

Eye Tracking Based Augmented Reality Human-Computer Interaction Method

Weixing Chen, Xinbo Ji, Xiaoyu Cui*

Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang Liaoning
Email: chen867820261@gmail.com, amani.java@gmail.com, *cuixy@bmie.neu.edu.cn

Received: May 11th, 2019; accepted: May 24th, 2019; published: May 31st, 2019

Abstract

In order to improve the human-computer interaction (HCI) of augmented reality, we combined with the existing optical see-through head-mounted display (OST-HMD). Inspired by the Resnet network idea, we proposed a Heterogeneous nested neural network, which outputs 55 feature points of an eye. On this basis, we respectively get the line of sight vector and the center of the pupil. Two Non-IR cameras are used to track the movement of both eyes in the near eye area. The corneal center error was reduced to 0.66 mm and the angular error was reduced to 0.89°. The method has achieved good results, which lays a foundation for the eye movement interaction of the augmented reality system.

Keywords

Eye Tracking, Deep Learning, Augmented Reality

基于视线追踪的增强现实人机交互方法

陈卫兴, 纪欣伯, 崔笑宇*

东北大学, 中荷生物医学与信息工程学院, 辽宁 沈阳
Email: chen867820261@gmail.com, amani.java@gmail.com, *cuixy@bmie.neu.edu.cn

收稿日期: 2019年5月11日; 录用日期: 2019年5月24日; 发布日期: 2019年5月31日

摘要

本文结合现有光学透视头戴显示设备, 提出了基于视线进行交互的方法。受到Resnet网络思想的启发, 提出了一种异构嵌套的卷积神经网络(Heterogeneous nested neural, HNN), 通过12层卷积和3层全连

*通讯作者。

接层得到一个眼睛的55个特征点,在此基础上分别得出视线向量和瞳孔中心,进而得到了用户视线。在近眼部位采用两个非红外光源相机分别追踪双眼的运动,角膜中心误差降到了0.66 mm,角度误差降到了0.89°。最后本文在HNN中做了不同通道数的对比以及HNN和Resnet网络的对比。实验结果表明,相比于传统方法和普通的卷积神经网络,本文提出的HNN网络对眼睛追踪效果有明显提升。

关键词

视线追踪, 深度学习, 增强现实

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

增强现实是一项将虚拟与现实相结合的技术,它可以应用于医疗、生产生活、工业设计和娱乐等各个行业[1]。随着计算能力的提高和5G技术的发展,增强现实有望成为未来通用的计算平台。光学透视头戴式显示器(Optical See Through-Head Mounted Display, OST-HMD)作为增强现实系统的主要载体,近几年国内外发布的面向消费者的产品日益增多[2],在扩大普及度的同时,如何在OST-HMD上更好地进行人机交互成为目前增强现实领域需要解决的问题。目前增强现实的人机交互(Human-Computer Interaction, HCI)在手势和语音构建得较为完善,微软在2016年推出的HoloLens已经支持了语音交互和手势识别[3],同年德国的pupil labs开发出可装置在HoloLens上的眼追踪相机。2018年magic leap公司发布的混合现实(MR)头显在眼镜内测装置了眼追踪摄像头,但交互方式仍以手势操作和手柄控制为主。

视线追踪技术的不断改善在于提高准确率和帧率的同时降低侵入性和减少数据带宽,从而优化设备的穿戴体验和交互体验。移动技术的发展使视线追踪从桌面端的研究和应用转移到移动端,从而发展普适的视线追踪方法,将视线追踪用于OST-HMD的交互,可以弥补手势识别和语音识别的不足。一方面,佩戴在人眼前的增强现实眼镜与眼动交互具有很高的契合度,通过视线追踪进行交互也将大大提高自由度;另一方面,在手势和语音的操作受到限制的情况下,譬如在一些特殊的应用场景:医生在为病人做手术时,通过眼睛控制OST-HMD的手术导航系统来完成更多必要的操作。此外,视线追踪可以帮助佩戴增强现实眼镜的残障人士(渐冻症患者只有眼球可以运动)完成一些必要的操作,作为提高生活质量的媒介。近几年随着硬件计算能力的提高,增强现实眼镜以及头戴式增强现实设备体积在逐步缩小,增强现实有望成为未来通用的计算平台。眼镜的后置相机捕捉用户视线位置,前置相机捕捉注视环境,进而得到用户在环境中的注视点。用户可以通过佩戴增强现实眼镜进行操作来即时获取所需要的信息,跳过了用户手动查找信息这一步骤,将使人们的工作和生活效率发生质的变化。

本文提出视线追踪方案是通过异构嵌套的神经网络模型解决了上述两个缺点,用普通近焦相机在眼前2 cm的位置拍照,避免了头部运动的干扰的同时,减少了眼部区域的外部图像参数的评定,无需在进行标定,可以直接进行追踪。本文的方法结合了深度学习与神经网络,取得了较好的效果,为增强现实系统的眼动交互奠定了基础。

2. 研究背景

视线追踪的方法可分为侵入式和非侵入式,最早的眼追踪是用巩膜搜索线圈嵌入到眼部区域[4]这种

方法具有高侵入性，对使用者的干扰较大。在过去的几十年中，随着图像处理算法的发展和计算能力的提高，非侵入式的视线追踪在注释分析、情感分析、认知研究等领域和固定于桌面系统的视线追踪方法已经取得了一系列的成果[5]。基于数字视频分析的视线追踪(Video Oculographic, VOG) [5]，仅通过摄像机追踪视线的变化逐渐成为非侵入式的热点研究方法，此后衍生出一系列眼球追踪的图像学方法。其中，使用瞳孔跟踪或角膜反射跟踪(PCCR)已被广泛应用，而扭转分量的测量通常被认为是计算上更为困难的任務，解决该问题的方法包括极坐标互相关方法和虹膜的匹配和跟踪[6]。典型非侵入式的视线追踪包括基于外观的方法和基于形状的方法[5]，以及近些年受到关注的基于深度学习的方法。

2.1. 基于外观的方法

基于外观的方法与眼睛的实际几何特征无关，这种方法通过眼睛区域的光度特性进行检测保留眼睛图像像素的空间和强度信息。依赖于直接建立在眼睛区域外观上的模型，在概念上涉及通过构建图像块模型，通过使用相似性度量的模型匹配来执行眼睛检测来进行模板匹配[7]。基于外观的方法可以进一步分为基于强度和子空间的方法。基于强度的方法直接使用强度或滤波强度图像作为模型，而子空间方法假设眼睛图像的重要信息在较低维度子空间中定义。

2.2. 基于形状的方法

基于形状的方法的分类取决于先验模型是简单的椭圆还是更复杂的性质。一般的边缘检测技术用于提取角膜缘或瞳孔边界[8]。图像中的若干区域可具有与虹膜和瞳孔区域类似的强度分布，因此阈值仅适用于约束设置。通过明确的特征提取后，霍夫变换在一些静态图像中可以有效地检测虹膜和瞳孔边界，但在视频流中产生的效果却不佳，因为视线的变化会导致眼睛的形状产生明显的改变，另一方面，光照强度的变化也会对检测效果产生很大的影响[9]。现有的二维视线追踪系统都存在两个共同的缺点：1) 用户在使用前都需要进行个体标定，以确定视线落点与用户依赖参数的关系；2) 同时由于头部运动和眼球运动的随机性，以及瞳孔运动的非线性，用户需要保持头部静止，没有明显的头部运动[8]。

2.3. 基于深度学习的方法

近年来深度学习与神经网络的发展得较为迅速，以卷积神经网络为代表的面向图像的深度学习算法得到了广泛的应用和优化。因此，基于深度学习的视线追踪方法开始受到人们的关注。台湾大学的 Chih-Fan Hsu 等人[10]通过 CNN 对眼动视频进行了分析和研究，剑桥大学 Erroll Wood 等人[11]开发的 UnityEyes 为训练神经网络提供了完备的数据集。

3. 基于异构嵌套神经网络的视线交互

本文提出了一个异构嵌套的卷积神经网络模型(HNN)。该网络采用 Resnet 网络的思想，基于 tensorflow 1.8 的 GPU 版本进行构建，网络的详细结构如 3.2 所述。数据集来源上文提到的 UnityEyes，根据相机位置和角度调节生成器的参数，共生成 70,000 张眼睛图片。

3.1. 图像预处理

训练前首先进行图像的预处理，两个普通光学相机分别采集左右眼部的原始红绿蓝三通道图像，对图像中的红色通道进行直方图均衡化，增强大多数场景下的图像细节。再在 RGB 空间利用公变换式

$$g(I, j) = a \times f(I, j) + b \quad (1)$$

式(1)中 $a > 1$ 时图像对比度增强，从而突出皮肤与眼球以及眼白与虹膜的色彩差别。最后经过锐化处理，

突出边缘特征，以便于网络的学习，同时调节图像尺寸为 256×256 像素值。

3.2. 异构嵌套神经网络

步骤一：输入图像先经过一层 Batch Normalization(BN)层，一层 7×7 卷积核的卷积(CONV)层，一层修正线性单元(relu)层，再进入由 4 个模块组成的卷积网络中。每个模块都由若干个小网络组成，除了第一个模块的第一个网络外，其余网络的输入皆为上个模块的输出与输入的和。

步骤二：模块中的各个网络由一层 BN 层，一层 3×3 CONV 层与一层 relu 层依次连接而成的结构组成。最后一个模块的输出，一方面经过降维并通过全连接(Fully Connected, FC)层得出 22 个特征点中的虹膜边缘轮廓特征点，另一方面通过一层 BN 层，一层 3×3 CONV 层与一层 relu 层，降维后经过全连接层，得出余下 33 个眼睛轮廓特征点。在得到特征点后再以这 55 个点坐标作为输入，经过 3 个 FC 层得出视线向量。

步骤三：根据 22 个虹膜边缘轮廓的特征点可以得出瞳孔中心，据所述的眼睛 33 个眼睛轮廓特征点识别眼部动作;以全部的 55 个特征点作为输入，经过 3 个 FC 层得出 2 个视线向量;以两视线向量交点确定为空间上人眼的视线焦点的位置。输入到训练好的卷积神经网络中得到各个眼睛的 55 个眼部特征点，瞳孔中心以及视线向量。HNN 网络结构、相应的瞳孔特征点以及输出的视线结果如图 1 所示。

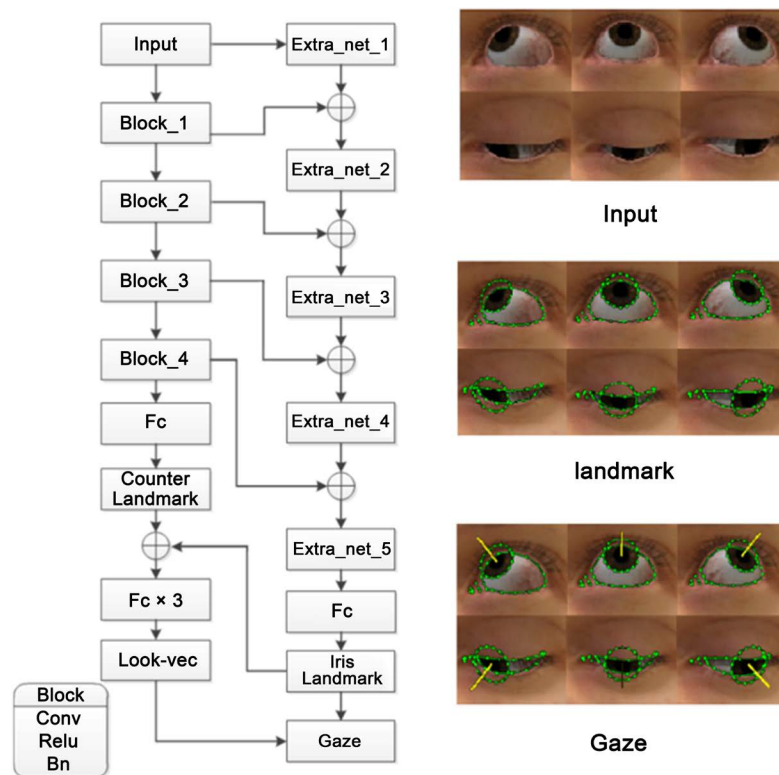


Figure 1. HNN network and eye tracking result
图 1. 异构嵌套神经网络及视线追踪效果

4. 实验环境及结果

4.1. 硬件环境

视线追踪相机为 5 mm 微型 USB 接口摄像头, Sensor 感光面积为 0.2 英寸, 像素尺寸是 $1.75 \mu\text{m} \times 1.75$

μm , 视场角 60° , 分辨率为 1280×780 , 焦距范围在 2 cm 到 5 cm 之间。我们选用 Nvidia Jetson TX2 作为嵌入式开发平台, 增强现实模块选用灵犀微光的光波导光机, 并以此设计出 3D 打印的眼镜模型, 光机装置在镜架两侧, 通过驱动板连接 TX2 的 HDMI 接口, 如图 2 所示。



Figure 2. Hardware environment with OST-HMD and TX2
图 2. 光学头戴式显示器和 TX2 的硬件设备

4.2. 实验结果

我们把训练好的网络在 Nvidia Jetson TX2 进行了测试, 速度达到了 71 fps, 训练过程中的损失函数和特征如图 3 所示。

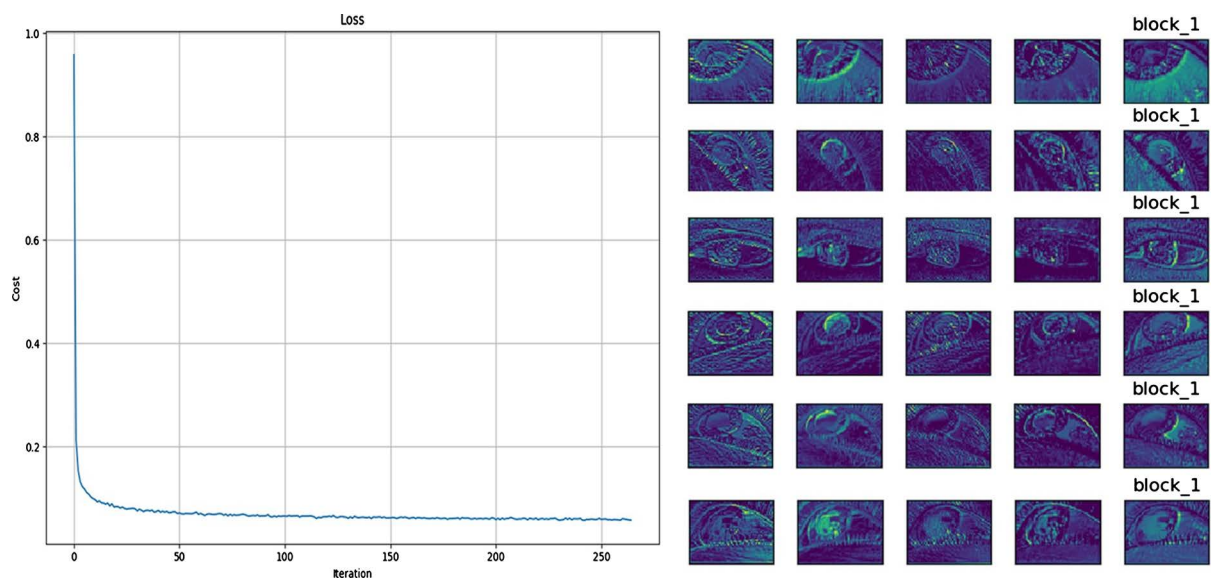


Figure 3. Loss function and feature of training
图 3. 损失函数和训练特征

5. 精度测量和对比分析

5.1. 精度测量

对精确眼动追踪系统的需求也提高了对可靠精确测量方法的需求, 精度测量对于眼动追踪系统的开发以及设备的性能规范至关重要。我们根据[12]提出的眼追踪精度测量方法进行了一系列的实验来评估所

训练不同层数的神经网络的精度误差, 选用大恒 GCM-1101M 系列旋转台搭建光学测试平台来进行精度的测量工作, 具体流程如图 4 所示, 实物如图 5 所示。

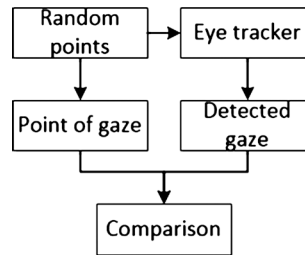


Figure 4. Error measurement flow chart

图 4. 误差测量流程图

测试者在距屏幕前 Z 的距离进行测试, 首先在屏幕上生成两条相交的直线, 在直线上随机取 n 个不等间隔的点作为实际注视点 $a = \{a_0, a_1, a_2, a_3, \dots\}$, 再通过模型的检测输出计算的注视 $b = \{b_0, b_1, b_2, b_3, \dots\}$, 将实际注视点与检测注视点相连接, 计算各组点之间的欧式距离 $L = \{L_0, L_1, L_2, L_3, \dots\}$, 这样得出的相对位移就避免了测量中注视点改变而产生的二次误差。然后测试者扫过各组的连线, 并测量屏幕上瞳孔中心的位移 $x = \{x_0, x_1, x_2, x_3, \dots\}$ 。采集测试者的眼球半径 r 和相机捕捉到的图像中眼球半径 R , 比例尺的计算公式为

$$\gamma = \frac{rx}{RL} \quad (2)$$

眼睛实际的运动距离 $y = \{y_0, y_1, y_2, y_3, \dots\}$

$$y = \frac{rx}{R} \quad (3)$$

角膜的中心误差 α

$$\alpha = \frac{1}{n} \sum_{i=1}^n y_i \quad (4)$$

角度误差 β

$$\beta = \frac{1}{2} \left(\arctan \frac{\sum_{i=1}^n L_i}{nz} + \arctan \frac{\alpha}{\gamma z} \right) \quad (5)$$

测得相机捕捉到的眼球平均半径 $\bar{R} = 4.5 \text{ cm}$, 测试者眼球平均半径 $\bar{r} = 0.5 \text{ cm}$, 再将表 1 中的数据带入(2)式、(4)式和(5)式, 计算出角度误差为 0.89° , 角膜中心误差为 0.66 mm , 明显优于现有的视线追踪方法。

5.2. 对比分析

我们按照最优网络生成的方法进行了参照和对比, 分别具有相同模型不同通道数的对比和不同模型之间的对比。其中, 异构嵌套网络(HNN)分为 16 通道, 32 通道和 64 通道进行对比, Resnet 网络分为 18 通道和 34 通道, 五个模型的对比显示, 32 通道为最优解, 如图 4 所示。检测方法参照[12], 注释点距屏幕的距离为 55 cm , 模拟人眼位于平台中心点, 即在平面坐标轴的原点处, 通过测量模拟眼球分别对准检测位置和注视位置的水平和垂直旋转角度计算眼球实际的移动距离, 对比结果显示 32 通道数的 HNN 网络具有最优效果, 平均误差降到 10%以下, 如图 6 所示。



Figure 5. Error measurement hardware
图 5. 误差测量实物图

Table 1. Corneal center error assessment
表 1. 角膜中心误差评价

标记距离 L (cm)	测量距离 X (cm)	实际距离 Y (cm)
3.04	0.9	0.1
3.05	0.81	0.09
1.72	0.51	0.5667
1.58	0.48	0.5333
0.81	0.3	0.2556
1.65	0.51	0.0567
2.23	0.66	0.0733
2.21	0.65	0.0722
2.71	0.83	0.0922
1.89	0.56	0.0622
1.92	0.53	0.0589
2.88	0.87	0.0967
0.46	0.132	0.0147
2.53	0.75	0.0833
1.84	0.55	0.0617

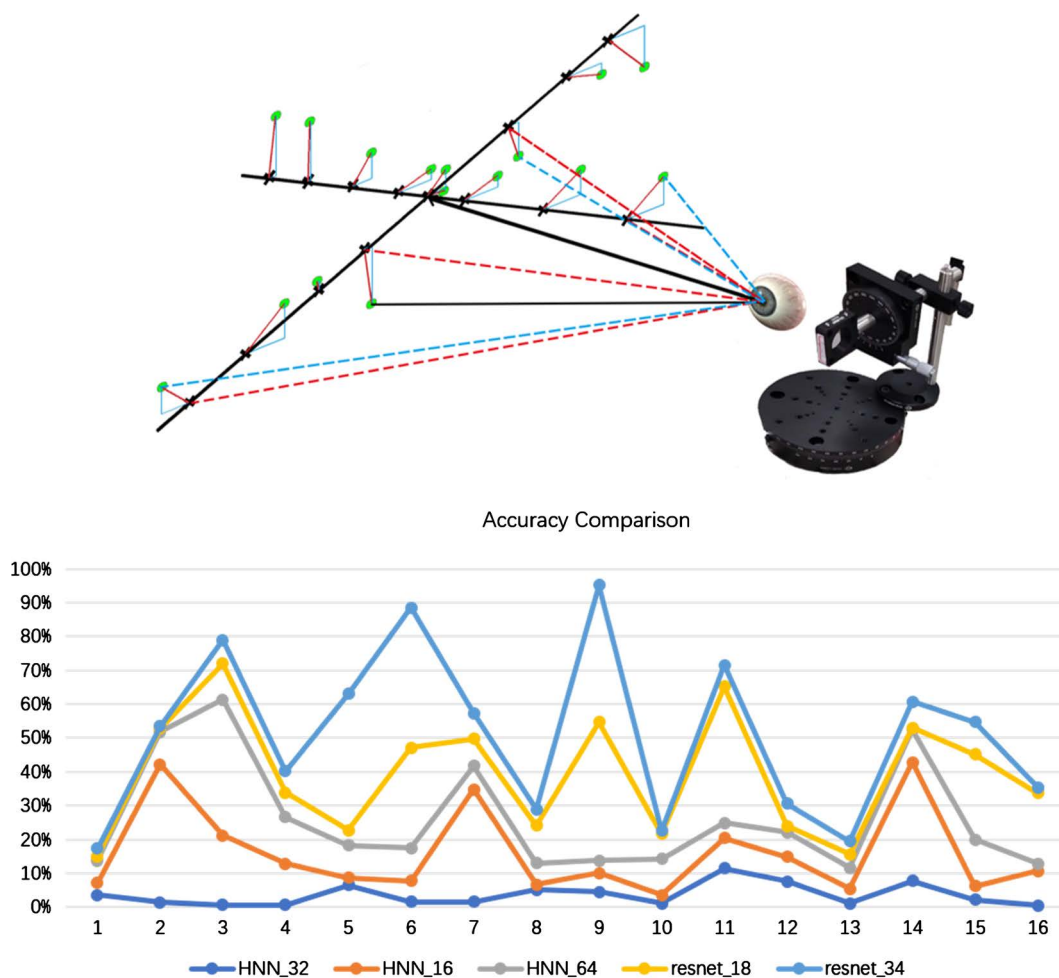


Figure 6. Comparative experiment and accuracy result
图 6. 对比实验和准确率结果

6. 总结

本文针对头戴式增强现实系统的交互需求和现有的不足, 将视线引入 OST-HMD 的交互设计。本文对比了现有的一些视线追踪的方法, 采用深度学习中残差网络 Restnet 网络的思想, 提出了可高效识别瞳孔中心和视线的异构嵌套神经网络, 省去了添加红外光源、标定等步骤的同时将角度误差降到了 0.89° 。我们计划设计 20 种以上眼球运动轨迹交互指令, 同时实现 OST-HMD 系统的自动标定, 并 FPGA 上设计出高集成度的嵌入式增强现实人机交互系统。未来我们会将视线注视点及其所在的场景相结合, 并进一步研究 SLAM 算法和 CBIR 算法, 将所看到的事物的相关信息展示在用户眼前。

基金项目

本文获得东北大学 2018 国家级大学生创新创业计划资助项目(2018101180200)资助。

参考文献

- [1] Grubert, J., Itoh, Y., Moser, K. and Swan, J.E. (2018) A Survey of Calibration Methods for Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics*, **24**, 2649-2662. <https://doi.org/10.1109/TVCG.2017.2754257>

- [2] del Amo, I.F., Erkoyuncu, J.A., Roy, R. and Wilding, S. (2018) Augmented Reality in Maintenance: An Information-Centred Design Framework. *Procedia Manufacturing*, **19**, 148-155. <https://doi.org/10.1016/j.promfg.2018.01.021>
- [3] Kassner, M., Patera, W. and Bulling, A. (2014) Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-Based Interaction. *CoRR*. <https://doi.org/10.1145/2638728.2641695>
- [4] Cristina, S. and Camilleri, K.P. (2018) Unobtrusive and Pervasive Video-Based Eye-Gaze Tracking. *Image and Vision Computing*, **74**, 21-40. <https://doi.org/10.1016/j.imavis.2018.04.002>
- [5] 张闯, 迟健男, 张朝晖, 王志良. 一种新的基于瞳孔 - 角膜反射技术的视线追踪方法[J]. 计算机学报, 2010, 33(7): 1272-1285.
- [6] Schreiber, K. and Haslwanter, T. (2004) Improving Calibration of 3-D Video Oculography Systems. *IEEE Transactions on Biomedical Engineering*, **51**, 676-679. <https://doi.org/10.1109/TBME.2003.821025>
- [7] Hansen, D.W. and Ji, Q. (2010) In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**, 478-500. <https://doi.org/10.1109/TPAMI.2009.30>
- [8] Zhu, Z., Ji, Q. and Fujimura, K. (2002) Combining Kalman Filtering and Mean Shift for Real Time Eye Tracking. *Proceedings of the International Conference on Pattern Recognition*, **4**, 318-321.
- [9] Sasson, N.J. and Elison, J.T. (2012) Eye Tracking Young Children with Autism. *Journal of Visualized Experiments*, **61**, 3675. <https://doi.org/10.3791/3675>
- [10] Hsu, C.-F., Chen, Y.-C., Wang, Y.-S., Lei, C.-L. and Chen, K.-T. (2018) Realizing the Real-Time Gaze Redirection System with Convolutional Neural Network. *MMSys'18: 9th ACM Multimedia Systems Conference*, Amsterdam, 4 p.
- [11] Wood, E., Baltrušaitis, T., Morency, L.-P., Robinson, P. and Bulling, A. (2016) Learning an Appearance-Based Gaze Estimator from One Million Synthesised Images. *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA'16)*, ACM, New York, 131-138. <https://doi.org/10.1145/2857491.2857492>
- [12] Wyder, S. and Cattin, P.C. (2018) Eye Tracker Accuracy: Quantitative Evaluation of the Invisible Eye Center Location. *International Journal of Computer Assisted Radiology and Surgery*, **13**, 1651-1660. <https://doi.org/10.1007/s11548-018-1808-5>

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: csa@hanspub.org