员在驾驶过程中的注意力变换进行建模,探究人类在驾驶过程的认知变化过程,从而构建融合人类驾驶经验的智能驾驶系统.

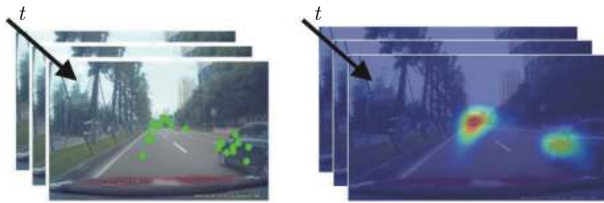


图13 基于注视点的驾驶注意力分析^[140]

Fig.13 Driving attention analysis based on the gaze points^[140]

5.4 医疗健康

眼动跟踪研究在医疗健康方面发挥着重大作用.作为一种生理指标,眼动在精神分裂、抑郁症、睡眠障碍、多动症等疾病的研究与分析中发挥着重要的作用,这部分研究开展较早,相对较为成熟.近年来,眼动追踪研究进一步拓展至医疗图像分析、医疗设备和医疗决策等领域,例如在腹腔镜手术系统中,可以根据主治外科医生当前的视线方向来确定手术电脑屏幕中需要重点关注的区域,从而辅助腹腔镜的自动操作^[141].Chen等^[142]利用放射科医生进行乳腺癌诊断的视线注视点变化来建模描述有经验专家的诊疗过程,从而辅助机器或者临床经验不足的医生进行有效诊断.Khosravan等^[143]提出利用眼动跟踪系统获取放射科医生的眼动数据,构建一个基于注意力机制的方法来进行局部和全局的图像分析,并实际应用于肺癌诊断中.眼动也为医学从业人员培训提供了一种新的角度,医学规培生和专家的视线追踪结果的对比有利于分析诊断决策的过程.Jarodzka等^[144]对专家的眼动进行建模,构建了一个EMME (Eye movement modeling examples)训练系统.该系统可以对专家诊断过程中的眼动进行回放,通过闪光点给予学员提示.研究结论表明,使用EMME的学员在学习过程中注意力得到更好的引导,在测试中表现更为出色.

5.5 人因工程

眼动作为认知科学、行为科学等的重要研究对象,在人因工程应用中越来越受到重视.与传统的心理实验设计相比,眼动跟踪更加具有技术和方法上的优势,可以更加直观、具体地展示人类的兴趣

点与注意力分布情况.有学者研究开发了软件平台或眼动实验设计,尝试进一步推动眼动追踪技术在人因工程上的应用,例如PyGaze^[145]、Eye-MMV^[146]等.在教育领域,Halszka等^[147]总结了众多教育科学的眼动研究后指出,眼动追踪技术在改善基于电脑的多媒体教学课程设计、揭示视觉相关领域专业知识及其发展(例如棋类运动、医学等)、眼动建模学习专业知识三个方面应用较为成功.总之,眼动信息可以表达人的注意力、传达人的情感、表征人类推理过程;较好的眼动分析可以帮助探究人类认知机理,从而构建人在回路的混合增强平行智能^[43, 148-151].

6 结束语

本文对眼动跟踪进行了研究综述,重点阐述了基于视频图像分析的相关研究进展,分别从眼动跟踪发展概述、瞳孔位置检测、视线估计3个方面的研究现状进行详细论述与分析,并介绍了相关应用现状.眼动跟踪是实现强人工智能的基础,可以为探究人类认知机理提供支撑,具有重要的理论意义和良好的应用前景.尽管该领域取得了一系列研究进展,但由于自然环境多样、数据获取与标注费时耗力、个体眼部外观结构差异、姿态动态变化等,眼动跟踪目前仍然是一个具有挑战性的核心难点问题,未来的研究工作可以从以下几个方面展开:

1) 复杂环境下多模态数据融合:在极端环境下,比如低清晰度、低光照等,获取的视频图像数据信息有限,传统的基于图像数据驱动的深度学习方法无法较好实现眼动跟踪与视线估计.其他传感器数据,例如红外图像、深度图像、激光雷达、音频等往往可以提供更多信息,通过多模态数据融合可以大大提升眼动跟踪的精度,从而满足实际应用场景的时效性需求.

2) 小样本下的新型学习理论:实际应用场景中,为了使复杂的深度模型方法更为鲁棒有效,往往需要大规模标注数据集.然而,眼动视线数据的采集标注非常耗时费力.研究小样本情况下的新型学习理论与方法,例如元学习、平行学习、知识迁移、数据对抗生成等方法,对解决小样本情况下的眼动跟踪与视线估计具有重要的研究意义.

3) 注意力建模及认知机理分析:通过深入分析人类视线的注视点,对其注意力变化动态建模,然后基于现有注意力机制实现机器注意力估计,将人类注意力模型和机器注意力模型对比耦合,可以帮助进一步探究人类在感知场景及任务驱动下的认知机理,从而实现人机混合增强智能,推进自动驾驶、

机器人、人机交互的发展与应用.

References

- 1 Hansen D W, Ji Q. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(3): 478–500
- 2 Klaib A F, Alsrehin N O, Melhem W Y, Bashtawi H O, Magableh A A. Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies. *Expert Systems With Applications*, 2021, **166**: 114037
- 3 Kar A, Corcoran P. A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms. *IEEE Access*, 2017, **5**: 16495–16519
- 4 Modi N, Singh J. A review of various state of art eye gaze estimation techniques. *Advances in Computational Intelligence and Communication Technology*. Singapore: Springer, 2021. 501–510
- 5 Liu Jia-Hui, Chi Jian-Nan, Yin Yi-Xin. A review of feature-based gaze tracking methods. *Acta Automatica Sinica*, 2021, **47**(2): 252–277
(刘佳惠, 迟健男, 尹怡欣. 基于特征的视线跟踪方法研究综述. 自动化学报, 2021, **47**(2): 252–277)
- 6 Yu Y, Odobez J M. Unsupervised representation learning for gaze estimation. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 7312–7322
- 7 Zhang R H, Walshe C, Liu Z D, Guan L, Muller K, Whritner J, et al. Atari-head: Atari human eye-tracking and demonstration dataset. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI, 2020, **34**(4): 6811–6820
- 8 Robinson D A. A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Transactions on Bio-Medical Electronics*, 1963, **10**(4): 137–145
- 9 Eibenberger K, Eibenberger B, Rucci M. Design, simulation and evaluation of uniform magnetic field systems for head-free eye movement recordings with scleral search coils. In: Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Orlando, USA: IEEE, 2016. 247–250
- 10 Eibenberger K, Eibenberger B, Roberts D C, Haslwanter T, Carey J P. A novel and inexpensive digital system for eye movement recordings using magnetic scleral search coils. *Medical & Biological Engineering & Computing*, 2016, **54**(2–3): 421–430
- 11 Martinikorena I, Cabeza R, Villanueva A, Urtasun I, Larumbe A. Fast and robust ellipse detection algorithm for head-mounted eye tracking systems. *Machine Vision and Applications*, 2018, **29**(5): 845–860
- 12 Mestre C, Gautier J, Pujol J. Robust eye tracking based on multiple corneal reflections for clinical applications. *Journal of Biomedical Optics*, 2018, **23**(3): 035001
- 13 Santini T, Fuhl W, Geisler D, Kasneci E. EyeRecToo: Open-source software for real-time pervasive head-mounted eye tracking. In: Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2017). Porto, Portugal: SciTePress, 2017. 96–101
- 14 Fuhl W, Santini T C, Kübler T, Kasneci E. Else: Ellipse selection for robust pupil detection in real-world environments. In: Proceedings of the 9th Biennial ACM Symposium on Eye Tracking Research & Applications. South Carolina, Charleston: ACM, 2016. 123–130
- 15 Păsărică A, Bozomitu R G, Cehan V, Lupu R G, Rotariu C. Pupil detection algorithms for eye tracking applications. In: Proceedings of the 21st International Symposium for Design and Technology in Electronic Packaging (SIITME). Brasov, Romania: IEEE, 2015. 161–164
- 16 Chennamma H R, Yuan X H. A survey on eye-gaze tracking techniques. *Indian Journal of Computer Science and Engineering (IJCSSE)*, 2013, **4**(5): 388–393
- 17 Bulling A, Gellersen H. Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing*, 2010, **9**(4): 8–12
- 18 Khamis M, Hoel A, Klimczak A, Reiss M, Alt F, Bulling A. EyeScout: Active eye tracking for position and movement independent gaze interaction with large public displays. In: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology. Québec City, Canada: ACM, 2017. 155–166
- 19 Kong Y S, Lee S, Lee J, Nam Y. A head-mounted goggle-type video-oculography system for vestibular function testing. *EURASIP Journal on Image and Video Processing*, 2018, **2018**(1): 28
- 20 Brousseau B, Rose J, Eizenman M. Hybrid eye-tracking on a smartphone with CNN feature extraction and an infrared 3D model. *Sensors*, 2020, **20**(2): 543
- 21 Yuille A L, Hallinan P W, Cohen D S. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 1992, **8**(2): 99–111
- 22 Hansen D W, Pece A E C. Eye tracking in the wild. *Computer Vision and Image Understanding*, 2005, **98**(1): 155–181
- 23 Valenti R, Gevers T. Accurate eye center location and tracking using isophote curvature. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE, 2008. 1–8
- 24 Valenti R, Gevers T. Accurate eye center location through invariant isocentric patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(9): 1785–1798
- 25 Tan X Y, Song F Y, Zhou Z H, Chen S C. Enhanced pictorial structures for precise eye localization under uncontrolled conditions. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 1621–1628
- 26 Felzenszwalb P F, Huttenlocher D P. Pictorial structures for object recognition. *International Journal of Computer Vision*, 2005, **61**(1): 55–79
- 27 Wang C Y, Zhu Y H, Liu Y F, He R, Sun Z N. Joint iris segmentation and localization using deep multi-task learning framework. arXiv preprint arXiv: 1901.11195, 2019
- 28 Gowroju S, Aarti, Kumar S. Robust deep learning technique: U-net architecture for pupil segmentation. In: Proceedings of the 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). Vancouver, Canada: IEEE, 2020. 609–613
- 29 Araujo G M, Ribeiro F M L, Silva E A B, Goldenstein S K. Fast eye localization without a face model using inner product detectors. In: Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP). Paris, France: IEEE, 2014. 1366–1370
- 30 Wu Y, Ji Q. Learning the deep features for eye detection in uncontrolled conditions. In: Proceedings of the 22nd International Conference on Pattern Recognition. Stockholm, Sweden: IEEE, 2014. 455–459

- 31 Fuhr W, Santini T, Kasneci G, Rosenstiel W, Kasneci E. PupilNet v2.0: Convolutional neural networks for CPU based real time robust pupil detection. arXiv preprint arXiv: 1711.00112, 2017
- 32 Chen S, Liu C J. Clustering-based discriminant analysis for eye detection. *IEEE Transactions on Image Processing*, 2014, **23**(4): 1629–1638
- 33 Lucio D R, Laroca R, Zanlorensi L A, Moreira G, Menotti D. Simultaneous iris and periocular region detection using coarse annotations. In: Proceedings of the 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). Rio de Janeiro, Brazil: IEEE, 2019. 178–185
- 34 Xia Y F, Yu H, Wang F Y. Accurate and robust eye center localization via fully convolutional networks. *IEEE/CAA Journal of Automatica Sinica*, 2019, **6**(5): 1127–1138
- 35 Hsu W Y, Chung C J. A novel eye center localization method for head poses with large rotations. *IEEE Transactions on Image Processing*, 2020, **30**: 1369–1381
- 36 Zhou M C, Wang X Y, Wang H T, Heo J, Nam D. Precise eye localization with improved SDM. In: Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP). Quebec City, Canada: IEEE, 2015. 4466–4470
- 37 Xiong X H, De la Torre F. Supervised descent method and its applications to face alignment. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE, 2013. 532–539
- 38 Gou C, Wu Y, Wang K, Wang F Y, Ji Q. Learning-by-synthesis for accurate eye detection. In: Proceedings of the 23rd International Conference on Pattern Recognition (ICPR). Cancun, Mexico: IEEE, 2016. 3362–3367
- 39 Gou C, Wu Y, Wang K, Wang K F, Wang F Y, Ji Q. A joint cascaded framework for simultaneous eye detection and eye state estimation. *Pattern Recognition*, 2017, **67**: 23–31
- 40 Xiao F, Huang K J, Qiu Y, Shen H B. Accurate iris center localization method using facial landmark, snake, circle fitting and binary connected component. *Multimedia Tools and Applications*, 2018, **77**(19): 25333–25353
- 41 Gou C, Zhang H, Wang K F, Wang F Y, Ji Q. Cascade learning from adversarial synthetic images for accurate pupil detection. *Pattern Recognition*, 2019, **88**: 584–594
- 42 Wang Kun-Feng, Gou Chao, Wang Fei-Yue. Parallel vision: An ACP-based approach to intelligent vision computing. *Acta Automatica Sinica*, 2016, **42**(10): 1490–1500
(王坤峰, 苟超, 王飞跃. 平行视觉: 基于ACP的智能视觉计算方法. *自动化学报*, 2016, **42**(10): 1490–1500)
- 43 Wang K F, Gou C, Zheng N N, Reh J M, Wang F Y. Parallel vision for perception and understanding of complex scenes: Methods, framework, and perspectives. *Artificial Intelligence Review*, 2017, **48**(3): 299–329
- 44 Li Li, Lin Yi-Lun, Cao Dong-Pu, Zheng Nan-Ning, Wang Fei-Yue. Parallel learning—A new framework for machine learning. *Acta Automatica Sinica*, 2017, **43**(1): 1–8
(李力, 林懿伦, 曹东璞, 郑南宁, 王飞跃. 平行学习—机器学习的一个新型理论框架. *自动化学报*, 2017, **43**(1): 1–8)
- 45 Li L, Lin Y L, Zheng N N, Wang F Y. Parallel learning: A perspective and a framework. *IEEE/CAA Journal of Automatica Sinica*, 2017, **4**(3): 389–395
- 46 Shrivastava A, Pfister T, Tuzel O, Susskind J, Wang W D, Webb R. Learning from simulated and unsupervised images through adversarial training. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 2242–2251
- 47 Wang Kun-Feng, Gou Chao, Duan Yan-Jie, Lin Yi-Lun, Zheng Xin-Hu, Wang Fei-Yue. Generative adversarial networks: The state of the art and beyond. *Acta Automatica Sinica*, 2017, **43**(3): 321–332
(王坤峰, 苟超, 段艳杰, 林懿伦, 郑心湖, 王飞跃. 生成式对抗网络GAN的研究进展与展望. *自动化学报*, 2017, **43**(3): 321–332)
- 48 Choi J H, Lee K I, Kim Y C, Song B C. Accurate eye pupil localization using heterogeneous CNN models. In: Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP). Taipei, China: IEEE, 2019. 2179–2183
- 49 Lee K I, Jeon J H, Song B C. Deep Learning-Based Pupil Center Detection for Fast and Accurate Eye Tracking System In: Proceeding of European Conference on Computer Vision. Cham, Switzerland: Springer, 2020. 36–52
- 50 Wang K, Ji Q. 3D gaze estimation without explicit personal calibration. *Pattern Recognition*, 2018, **79**: 216–227
- 51 Guestrin E D, Eizenman M. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering*, 2006, **53**(6): 1124–1133
- 52 Villanueva A, Cabeza R. A novel gaze estimation system with one calibration point. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2008, **38**(4): 1123–1138
- 53 Hennessey C, Nouredin B, Lawrence P. A single camera eye-gaze tracking system with free head motion. In: Proceedings of the 2006 Symposium on Eye Tracking Research & Applications. San Diego, USA: Association for Computing Machinery, 2006. 87–94
- 54 Goss D A, West R W. *Introduction to the Optics of the Eye*. Boston: Butterworth-Heinemann Medical, 2001.
- 55 Liu J H, Chi J N, Hu W X, Wang Z L. 3D model-based gaze tracking via iris features with a single camera and a single light source. *IEEE Transactions on Human-Machine Systems*, 2021, **51**(2): 75–86
- 56 Zhou X L, Cai H B, Li Y F, Liu H H. Two-eye model-based gaze estimation from a Kinect sensor. In: Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA). Singapore: IEEE, 2017. 1646–1653
- 57 Yoo D H, Chung M J. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Computer Vision and Image Understanding*, 2005, **98**(1): 25–51
- 58 Yoo D H, Kim J H, Lee B R, Chung M J. Non-contact eye gaze tracking system by mapping of corneal reflections. In: Proceedings of the 5th IEEE International Conference on Automatic Face Gesture Recognition. Washington, USA: IEEE, 2002. 101–106
- 59 Cheng H, Liu Y Q, Fu W H, Ji Y L, Yang L, Zhao Y. Gazing point dependent eye gaze estimation. *Pattern Recognition*, 2017, **71**: 36–44
- 60 Coutinho F L, Morimoto C H. Free head motion eye gaze tracking using a single camera and multiple light sources. In: Proceedings of the 19th Brazilian Symposium on Computer Graphics and Image Processing. Amazonas, Brazil: IEEE, 2006. 171–178
- 61 Coutinho F L, Morimoto C H. A depth compensation method for cross-ratio based eye tracking. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications. Austin, USA: Association for Computing Machinery, 2010. 137–140
- 62 Hansen D W, Agustin J S, Villanueva A. Homography normalization for robust gaze estimation in uncalibrated setups. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications. Austin, USA: Association for Computing Ma-

- chinery, 2010. 13–20
- 63 Cerrolaza J J, Villanueva A, Villanueva M, Cabeza R. Error characterization and compensation in eye tracking systems. In: Proceedings of the 2012 Symposium on Eye Tracking Research and Applications. Santa Barbara, USA: Association for Computing Machinery, 2012. 205–208
- 64 Arar N M, Gao H, Thiran J P. Towards convenient calibration for cross-ratio based gaze estimation. In: Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE, 2015. 642–648
- 65 Coutinho F L, Morimoto C H. Improving head movement tolerance of cross-ratio based eye trackers. *International Journal of Computer Vision*, 2013, **101**(3): 459–481
- 66 Arar N M, Gao H, Thiran J P. Robust gaze estimation based on adaptive fusion of multiple cameras. In: Proceedings of the 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). Ljubljana, Slovenia: IEEE, 2015. 1–7
- 67 Sasaki M, Nagamatsu T, Takemura K. Screen corner detection using polarization camera for cross-ratio based gaze estimation. In: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications. Denver, USA: Association for Computing Machinery, 2019. 24
- 68 Kanai S, Date H. Homography-based low-cost gaze estimation and its application to the usability assessment of digital prototypes of information appliances. In: Proceedings of the 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Portland, USA: ASME, 2013. 12931
- 69 Ma C F, Baek S J, Choi K A, Ko S J. Improved remote gaze estimation using corneal reflection-adaptive geometric transforms. *Optical Engineering*, 2014, **53**(5): 053112
- 70 Shin Y G, Choi K A, Kim S T, Ko S J. A novel single IR light based gaze estimation method using virtual glints. *IEEE Transactions on Consumer Electronics*, 2015, **61**(2): 254–260
- 71 Luo K Q, Jia X, Xiao H, Liu D M, Peng L, Qiu J, et al. A new gaze estimation method based on homography transformation derived from geometric relationship. *Applied Sciences*, 2020, **10**(24): 9079
- 72 Huang J B, Cai Q, Liu Z C, Ahuja N, Zhang Z Y. Towards accurate and robust cross-ratio based gaze trackers through learning from simulation. In: Proceedings of the 2014 Symposium on Eye Tracking Research and Applications. Safety Harbor, USA: Association for Computing Machinery, 2014. 75–82
- 73 Yamazoe H, Utsumi A, Yonezawa T, Abe S. Remote gaze estimation with a single camera based on facial-feature tracking without special calibration actions. In: Proceedings of the 2008 Symposium on Eye Tracking Research & Applications. Savannah, USA: Association for Computing Machinery, 2008. 245–250
- 74 Chen J X, Ji Q. 3D gaze estimation with a single camera without IR illumination. In: Proceedings of the 19th International Conference on Pattern Recognition. Tampa, USA: IEEE, 2008. 1–4
- 75 Vicente F, Huang Z H, Xiong X H, De la Torre F, Zhang W D, Levi D. Driver gaze tracking and eyes off the road detection system. *IEEE Transactions on Intelligent Transportation Systems*, 2015, **16**(4): 2014–2027
- 76 Sesma L, Villanueva A, Cabeza R. Evaluation of pupil center-eye corner vector for gaze estimation using a web cam. In: Proceedings of the 2012 Symposium on Eye Tracking Research and Applications. Santa Barbara, USA: Association for Computing Machinery, 2012. 217–220
- 77 Sun L, Liu Z C, Sun M T. Real time gaze estimation with a consumer depth camera. *Information Sciences*, 2015, **320**: 346–360
- 78 Wood E, Baltrušaitis T, Morency L P, Robinson P, Bulling A. A 3D morphable eye region model for gaze estimation. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016. 297–313
- 79 Wang K, Ji Q. Real time eye gaze tracking with 3D deformable eye-face model. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 1003–1011
- 80 Park S, Zhang X C, Bulling A, Hilliges O. Learning to find eye region landmarks for remote gaze estimation in unconstrained settings. In: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications. Warsaw, Poland: Association for Computing Machinery, 2018. 21
- 81 Wood E, Baltrušaitis T, Zhang X C, Sugano Y, Robinson P, Bulling A. Rendering of eyes for eye-shape registration and gaze estimation. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 3756–3764
- 82 Yiu Y H, Aboulatta M, Raiser T, Ophey L, Flanagan V L, zu Eulenburg P, et al. DeepVOG: Open-source pupil segmentation and gaze estimation in neuroscience using deep learning. *Journal of Neuroscience Methods*, 2019, **324**: 108307
- 83 Świrski L, Dodgson N. A fully-automatic, temporal approach to single camera, glint-free 3D eye model fitting. *Proc. ECEM*. 2013. 1–10
- 84 Huang M X, Kwok T C K, Ngai G, Leong H V, Chan S C F. Building a self-learning eye gaze model from user interaction data. In: Proceedings of the 22nd ACM international conference on Multimedia. Orlando, USA: Association for Computing Machinery, 2014. 1017–1020
- 85 Zhang Y X, Bulling A, Gellersen H. Pupil-canthy-ratio: A calibration-free method for tracking horizontal gaze direction. In: Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces. Como, Italy: Association for Computing Machinery, 2014. 129–132
- 86 Wu Y L, Yeh C T, Hung W C, Tang C Y. Gaze direction estimation using support vector machine with active appearance model. *Multimedia Tools and Applications*, 2014, **70**(3): 2037–2062
- 87 Chuang M C, Bala R, Bernal E A, Paul P, Burry A. Estimating gaze direction of vehicle drivers using a smartphone camera. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Columbus, USA: IEEE, 2014. 165–170
- 88 Wang K, Ji Q. Real time eye gaze tracking with Kinect. In: Proceedings of the 23rd International Conference on Pattern Recognition (ICPR). Cancun, Mexico: IEEE, 2016. 2752–2757
- 89 Cao L, Gou C, Wang K F, Xiong G, Wang F Y. Gaze-aided eye detection via appearance learning. In: Proceedings of the 24th International Conference on Pattern Recognition (ICPR). Beijing, China: IEEE, 2018. 1965–1970
- 90 Park S, Spurr A, Hilliges O. Deep pictorial gaze estimation. In: Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018. 741–757
- 91 Zhang Y X, Bulling A, Gellersen H. Discrimination of gaze directions using low-level eye image features. In: Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-Based Interaction. Beijing, China: Association for

- Computing Machinery, 2011. 9–14
- 92 Wood E, Baltrušaitis T, Morency L P, Robinson P, Bulling A. Learning an appearance-based gaze estimator from one million synthesised images. In: Proceedings of the 9th Biennial ACM Symposium on Eye Tracking Research & Applications. Charleston, USA: Association for Computing Machinery, 2016. 131–138
- 93 Sugano Y, Matsushita Y, Sato Y. Learning-by-synthesis for appearance-based 3D gaze estimation. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014. 1821–1828
- 94 Wang Y F, Shen T Y, Yuan G L, Bian J M, Fu X P. Appearance-based gaze estimation using deep features and random forest regression. *Knowledge-Based Systems*, 2016, **110**: 293–301
- 95 Kacete A, Séguier R, Collobert M, Royan J. Unconstrained gaze estimation using random forest regression voting. In: Proceedings of the 13th Asian Conference on Computer Vision. Taipei, China: Springer, 2017. 419–432
- 96 Huang Q, Veeraraghavan A, Sabharwal A. TabletGaze: Dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets. *Machine Vision and Applications*, 2017, **28**(5–6): 445–461
- 97 Funes-Mora K A, Odobez J M. Gaze estimation in the 3D space using RGB-D sensors. *International Journal of Computer Vision*, 2016, **118**(2): 194–216
- 98 Yu Y, Liu G, Odobez J M. Deep multitask gaze estimation with a constrained landmark-gaze model. In: Proceedings of the 2018 European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2019. 456–474
- 99 Wang K, Zhao R, Su H, Ji Q. Generalizing eye tracking with Bayesian adversarial learning. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 11899–11908
- 100 Ali A, Kim Y G. Deep fusion for 3D gaze estimation from natural face images using multi-stream CNNs. *IEEE Access*, 2020, **8**: 69212–69221
- 101 Zhou X L, Jiang J Q, Liu Q Q, Fang J W, Chen S Y, Cai H B. Learning a 3D gaze estimator with adaptive weighted strategy. *IEEE Access*, 2020, **8**: 82142–82152
- 102 Wang Z C, Zhao J, Lu C, Huang H, Yang F, Li L J, et al. Learning to detect head movement in unconstrained remote gaze estimation in the wild. In: Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Snowmass, USA: IEEE, 2020. 3432–3441
- 103 Chen Z K, Shi B E. Towards high performance low complexity calibration in appearance based gaze estimation. arXiv preprint arXiv: 2001.09284, 2020
- 104 Krafka K, Khosla A, Kellnhöfer P, Kannan H, Bhandarkar S, Matusik W, et al. Eye tracking for everyone. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 2176–2184
- 105 Jesorsky O, Kirchberg K J, Frischholz R W. Robust face detection using the hausdorff distance. In: Proceedings of the 3rd International Conference on Audio-and Video-Based Biometric Person Authentication. Halmstad, Sweden: Springer, 2001. 90–95
- 106 Smith B A, Yin Q, Feiner S K, Nayar S K. Gaze locking: passive eye contact detection for human-object interaction. In: Proceedings of the 26th annual ACM symposium on User interface software and technology. New York, USA: Association for Computing Machinery, 2013. 271–280
- 107 Alaslani M G, Elrefaai L A. Convolutional neural network based feature extraction for iris recognition. *International Journal of Computer Science & Information Technology (IJCSIT)*, 2018, **10**(2): 65–78
- 108 Fuhr W, Kübler T, Sippel K, Rosenstiel W, Kasneci E. Ex-CuSe: Robust pupil detection in real-world scenarios. In: Proceedings of the 16th International Conference on Computer Analysis of Images and Patterns. Valletta, Malta: Springer, 2015. 39–51
- 109 Tonsen M, Zhang X C, Sugano Y, Bulling A. Labelled pupils in the wild: A dataset for studying pupil detection in unconstrained environments. In: Proceedings of the 9th Biennial ACM Symposium on Eye Tracking Research & Applications. Seattle, USA: Association for Computing Machinery, 2016. 139–142
- 110 Garbin S J, Shen Y R, Schuetz I, Cavin R, Hughes G, Talathi S S. OpenEDS: Open eye dataset. arXiv preprint arXiv: 1905.03702, 2019
- 111 Fuhr W, Kasneci G, Kasneci E. TEyeD: Over 20 million real-world eye images with pupil, eyelid, and iris 2D and 3D segmentations, 2D and 3D landmarks, 3D eyeball, gaze vector, and eye movement types. In: Proceedings of the 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Bari, Italy: IEEE, 2021. 367–375
- 112 Ahuja K, Banerjee R, Nagar S, Dey K, Barbhuiya F. Eye center localization and detection using radial mapping. In: Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP). Phoenix, USA: IEEE, 2016. 3121–3125
- 113 Choi I, Kim D. A variety of local structure patterns and their hybridization for accurate eye detection. *Pattern Recognition*, 2017, **61**: 417–432
- 114 Cai H B, Liu B L, Ju Z J, Thill S, Belpaeme T, Vanderborght B, et al. Accurate eye center localization via hierarchical adaptive convolution. In: Proceedings of the 29th British Machine Vision Conference. Newcastle, UK: British Machine Vision Association Press, 2018. 284
- 115 Levinshstein A, Phung E, Aarabi P. Hybrid eye center localization using cascaded regression and hand-crafted model fitting. *Image and Vision Computing*, 2018, **71**: 17–24
- 116 Villanueva A, Ponz V, Sesma-Sanchez L, Ariz M, Porta S, Cabeza R. Hybrid method based on topography for robust detection of iris center and eye corners. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 2013, **9**(4): 1–20
- 117 Funes Mora K A, Monay F, Odobez J M. EYEDIAP: A database for the development and evaluation of gaze estimation algorithms from RGB and RGB-D cameras. In: Proceedings of the 2014 Symposium on Eye Tracking Research and Applications. Safety Harbor, USA: Association for Computing Machinery, 2014. 255–258
- 118 Zhang X C, Sugano Y, Fritz M, Bulling A. MP II Gaze: Real-world dataset and deep appearance-based gaze estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, **41**(1): 162–175
- 119 Fischer T, Chang H J, Demiris Y. RT-GENE: Real-time eye gaze estimation in natural environments. In: Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018. 339–357
- 120 Kellnhöfer P, Recasens A, Stent S, Matusik W, Torralba A. Gaze360: Physically unconstrained gaze estimation in the wild. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, 2019. 6911–6920

- 121 Porta S, Bossavit B, Cabeza R, Larumbe-Bergera A, Garde G, Villanuova A. U2Eyes: A binocular dataset for eye tracking and gaze estimation. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshops (IC-CVW). Seoul, Korea: IEEE, 2019. 3660–3664
- 122 Zhang X C, Park S, Beeler T, Bradley D, Tang S, Hilliges O. ETH-XGaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation. In: Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer, 2020. 365–381
- 123 Park S, De Mello S, Molchanov P, Iqbal U, Hilliges O, Kautz J. Few-shot adaptive gaze estimation. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, 2019. 9367–9376
- 124 Cheng Y H, Zhang X C, Lu F, Sato Y. Gaze estimation by exploring two-eye asymmetry. *IEEE Transactions on Image Processing*, 2020, **29**: 5259–5272
- 125 Cheng Y H, Huang S Y, Wang F, Qian C, Lu F. A coarse-to-fine adaptive network for appearance-based gaze estimation. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI, 2020. 10623–10630
- 126 Wang K, Zhao R, Ji Q. A hierarchical generative model for eye image synthesis and eye gaze estimation. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 440–448
- 127 Chen Z K, Shi B E. Appearance-based gaze estimation using dilated-convolutions. In: Proceedings of the 14th Asian Conference on Computer Vision. Perth, Australia: Springer, 2019. 309–324
- 128 Zhang X C, Sugano Y, Fritz M, Bulling A. Appearance-based gaze estimation in the wild. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 4511–4520
- 129 Santini T, Fuhr W, Kasneci E. CalibMe: Fast and unsupervised eye tracker calibration for gaze-based pervasive human-computer interaction. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. Denver, USA: Association for Computing Machinery, 2017. 2594–2605
- 130 Wood E, Bulling A. EyeTab: Model-based gaze estimation on unmodified tablet computers. In: Proceedings of the 2014 Symposium on Eye Tracking Research and Applications. Safety Harbor, USA: Association for Computing Machinery, 2014. 207–210
- 131 Klabi A F, Alsrehin N O, Melhem W Y, Bashtawi H O. IoT smart home using eye tracking and voice interfaces for elderly and special needs people. *Journal of Communications*, 2019, **14**(7): 614–621
- 132 Lim Y, Ramasamy S, Gardi A, Kistan T, Sabatini R. Cognitive human-machine interfaces and interactions for unmanned aircraft. *Journal of Intelligent & Robotic Systems*, 2018, **91**(3–4): 755–774
- 133 Pongsakornsathien N, Lim Y X, Gardi A, Hilton S, Planke L, Sabatini R, et al. Sensor networks for aerospace human-machine systems. *Sensors*, 2019, **19**(16): 3465
- 134 Lee J, Ahn J H. Attention to banner ads and their effectiveness: An eye-tracking approach. *International Journal of Electronic Commerce*, 2012, **17**(1): 119–137
- 135 Higgins E, Leininger M, Rayner K. Eye movements when viewing advertisements. *Frontiers in Psychology*, 2014, **5**: 210
- 136 Chen Y Y, Lv Y S, Li Z J, Wang F Y. Long short-term memory model for traffic congestion prediction with online open data. In: Proceedings of the 19th International Conference on Intelligent Transportation Systems (ITSC). Rio de Janeiro, Brazil: IEEE, 2016. 132–137
- 137 Rezaei M, Klette R. Novel adaptive eye detection and tracking for challenging lighting conditions. In: Proceedings of the 2013 Asian Conference on Computer Vision. Daejeon, Korea: Springer, 2013. 427–440
- 138 Mandal B, Li L Y, Wang G S, Lin J. Towards detection of bus driver fatigue based on robust visual analysis of eye state. *IEEE Transactions on Intelligent Transportation Systems*, 2017, **18**(3): 545–557
- 139 Zeeb K, Buchner A, Schrauf M. What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accident Analysis & Prevention*, 2015, **78**: 212–221
- 140 Deng T, Yan H M, Qin L, Ngo T, Manjunath B S. How do drivers allocate their potential attention? Driving fixation prediction via convolutional neural networks. *IEEE Transactions on Intelligent Transportation Systems*, 2020, **21**(5): 2146–2154
- 141 Li S P, Zhang X L, Kim F J, da Silva R D, Gustafson D, Molina W R. Attention-aware robotic laparoscope based on fuzzy interpretation of eye-gaze patterns. *Journal of Medical Devices*, 2015, **9**(4): 041007
- 142 Chen Y, Gale A. Using eye gaze in intelligent interactive imaging training. In: Proceedings of the 2010 Workshop on Eye Gaze in Intelligent Human Machine Interaction. Hong Kong, China: Association for Computing Machinery, 2010. 41–44
- 143 Khosravan N, Celik H, Turkbey B, Jones E C, Wood B, Bagci U. A collaborative computer aided diagnosis (C-CAD) system with eye-tracking, sparse attentional model, and deep learning. *Medical Image Analysis*, 2019, **51**: 101–115
- 144 Jarodzka H, van Gog T, Dorr M, Scheiter K, Gerjets P. Learning to see: Guiding students' attention via a Model's eye movements fosters learning. *Learning and Instruction*, 2013, **25**: 62–70
- 145 Dalmaijer E S, Mathôt S, Van der Stigchel S. PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, 2014, **46**(4): 913–921
- 146 Krassanakis V, Filippakopoulou V, Nakos B. EyeMMV toolbox: An eye movement post-analysis tool based on a two-step spatial dispersion threshold for fixation identification. *Journal of Eye Movement Research*, 2014, **7**(1): 1–10
- 147 Halszka J, Holmqvist K, Gruber H. Eye tracking in Educational Science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research*, 2017, **10**(1), doi: 10.1.3. doi: 10.16910/jemr.10.1.3
- 148 Zheng X L, Zeng D, Li H Q, Wang F Y. Analyzing open-source software systems as complex networks. *Physica A: Statistical Mechanics and its Applications*, 2008, **387**(24): 6190–6200
- 149 Wang Fei-Yue. Parallel system methods for management and control of complex systems. *Control and Decision*, 2004, **19**(5): 485–489
(王飞跃. 平行系统方法与复杂系统的管理和控制. *控制与决策*, 2004, **19**(5): 485–489)
- 150 Wang F Y, Wang X, Li L X, Li L. Steps toward parallel intelligence. *IEEE/CAA Journal of Automatica Sinica*, 2016, **3**(4): 345–348
- 151 Zheng N N, Liu Z Y, Ren P J, Ma Y Q, Chen S T, Yu S Y. Hybrid-augmented intelligence: Collaboration and cognition. *Frontiers of Information Technology & Electronic Engineering*, 2017, **18**(2): 153–179



苟 超 中山大学智能工程学院副教授. 中国科学院大学与美国伦斯勒理工学院联合培养博士. 主要研究方向为计算机视觉和机器学习. 本文通信作者.

E-mail: gouchao@mail.sysu.edu

(**GOU Chao** Associate professor at

the School of Intelligent Engineering, Sun Yat-sen University. Ph.D. at University of Chinese Academy of Sciences, and jointly trained by Rensselaer Polytechnic Institute. His research interest covers computer vision and machine learning. Corresponding author of this paper.)

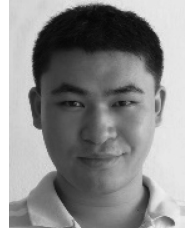


卓 莹 中山大学智能工程学院硕士研究生. 2019 年获西南交通大学交通运输与物流学院学士学位. 主要研究方向为视线估计.

E-mail: zhuoy8@mail2.sysu.edu.cn

(**ZHUO Ying** Master student at the School of Intelligent Engineering,

Sun Yat-sen University. She received her bachelor degree from the School of Transportation and Logistics, Southwest Jiaotong University in 2019. Her main research interest is gaze estimation.)



王 康 英伟达高级算法工程师. 2019 年获美国伦斯勒理工学院电子信息与计算专业博士学位. 主要研究方向为自动驾驶相关的视觉算法.

E-mail: kangwang.kw@gmail.com

(**WANG Kang** Senior software engineer at Nvidia corporation. He received his Ph.D. degree in electrical, computer & system engineering from Rensselaer Polytechnic Institute in 2019. His research interest covers computer vision algorithms for autonomous cars.)



王飞跃 中国科学院自动化研究所研究员, 复杂系统管理与控制国家重点实验室主任, 中国科学院大学中国经济与社会安全研究中心主任, 青岛智能产业技术研究院院长. 主要研究方向为平行系统的方法与应用, 社会计算, 平行智能以及知识自动化.

E-mail: feiyue.wang@ia.ac.cn

(**WANG Fei-Yue** Professor at Institute of Automation, Chinese Academy of Sciences, director of the State Key Laboratory for Management and Control of Complex Systems. Director of China Economic and Social Security Research Center at University of Chinese Acad. President of Qingdao Academy of Intelligent Industries. His research interest covers methods and applications for parallel systems, social computing, parallel intelligence, and knowledge automation.)