

基于动态轨迹的眼动跟踪隐式标定方法

程时伟 齐文杰

(浙江工业大学计算机科学与技术学院 杭州 310023)

摘要 针对现有多点标定方案耗时长和简化标定方案注视精度较差的局限,提出一种眼动跟踪隐式标定方法,使得眼动跟踪系统只需采集少量样本即可建立准确的映射关系。该方法分为 3 个步骤:首先,标定数据采集,让用户视线跟随动态轨迹运动,记录这一过程中用户眼部图像特征和标定点之间的映射点对。然后,提出合理化的异常值去除方法以自动消除样本噪声,并选择最佳点对集合建立映射模型。对眼动跟踪数据的采集进行延时处理,减少了运动轨迹产生的误差。进一步对样本进行降噪时,排除瞳孔误差数据,并采用随机采样一致算法进一步筛选样本。最后,结合免标定和单点标定这两种方法,在后续标定过程中进一步简化隐式标定过程,并测试了隐式标定的最佳参数。实验表明,在视距为 60 cm 时,该方法的标定时间为 8 s,平均精度为 2.0° ;在隐式标定原型系统中,对于已标定的用户,通过读取其映射模型,即可免标定地快速获取注视点坐标,所需时间为 2 s,平均标定精度为 2.47° ;对于进行眼动跟踪的新用户,通过单点标定方法计算个体差异补偿模型,获取注视点坐标,所需时间为 3 s,平均标定精度为 2.49° ,进一步提高了该方法的实用性。

关键词 眼动跟踪,隐式标定,免标定,注视点,人机交互

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.08.047

Dynamic Trajectory Based Implicit Calibration Method for Eye Tracking

CHENG Shi-wei QI Wen-jie

(School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract Aiming at the limitations of the existing multi-point calibration schemes, such as time-consuming and poor gaze accuracy of simplified calibration schemes, this paper proposed an implicit calibration method for eye tracking, which makes the eye tracking system only need a few samples to establish an accurate mapping relationship. The method has three steps. Firstly, it collects calibration data. With user's gaze following the dynamic trajectory, it records the mapping points between the image of user's eyes and the calibration points in this process. Then, a rationalized outlier removal method is proposed to automatically eliminate the sample noise and select the best point pair to establish a mapping model. The acquisition of eye movement data is delayed, which can reduce the error caused by the dynamic trajectory. Furthermore, when the noise data of the sample are removed, a method for eliminating the pupil error data is proposed, and the sample data are further filtered by Random Sample Consensus (RANSAC) algorithm. Finally, the two methods of calibration-free and single-point calibration are combined to simplify the subsequence implicit calibration process. Experiment results show that the average calibration time is 8 s, and the average accuracy is 2.0° of visual angle when visual distance is 60 cm. In the simplified implicit calibration prototype system, for the user who is calibrated, the coordinates of the fixation are obtained by the calibration-free method. The average calibration time is 2 s, and the average calibration accuracy is 2.47° of visual angle. For the new user who performs eye tracking, the individual difference compensation model is calculated by the single point calibration method to obtain the coordinates of the fixation. The average calibration time is 3 s, and the average calibration accuracy is 2.49° of visual angle, which further improves the practicality of the implicit calibration method.

Keywords Eye tracking, Implicit calibration, Calibration free, Fixation, Human-computer interaction

1 引言

随着计算设备性能的提高和成本的降低,眼动跟踪技术

已经日渐成熟。例如,头戴式眼动跟踪系统由于具有非侵入性、灵活性和移动性,已成为面向普适计算应用的理想选择^[1]。然而,已有的眼动跟踪系统仍然需要在每次使用之前

进行多点标定,标定过程需要耗费相当多的时间来记录标定数据,且用户需要频繁注视不同的标定点,长时间的标定容易导致用户眼睛疲劳,产生错误标定信息,这些已经成为妨碍眼动跟踪技术广泛应用的主要因素之一^[2]。

在多点标定方案存在各类问题的情况下,减少标定点数量、简化标定模型的方案被提出。减少标定点数量,即通过少量标定点来获取足够多的标定数据,已有研究通过使用基于回归^[2]、基于外观^[3]的方法,或使用眼球模型^[4],将标定点数量减少到两点、单点,或是免除标定点;简化标定模型,即使用隐藏的标定点,结合基于回归^[5-6]、基于外观^[7-8]的方法,采集注视-标定点关系的映射点对,计算出标定点坐标。但是,目前的简化标定的方案存在各种问题,例如,减少标定点数量后,注视精度较差,需要的硬件设备增多;采用隐藏的标定点时,由于用户轨迹跟踪能力的约束,注视点在轨迹高速移动时滞后于标定点;由于设备计算能力的限制,无法实时处理大量点对。

在上述研究背景下,针对多点标定方案和已有简化标定方案的局限性,本文提出了基于动态轨迹的隐式标定方法。用户视线跟随标定点的运动轨迹,使得眼动跟踪系统只需采集少量样本即可建立准确的映射关系。对于数据中包含的噪声数据,通过研究瞳孔变化进行降噪,并使用改进的随机采样一致(Random Sample Consensus, RANSAC)算法进一步排除噪声数据,最后将免标定方法和单点标定方法应用到隐式标定原型系统中。可以根据用户是否标定过,选择不同的标定方式:对于已经隐式标定过的用户,在初次标定之后可以免除后续的标定过程;对于尚未标定的用户,只需要单点标定即可计算出用户注视点坐标,有效简化了隐式标定过程。本文通过实验探索了动态轨迹的最优参数设置,实现了准确的标定。快速和准确的标定是实现自然人机交互的关键^[9],本文方法简化了标定过程,给用户带来了更加便捷的交互体验。

2 相关工作

眼动跟踪的核心步骤是标定,它将产生用户眼部图像特征到注视点坐标的映射函数^[10]。在标定过程中,绝大多数研究采用多点标定法,传统隐式标定方法的优点是解决了在多点标定过程中看错某个标定点会影响整个映射模型精确性的问题,具体说明如下。

Pfeuffer 等^[11]提出了一种平滑跟随的标定方法,在标定过程中用户眼睛跟随沿着轨迹运动的图标,摄像头会拍下此过程中的眼部图像。通过计时器来保证采集眼部图像的同时采集相应时刻的目标轨迹的坐标,然后根据时间戳对齐眼部图像与目标轨迹坐标。Sugano 等^[7]提出一种注视点估计方法 AggreGaze。首先在公共大屏端播放视频片段,使用基于外观模型的方法记录用户脸部数据,然后基于不同标定点轨迹收集映射点对,最终得到的注视精度在 $0.6^{\circ} \sim 1.4^{\circ}$ 不等。Sugano 和 Bulling^[12]使用眼部输入特征加上在场景中被计算出的显著图来完成标定。

另一方面,一些标定方法试图通过改变标定程序来适应

用户的需求。例如,Pfeuffer 等^[11]在显示器中以已知轨迹移动的物体来标定放置在屏幕下的桌面式眼动仪,达到了 0.6° (标准差为 0.1°) 的平均精度。与其相似,Huang 等^[13]提出利用用户和计算设备之间的交互事件来收集标定点样本,达到了 2.56° 的精度。在驾驶场景中,Bernet 等^[14]采用自动检测的自定义标定点(由两个嵌套的黑色方块组成),避免了助理和实验用户之间的协调,但是该方法没有考虑深度变化或头部旋转。对于无噪声的模拟数据,其平均精度为 2.22 pixel (标准差为 1.42 pixel),视距为 2.5 m ;实际场景中的精度未给出。

Santini 等^[15]介绍了一种头戴式跟踪系统隐式标定的方法 CalibMe,这是一种利用收集标记(自动检测基准标记)的新方法,用户盯着屏幕中标定点的轨迹,可以通过眼动仪收集大量的标定点对,然后移除其中误差较大的点对,并以快速无监督的方法对保留的点对进行评估。首先,放置的参考点覆盖受试者预期的视觉运动范围;之后,助理和用户合作收集标定点,助理负责检查在整个标定过程中是否正确检测到眼睛特征(主要是瞳孔中心)并计算标定后的注视估计精度。CalibMe 达到了 0.59° (标准差为 0.23°) 的平均精度,对 9 点标定的精度达到了 0.82° (标准差为 0.15°),从而验证了该方法的有效性。其不足之处是当轨迹运动过快时,注视误差较大。现有工作还存在一些局限:1)通过大量注视-标定点点对来获取足够多的标定数据,计算资源消耗大;2)没有深入研究如何高效地对噪声数据进行剔除;3)在后续的标定过程中,对于能否简化隐式标定过程,没有展开深入研究。针对上述局限,本文提出一种基于动态轨迹的隐式标定方法,用户眼睛跟随动态轨迹运动,只需采集少量样本即可建立准确的映射关系;为了提高隐式标定系统的精度,对于映射点对中包含的噪声数据,提出两种剔除噪声数据的方法;为了简化隐式标定系统的使用,将免标定方法和单点标定方法应用到隐式标定原型系统中。与其他方法的比较实验,验证了本文方法不仅简化了眼动数据的收集,还能减少标定时间,保证足够的准确度。

3 隐式标定方法的原理

隐式标定的流程框架如图 1 所示,主要分为以下 3 个步骤:标定数据采集、噪声数据过滤、注视点计算。首先,将标定点定义为特征鲜明的标记,动态收集用户眼睛位置和注视点之间的映射点对,以进行误差修正和评估,同时对眼动数据进行延时采集,减小注视-标定点关系的误差;然后,允许用户在标定过程中眨眼,或是中断系统标定,提出合理化的异常值去除方法以自动消除样本噪声,并自动选择最佳点对集合建立映射模型;最后,通过多项式方法计算注视点,并进行一系列实验来测试动态轨迹的最佳参数。

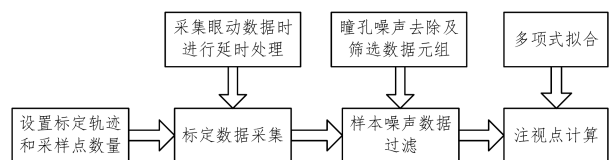


图 1 隐式标定的流程框架

Fig. 1 Process framework of implicit calibration

3.1 标定数据采集

3.1.1 标定点的设置

在标定过程中,用户眼睛跟随屏幕中标定点的轨迹运动。在注视-标定点关系中要使用的标定点应具有以下属性:1)用户应该能够容易地定位和区分要注视的标定点;2)标定点检测应该准确并且只需要较少的计算资源,因为它必须与其他图像处理算法结合在标定系统中实时运行;3)如果对标定点的移动速度未做限制,当其运动速度过快时会出现模糊效果。标定点边框越厚,允许的运动速度越快,因此标定过程也越快。

标定过程中设计有专用的标定点,此标定点可以通过特定的图像处理算法,例如霍夫变换^[16],来保证自动检测到标定点在场景图像中的位置。本文选用了4种标定图标作为标定点,如图2所示。此外,本文还通过增加标定点边框的厚度来减轻移动过程中标定点的模糊程度。

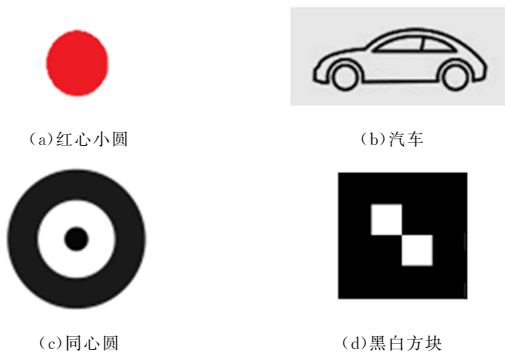


图2 选用的标定点(电子版为彩色)

Fig. 2 Selected calibration point

3.1.2 动态轨迹的选择

如果要获取多组注视-标定点信息,标定点或头部必然有一方需要移动,以获取眼部图像到标定点的对应信息。本文选取移动标定点,通过设置轨迹路径,使得映射点对的分布较为均匀。

本文设计了几种标定点动态轨迹,收集用户眼部特征和注视点坐标之间的关系点阵列,用于进行标定和评估。主要从以下两个方面来设计动态轨迹:1)由于在轨迹运动的过程中,用户眼睛需要一直注视屏幕中的标定点,因此动态轨迹需要考虑用户眼睛的舒适性;2)标定点可以沿着任意路径运动,为了保证标定系统的精确性,动态轨迹的运动范围最好能覆盖整块屏幕,而且要保证场景摄像头能检测到屏幕中的标定点。在点对选择的问题上,采样点应具有代表性,即采样点应该包含屏幕边缘和中间区域,并且采样点应该在用户注视误差较小的区域。

标定开始前,用户头部正对并注视屏幕中的标定点,然后标定点沿轨迹运动。隐式标定选用的动态轨迹如图3所示。

在标定开始前,用户头部先微微抬高,以注视屏幕左上角位置。

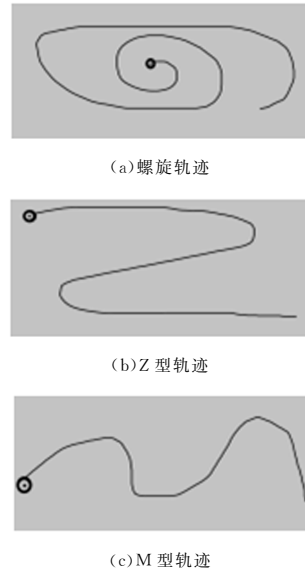


图3 隐式标定选用的动态轨迹

Fig. 3 Selected dynamic trajectory of implicit calibration

标定时间与摄像机帧率相关,如果摄像机帧率较高,则可以在较短时间内采集大量注视-标定点数据^[17]。本文使用的眼动仪的采样频率为每秒15~25帧,大概需要8~12s的时间采集标定数据,然后根据动态轨迹设置标定数据的采集位置。可以在标定期间动态收集120~300个采样点,为了减少设备的计算量,本文只采集少量样本进行映射关系计算,数量在40个以内。

3.2 噪声数据过滤

眼部摄像头和场景摄像头在时间戳 t 处采集数据,从眼部图像中检测到了瞳孔中心坐标 (p_x, p_y) 、普洱钦斑坐标 (r_x, r_y) ,以及瞳孔椭圆的长半轴 L_{axis} 、短半轴 S_{axis} ;类似地,从场景图像中采集轨迹标定点,其中心坐标为 (t_x, t_y) 。设 D 为系统在每个时间戳生成的映射点对:

$$D = (t, p_x, p_y, r_x, r_y, L_{axis}, S_{axis}, t_x, t_y) \quad (1)$$

标定数据采集阶段收集了包含瞳孔中心 (p_x, p_y) 和标定点 (t_x, t_y) 的点对,以此建立瞳孔映射到注视点位置的函数。用户1采集到的部分映射点对数据如表1所列,其中第一列是标定点坐标,在轨迹运行期间采样;第四列为采样的时间,前一个数据为屏幕图像的采样时间,后一个数据为眼部图像的采样时间,对屏幕图像采样时间与眼睛图像采样时间进行同步处理,以防止图像采集时间的错位。标定时间为第一个到最后一个映射点对的时间戳差值。

表1 用户1采集到的部分映射点对

Table 1 Partial mapping point pairs collected by user 1

标定点坐标	瞳孔中心/反光点中心	PCCR 向量	采样时间	映射点对 D
第1个 (683, 384)	$e_1(369.21, 220.91)$	$(19.21, -1.09)$	2019-03-03 20:44:39	$D_1 = (2019-03-03 20:44:39, 369.21, 220.91, 350.00, 222.00, 33.68, 26.10, 683, 384)$
	$r_1(350.00, 222.00)$		2019-03-03 20:44:39	
第2个 (20, 20)	$e_2(379.92, 257.28)$	$(24.92, 16.78)$	2019-03-03 20:44:40	$D_2 = (2019-03-03 20:44:40, 379.92, 257.28, 355.00, 240.50, 35.93, 25.70, 20, 20)$
	$r_2(355.00, 240.50)$		2019-03-03 20:44:40	

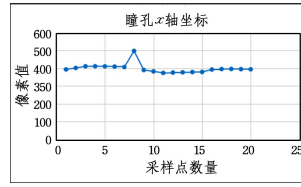
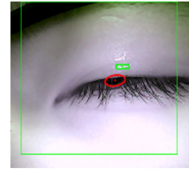
由于受人眼轨迹跟踪能力约束、眼球抖动、瞳孔定位误差等各方面因素的影响,采集到的映射点对有以下缺点:1)含有眨眼、闭眼导致的噪声数据,如果使用这些噪声数据计算注视点坐标,将会造成错误的映射关系;2)映射点对本身也会产生误差,会影响注视点的计算精度;3)由于存在误差点对,采样过少也会导致映射误差进一步扩大。

在进行隐式标定时,每个时间戳只能采集一次数据,无法在数据采集阶段避免误差点对的问题,因此必须对噪声数据过滤。对于采集到的映射点对,首先去除瞳孔误差数据,然后在利用最小二乘法计算映射关系之前,使用 RANSAC 算法迭代寻找最优映射模型,具体阐述如下。

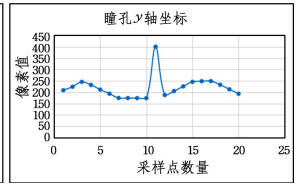
3.2.1 瞳孔误差数据的排除

瞳孔产生误差数据有两方面原因。第一种是眨眼或闭眼造成的,如图 4(a)和图 4(b)所示,采集到用户 1 标定时存在异常的瞳孔数据。在标定期间,瞳孔大小可能由于生理或心理因素(如受到惊吓)而改变,或者由于环境因素(如光照强度)而随之改变。正常情况下,瞳孔中心坐标的梯度变化范围很小,如果梯度变化幅度较大,说明采集到的瞳孔坐标为噪声数据,对于这类噪声数据,可以直接排除当前瞳孔坐标所对应的映射点对。第二种是由于用户走神导致某段眼动数据采集滞后或丢失造成的。因此,如果在顺序采样序列中某映射点对和下一映射点对的梯度相差较大,不能直接去除这两组数据。如图 4(c)和图 4(d)所示,用户在第 15 组和第 16 组数据中,瞳孔 x 轴和 y 轴坐标的梯度变大,到第 19 组和 20 组数据时梯度恢复正常,此时的处理方法是保留第 15 组数据,剔除从第 16 组到第 18 组的数据,因为这段采样数据是瞳孔数据从误差产生到消除的过程。初始时,所有采样得到的映射点对都被认为是没有误差数据的,然后,本文计算映射点对中瞳孔坐标的平均值 u 和标准差 α ,如果坐标出现在 $u \pm 3\alpha$ 范围之外,其会被认为是噪声。重复该判断过程,直到所有采样点检测完成。

用户 1 眨眼造成瞳孔异常数据

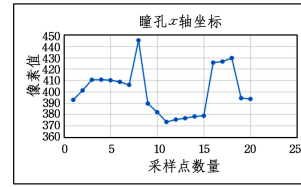
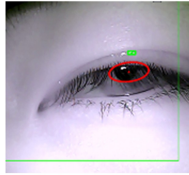


(a) x 轴坐标

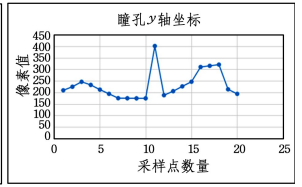


(b) y 轴坐标

用户 2 标定时走神造成某段样本出现噪声



(c) x 轴坐标



(d) y 轴坐标

图 4 用户瞳孔异常值的排除

Fig. 4 Exclusion of abnormal values for user pupils

3.2.2 基于 RANSAC 的数据过滤

仅通过排除瞳孔坐标噪声数据并不能完全剔除误差映射点对,误差较小的映射点对仍被保留下来。对于包含噪声数据的样本,有不同的处理方法,如最小二乘法、随机采样一致性法等,各处理方法的对比如表 2 所列。

表 2 不同方法处理噪声数据的比较

Table 2 Comparison of different methods for processing noise data

处理方法	特点	适用场景	优缺点
最小二乘法 ^[18]	是对所有点的最佳拟合	测量误差符合期望为 0 的高斯分布	如果噪声数据较多,计算出的映射模型不是最优的
随机采样一致性 ^[8,15]	迭代选择数据集中的随机子集来得到数学模型	在包含一定噪声点的数据集中获得最优解	若迭代次数设置上限,得到的模型可能不是最优的;已经有足够好的模型可以结束迭代过程,节省时间
总结	1)当映射点对中中局外点(即噪声点)比例较小且噪声点的偏差较小时,两种方法的注视点估计精度相差不大 2)当映射点对中中局外点较多时,使用 RANSAC 算法剔除噪声点的效果较好		

在剔除明显的误差数据之后,本文采用随机抽样一致性方法对映射点对再次进行筛选。因为它能在包含一定噪声点的数据集中获得最优解^[19],很适合本文的数据集。算法思想是通过迭代数据集中的随机子集估计映射模型,计算满足映射模型的内部点,然后根据内部点继续评估样本集,直至找到被评估为最优的数学模型。本文以 40 组映射点对为例来说明该算法的主要步骤。

- 1)从映射点对 D 中随机选择 10 组作为评估样本;
- 2)从剩下的 30 组映射点对中随机选择 10 组,根据这 10

组点对计算映射模型;

3)使用上述映射模型测试其余 20 组映射点对,误差在阈值内的点则为局内点;

4)如果局内点数量大于测试数据的 50%,则使用所有局内点再次计算映射模型,否则转到步骤 2)继续执行;

5)随机选择 10 组评估样本在新模型中测试,并以标准差评估该模型,保存最优的点,否则转步骤 2)继续执行。

该算法运行的结果要么因为局内点太少表现不及上一次的模型而被舍弃,要么因为比现有的模型更好而被选用。在

迭代过程中,若已经有足够好的映射模型,则可以结束算法迭代过程,以节省时间。隐式标定中映射点对的噪声数据去除效果如表3所列,本实验比较了排除瞳孔误差与RANSAC排除样本误差对最终得到的映射点对的影响。实验中,动态轨迹的运动速度为400 pixel/s,此速度通过标定点运动像素值

乘以帧率换算得到,轨迹为螺旋型。共有3名用户参与实验,在标定过程中,允许用户在某段时间眨眼或是走神,眼睛不再跟随目标轨迹运动,以便收集带噪声数据的样本。在标定结束后,用户依次注视屏幕中的16个测试点,计算标定误差。

表3 隐式标定的映射点对中噪声数据的移除

Table 3 Removal of noise data in implicitly calibrated data tuples

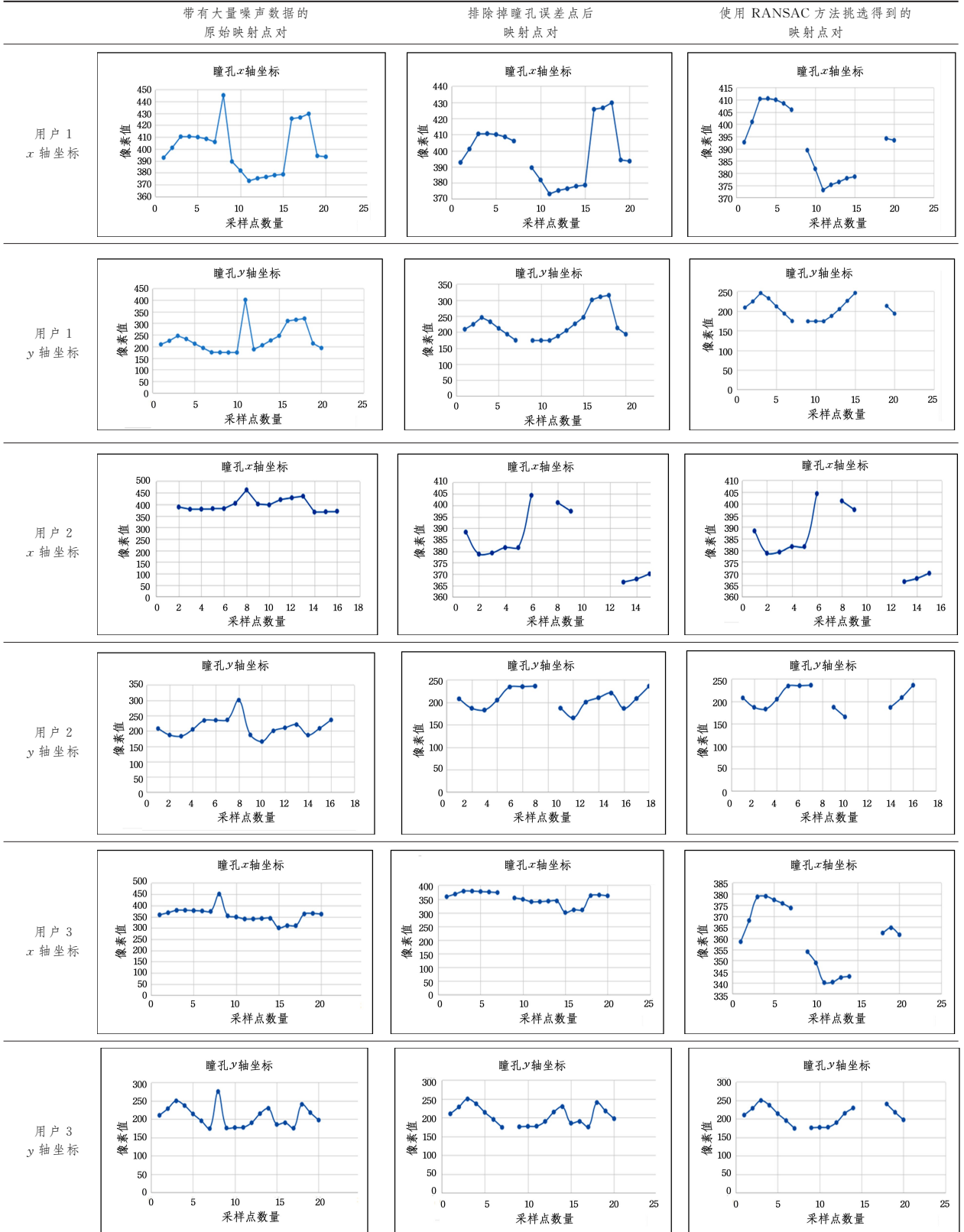


表 3 第二列为包含大量噪声数据的原始映射点对。第三列为排除瞳孔误差点对后的映射点对,此时瞳孔误差数据清除完成,部分误差较小的映射点对被保留下来。第四列使用随机采样一致性算法找出满足最优映射模型的映射点对。由表 3 可知,本文的噪声数据过滤方法具有良好的效果,排除误差点对后的映射点对能明显减少噪声数据。

3.3 注视点计算

注视点计算为隐式标定方法的第三个步骤,即对排除噪声数据的样本建立映射关系。常用的映射函数拟合方法有:直接线性变换法(Direct Linear Transformation, DLT)^[20]、多

项式法^[21]、神经网络法^[22-24]等。针对各类方法的适用场景,选择最适合的拟合方法。经过上一节的噪声数据过滤后,现在保留的映射点对的精确度较高。

3.3.1 映射关系的拟合

各类映射函数的拟合方法对比如表 4 所列。由表可知,最适合隐式标定映射函数的拟合方法是多项式法,从二次多项式开始可以反映部分注视-标定点数据的非线性关系,因此其具有较高的精度;但是高次多项式需要更多的标定点数目,标定时间长,而且容易过拟合,并不实用^[25]。本文选择二次多项式进行映射关系的拟合。

表 4 各类映射关系计算方法的对比

Table 4 Comparison of various mapping relationship calculation methods

映射计算方法	特点	细节	缺点
线性变换	不考虑非线性因素,计算代价小	$X'=HX$,至少需要 4 个映射点对(8 个方程)	只能求近似解,误差较大
多项式法	可以反映映射点对的非线性关系,精度较高	二次多项式被广泛应用于非线性关系的求解,12 个未知系数	高次多项式系数多,所需标定点数量多,容易过拟合
神经网络	对已有样本集进行自主学习,可以模拟非线性关系,容错性强	输入层:瞳孔坐标和普洱钦斑坐标;输出层:注视点在屏幕中的坐标;隐层神经元;计算映射关系,激活函数为 sigmoid	对激活函数依赖性强;样本集数据量大,训练时间长,无法满足实时性

Morimoto 等^[26]使用二次多项式模型模拟非线性函数,取得的精度较高,其映射模型如下:

$$\begin{cases} X = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy \\ Y = b_0 + b_1x + b_2y + b_3x^2 + b_4y^2 + b_5xy \end{cases} \quad (2)$$

其中, a_0, \dots, a_5 和 b_0, \dots, b_5 这 12 个参数为映射模型系数。

至少需要 6 个映射点对数据才能求解式(2)。将过滤了噪声的点数据代入映射模型,即可计算得到注视点信息;同时,提取初始标定过程中人眼图像瞳孔中心坐标 (x_1, y_1) 和普洱钦斑坐标 (rx, ry) ,并将这两个坐标及映射函数系数存入数据库,以用户姓名首字母命名创建该用户映射模型的记录。用户在每次使用此标定系统前,只要输入姓名首字母进行查询,即可确定其是否初次标定,从而确定有没有与其匹配的映射模型。在后续的标定过程中可以直接使用该映射模型,无需重新进行多项式拟合过程。

3.3.2 原型系统的实现

为了验证本文动态轨迹方法的可行性和有效性,本文设计开发了隐式标定原型系统。在使用隐式标定系统时,用户可以选择不同的标定点图标、标定速度和动态轨迹,如表 5 所列。

表 5 隐式标定系统的参数

Table 5 Implicit calibration parameters

标定点图标	标定速度/(pixel/s)	动态轨迹
红心小圆	200	螺旋轨迹
汽车	300	Z 型轨迹
同心圆	400	M 型轨迹
黑白方块	500	无

本文在隐式标定原型系统中使用了两种简化标定方法。

1)免标定方法,其原理是对于已经标定过的用户,通过匹配数据库中该用户的映射模型,免除后续的标定过程。方法执行流程如图 5 所示,首先,基于前期工作完成对眼部图像和场景图像的处理^[27];然后,通过查询数据库来判断用户是否

标定过,若标定过则免标定得到注视点信息;最后,建立头部运动模型,减小眼动跟踪过程中因头动造成的误差。

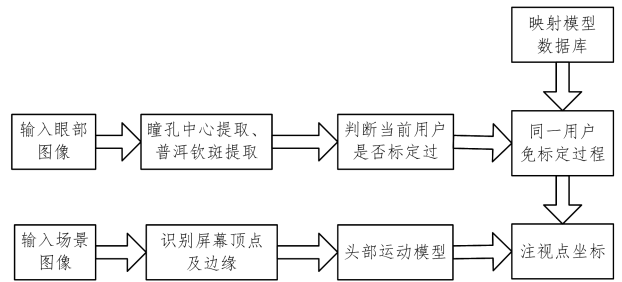


图 5 免标定方法的流程

Fig. 5 Process for calibration-free method

2)单点标定方法,其原理是通过单点标定采集用户眼部图像,并通过旋转、平移等操作补偿当前用户与已有基准用户瞳孔-普洱钦斑向量的差异,然后匹配基准用户的映射模型,快速计算出用户注视点。方法流程如图 6 所示。首先,新用户用户眼部图像时,通过单点标定提取用户的眼部图像特征,并将其与已经收集的基准用户眼部图像进行比较;然后,找出两者眼部特征的差异,从而建立个体差异模型;最后,当前用户使用基准用户的映射模型进行注视点的计算。

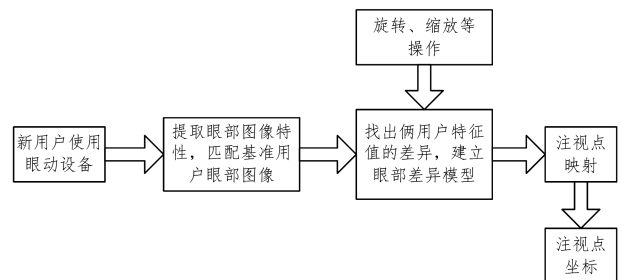


图 6 单点标定方法的流程

Fig. 6 Process for single-point calibration method

隐式标定原型系统的执行流程如图7所示。用户使用系统时可以分为3种情形:1)当用户B首次使用隐式标定系统时,系统通过动态轨迹记录注视-标定点点对,同时记录此用户注视初始标定点的眼动数据以及映射模型,将这些数据保存在数据库中,此时数据库中用户B的数据为眼部PCCR向量、映射模型的12个系数;2)新用户A使用该原型系统时,由于数据库中不存在该用户的眼动数据及映射模型,因此需要通过单点标定求出用户注视中间标定点时的眼部PCCR向

量,建立用户A与基准用户B的个体差异补偿模型,然后使用用户B的映射模型快速计算出用户注视点;3)当用户B再次使用原型系统时,通过调整头部位置,使用用户B回到初次隐式标定时的头部位置,然后在数据库中根据用户id查找当前用户的映射模型,免除标定过程。由上文可知,免标定方法和单点标定方法使用已标定用户的映射模型,因此能保持较高的精度。该原型系统有效简化了隐式标定过程,加强了系统的实用性。

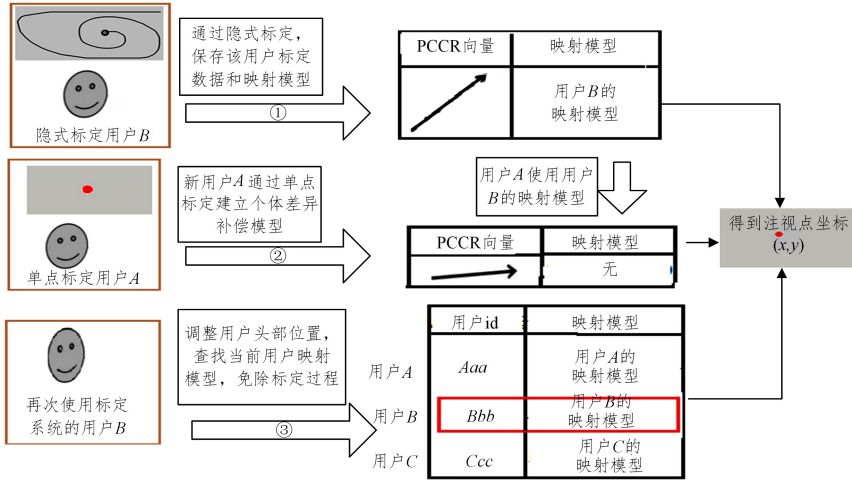


图7 隐式标定原型系统

Fig. 7 Implicit calibration prototype system

原型系统的精度如表6所列。共招募4名用户使用该原型系统,并记录用户使用3种标定方法的平均标定时间。由表6可知,当视距为60cm,标定点运动速度为400pixel/s时,用户初次使用该系统进行隐式标定的平均精度为2.01°,平均标定时间为8s;后续使用隐式标定原型系统时,对于保存了映

射模型的已标定用户,可以执行免标定方法获取注视点坐标,平均标定精度为2.47°,平均标定时间为2s;对于使用隐式标定系统的新用户,可以执行单点标定方法,平均标定精度为2.49°,平均标定时间为3s。可见:后续用户使用免标定方法或单点标定方法时,也能保持较高的标定精度,并能减少标定时间。

表6 隐式标定原型系统的精度(视距为60cm)

Table 6 Accuracy of prototype system of implicit calibration(visual distance is 60cm)

标定方法	平均标定时间/s	平均标定精度/pixel	平均标定精度/(°)	说明
隐式标定	8	85	2.01	初次标定的用户
免标定	2	105	2.47	保存了映射模型的已标定用户
单点标定	3	106	2.49	使用隐式标定系统的新用户

4 实验与分析

共招募3名用户(年龄23~27岁,无色盲、色弱等视力问题)来测试基于动态轨迹的隐式标定方法的精度,设置了动态轨迹的最优参数,并与其他隐式标定眼动跟踪方法进行了比较。

4.1 眼动跟踪精度测试

首先设置隐式标定中的各类参数,包括标定点图标的选择、动态轨迹的选择、轨迹运动的速度。视距为60cm时,标定点图标以及3种轨迹对精度的影响如表7所列,测试过程中保证了眼部图像与轨迹坐标的时间戳同步,同时因用户眨眼、走神导致的误差数据在噪声数据过滤阶段已被剔除。

分析实验结果可知:1)在各类标定点图标下,使用螺旋型或Z型得到的标定效果较好,平均精度可达到2°;2)标定点初始位置应该优先选择位于屏幕中心的位置,从而以更自然的位置开始标定。本文后续实验使用的动态轨迹默认为螺旋

型,并从屏幕中心开始向外扩散运动。

表7 标定点图标以及3种轨迹对精度的影响(视距为60cm)

Table 7 Effect of calibration icons and three tracks on accuracy

(visual distance is 60cm)

(单位:°)

图标	螺旋型精度	Z型精度	M型精度
红心小圆	2.01	2.11	3.42
汽车	2.82	2.49	3.41
双重圆	2.07	2.09	3.97
黑白方块	2.24	2.35	3.53

表8列出轨迹运动速度对精度的影响。前置条件是视距为60cm,采用螺旋型轨迹,并设置了9个采样点。人眼平滑尾随跟踪速度在一般情况下为40~1200pixel/s^[28],因此,如果动态轨迹速度超过1200pixel/s,人眼将无法顺利跟踪动态轨迹而完成标定。另一方面,除了平滑尾随跟踪外,人眼会试图追上轨迹的运动,存在追赶性眼跳。因此,为避免标定数据

中掺杂追赶性眼跳等噪声数据,需要将动态轨迹速度限定在 1200 pixel/s 的范围内。轨迹运动速度过快时,标定所需时间减少,但导致眼睛无法跟踪轨迹,误差增大,标定精度下降;轨迹运动速度过慢时,标定所需时间明显增加,用户会产生厌烦情绪,体验较差。因此,动态轨迹速度设置为 300 pixel/s 或 400 pixel/s,轨迹运动时间分别对应 10 s 和 8 s 时,标定效果最好。眼睛注视运动的轨迹时,由于人眼轨迹跟踪能力的约束,用户注视点在轨迹较高速移动时滞后于标定点,导致不太准确的注视-标定点关系。本文给出了隐式标定优化方法:1)降低轨迹运动速度;2)延迟 5~50 ms 采集当前眼部图像,结合两者求出现有条件下的最优解。如表 9 所列,本文测试了轨迹运动速度分别为 300 pixel/s 和 400 pixel/s 时,延时采集眼动数据的时间与标定精度的关系。

表 8 轨迹运动速度对精度的影响

Table 8 Effect of trajectory speed on accuracy

轨迹运动速度 / (pixel/s)	标定所需时间/s	标定精度/(°)
200	14	2.24
300	10	2.01
400	8	2.11
500	7	2.35(眼动数据延时采集 15 ms)
600	6	2.82(眼动数据延时采集 15 ms)

表 9 在轨迹运动速度固定时,眼睛参数延迟采集对精度的影响

Table 9 Effect of delayed acquisition of eye parameters on accuracy when trajectory velocity is fixed

轨迹运动速度 / (pixel/s)	采样延迟时间/ms	标定精度/pixel	标定精度/(°)
300	无	101	2.37
300	2	102	2.40
300	5	95	2.24
300	10	125	2.90
300	20	137	3.20
300	30	142	3.34
300	50	140	3.29
300	100	516	12.14
400	无	102	2.40
400	2	95	2.23
400	5	88	2.07
400	10	95	2.24
400	20	147	3.46
400	30	156	3.67
400	50	150	3.53

从表 8 和表 9 可以看出:1)轨迹运动速度与眼动数据的延迟采集时间成反比,轨迹运动速度越快,所需的延迟采集时间越短;2)考虑到用户标定时的舒适性,标定速度为 300 pixel/s 时,对应标定精度最高的延迟时间为 5 ms;3)在初始轨迹触发阶段,动态轨迹暂停 1 s 来让人眼捕获标定点,防止初始标定时眼睛跟不上轨迹。

本文进一步验证了映射点对数量对精度的影响,结果如表 10 所列,在轨迹运动速度为 400 pixel/s 时,采样点的个数与精度之间没有明显的对应关系;但是,若采样个数较少,映射点对中出现误差点对,会导致建立的映射关系出现误差的概率增大。

表 10 映射点对数量对精度的影响

Table 10 Effect of mapping point pairs on accuracy

映射点对数量/个	注视精度/pixel	注视精度/(°)
9	88	2.07
16	97	2.25
30	94	2.15

本文标定系统的硬件部分如图 8 所示,两个摄像头和红外光源被固定在眼镜架上,其中眼睛摄像头放置在设备左眼处的前方偏下位置,以确保采集到准确的眼部图像。场景摄像头放置在设备右眼处,且与视线方向平行,以便于拍摄准确的场景图像。标定系统软件基于 MFC+OpenCV 开发,该系统可以将眼部摄像头和场景摄像头采集到的数据传输至电脑端进行处理,电脑处理器为 i7 6700,内存为 8GB,实时图像处理的平均速度为 30 fps。



图 8 眼动设备的硬件部分

Fig. 8 Hardware of eye tracking system

眼动跟踪过程如图 9 所示,用户首先注视屏幕中沿螺旋型轨迹运动的标定点,在标定结束后,用户依次凝视屏幕中顺序出现的 16 个标定点用于精度测试。用户 1 在视距为 60cm 时隐式标定精度的测试结果如图 10 所示,线段长度表示误差范围,线段越长,误差越大。该用户隐式标定精度为 1.5°,标定时间为 8 s。



图 9 眼动跟踪过程

Fig. 9 Eye tracking process



图 10 离屏幕 60 cm 时的隐式标定精度

Fig. 10 Implicit calibration accuracy when visual distance is 60 cm

4.2 与其他方法的比较

如表 11 所列,将本文方法与已有的其他眼动跟踪方法进行比较。文献[15]的方法需要两名用户配合使用,通过标定

后判断数据集中心的评估点,达到了 0.8° 的平均精度,但无法在显示屏中展示标定点,当轨迹运动过快时,注视误差较大。文献[7]在初始时,使用基于外观模型眼动跟踪方法的精度较差,为 6.5° ,通过注视一段标定点轨迹实现了对注视位置的误差补偿,达到 1.4° 的平均精度,但相比基于回归的方法,基于外观的注视估计方法训练数据量大,使用标定系统的新用户

均需注视标定点轨迹实现误差补偿。文献[29]使用了基于模板匹配的方法,通过注视显著图,使用K-最近邻算法求出与新用户最相似的注视模板,然后对注视模板进行平移、缩放等操作,将当前用户注视点映射到其他个体,平均精度为 4.9° 。此方法的不足之处是通过显著图不能准确计算出用户注视模式的差异,导致精度较低。

表 11 所提方法与其他眼动跟踪方法的比较

Table 11 Comparison of this method with other eye tracking methods

设备类型	方法	平均精度/($^\circ$)	技术特征	标定点数量	头动范围
头戴式眼动仪+纸板	文献[15]	无法展示注视点	基于 AOI 的图标 + EyeRecToo 开源软件	注视标定点轨迹	通过移动头部建立映射关系
电脑端	文献[7]	1.4	外观模型 + 多模式卷积神经网络	注视标定点轨迹	允许头部小范围移动
电脑端	本文	2.0	头戴式双相机单光源 + 回归方法	注视标定点轨迹	头部固定
电脑端	文献[29]	4.9	桌面式眼动仪,单摄像机,基于用户注视模板匹配	注视标定点轨迹	头部固定
电脑端	文献[13]	2.56	利用用户和计算设备之间的交互事件来收集标定样本	交互事件	未作说明

本文提出的基于动态轨迹的隐式标定系统,突破了传统隐式标定方法存在的局限性,只需采集少量样本即可建立准确的映射关系,并将免标定方法和单点标定方法应用到隐式标定原型系统中,简化了后续的标定过程,具有较强的实用性。

结束语 本文提出一种隐式标定方法,用户视线跟随沿着动态轨迹运动的图标,使用眼动跟踪系统收集用户眼部图像特征和标定点之间的映射点对。对于映射点对中包含的噪声数据,通过研究瞳孔变化进行降噪,并使用改进的RANSAC算法进一步排除噪声数据,最后将免标定方法和单点标定方法应用到隐式标定原型系统中,有效简化了隐式标定过程,减少了标定时间,本文还通过实验探索了动态轨迹的最优参数设置,实现了准确的标定。与其他方法进行的比较实验,验证了本文方法不仅简化了眼动数据的收集,还能减少标定时间,增强了眼动跟踪系统的实用性。本文的隐式标定技术在精度、实时性等方面能满足实际交互应用的基本需求,但仍存在改进空间。例如,通过一段视频、动画来采集样本数据,能进一步提高用户体验;此外,在去除样本噪声数据时,可以选择性能最佳的噪声数据过滤算法。

参考文献

- [1] FUHL W, SANTINI T, GEISLER D, et al. Eyes wide open? eyelid location and eye aperture estimation for pervasive eye tracking in real-world scenarios[C]// Acm International Joint Conference on Pervasive & Ubiquitous Computing: Adjunct. ACM, 2016: 1656-1665.
- [2] MORIMOTO C H, MARCIO R M. Eye gaze tracking techniques for interactive applications[J]. Computer Vision and Image Understanding, 2005, 98(1): 4-24.
- [3] SUGANO Y, MATSUSHITA Y, SATO Y. Calibration-free gaze sensing using saliency maps[C]// Computer Vision and Pattern Recognition. IEEE, 2010: 2667-2674.
- [4] WOOD E, TADAS B, MORENCY L P, et al. A 3D Morphable Eye Region Model for Gaze Estimation[C]// European Conference on Computer Vision. Springer International Publishing, 2016: 297-313.
- [5] KHAMIS M, SALTUK O, HANG A, et al. TextPursuits: using text for pursuits-based interaction and calibration on public displays[C]// Acm International Joint Conference on Pervasive & Ubiquitous Computing. ACM, 2016: 274-285.
- [6] HUANG C W, TSENG S C, JIANG Z S, et al. Projective mapping compensation for the head movement during eye tracking[C]// IEEE International Conference on Consumer Electronics-taiwan. IEEE, 2014: 2243-2246.
- [7] SUGANO Y, ZHANG X, BULLING A. AggreGaze: Collective Estimation of Audience Attention on Public Displays[C]// Symposium on User Interface Software and Technology. ACM, 2016: 821-831.
- [8] STEIL J, HUANG M X, BULLING A. Fixation Detection for Head-Mounted Eye Tracking Based on Visual Similarity of Gaze Targets[C]// Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications. New York: ACM, 2018: 23.
- [9] PAIVI M, ANDREAS B. Eye Tracking and Eye-Based Human-Computer Interaction[M]// Advances in physiological computing. London: Springer, 2014: 39-65.
- [10] DONGHENG L, DAVID W, DERRICK J P. Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches[C]// Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2005: 2667-2674.
- [11] PFEUFFER K, VIDAL M, TURNER J, et al. Pursuit calibration: Making gaze calibration less tedious and more flexible[C]// Proceedings of the 26th annual ACM symposium on User interface software and technology. ACM, 2013: 261-270.
- [12] SUGANO Y, BULLING A. Self-Calibrating Head-Mounted Eye Trackers Using Ego-centric Visual Saliency[C]// The 28th Annual ACM Symposium. ACM, 2015: 363-372.
- [13] HUANG M X, KWOK T C K, NGAI G, et al. Building a Personalized, Auto-Calibrating Eye Tracker from User Interactions[C]// Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, 2016: 5169-5179.
- [14] BERNET S, CUDEL C, LEFLOCH D, et al. Autocalibration-

- based partitioning relationship and parallax relation for head-mounted eye trackers[J]. *Machine Vision and Applications*, 2013, 24(2):393-406.
- [15] SANTINI T, FUHL W, KASNECI E. CalibMe: Fast and Unsupervised Eye Tracker Calibration for Gaze-Based Pervasive Human-Computer Interaction[C]// *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017:2594-2605.
- [16] SUN F R, LIU J R. Fast Hough Transform Algorithm[J]. *Chinese Journal of Computers*, 2001, 24(10):1102-1109. (in Chinese)
孙丰荣, 刘积仁. 快速霍夫变换算法[J]. *计算机学报*, 2001, 24(10):1102-1109.
- [17] YANG H. A study of Calibration Techniques on Eye Gaze Tracking System with Head Fixed[D]. Xi'an: Xidian University, 2015. (in Chinese)
杨慧. 头部静止的视线跟踪系统中标定技术的研究[D]. 西安: 西安电子科技大学, 2015.
- [18] KOBAYASHI T, TOYAMA T, SHAFAIT F, et al. Recognizing Words in Scenes with a Head-Mounted Eye-Tracker[C]// *Iapr International Workshop on Document Analysis Systems*. IEEE Computer Society, 2012:2142-2149.
- [19] LIU K, GE J F, LUO Y P. Probability guided random sample consensus[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2009, 21(5):657-662. (in Chinese)
刘坤, 葛俊锋, 罗予频. 概率引导的随机采样一致性算法[J]. *计算机辅助设计与图形学学报*, 2009, 21(5):657-662.
- [20] LAWRENCE H Y, EIZENMAN M. A new methodology for determining point-of-gaze in head-mounted eye tracking systems [J]. *IEEE Trans. Biomed. Engineering*, 2004, 51(10):1765-1773.
- [21] CERROLAZA J J, VILLANUEVA A, CABEZA R. Study of Polynomial Mapping Functions in Video-Oculography Eye Trackers[J]. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2012, 19(2):1-25.
- [22] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine. Theory and applications [J]. *Neurocomputing*, 2006, 70:489-501.
- [23] JIANG G Y. Research on data fusion of head movement and gaze tracking[D]. Xi'an: Xi'an Technological University, 2017. (in Chinese)
蒋光毅. 头部运动与视线追踪数据融合技术的研究[D]. 西安: 西安工业大学, 2017.
- [24] ESSIG K, POMPLUN M, RITTER H. A neural network for 3D gaze recording with binocular eye trackers [J]. *International Journal of Parallel, Emergent and Distributed Systems*, 2006, 21(2):79-95.
- [25] HOU S W. The Research of the Calibration Method in the head-mounted Eye Tracking Systems[D]. Hefei: University of Science and Technology of China, 2013. (in Chinese)
侯树卫. 头戴式视线跟踪系统的标定方法研究[D]. 合肥: 中国科学技术大学, 2013.
- [26] MORIMOTO C H, MIMICA M R M. Eye gaze tracking techniques for interactive applications[J]. *Computer Vision & Image Understanding*, 2005, 98(1):4-24.
- [27] CHENG S W, SUN Z Q, LU Y H. An eye tracking approach to cross-device interaction [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2016, 28(7):1094-1104. (in Chinese)
程时伟, 孙志强, 陆煜华. 面向多设备交互的眼动跟踪方法[J]. *计算机辅助设计与图形学学报*, 2016, 28(7):1094-1104.
- [28] POLA J, WYATT H J. Target position and velocity: The stimuli for smooth pursuit eye movements[J]. *Vision Research*, 1980, 20(6):523-534.
- [29] ALNAJAR F, GEVERS T, VALENTI R, et al. Auto-Calibrated Gaze Estimation Using Human Gaze Patterns[J]. *International Journal of Computer Vision*, 2017, 124(2):223-236.