

视线跟踪技术在人机交互中的研究

廖卫华¹，严由伟²，徐子清¹

(1. 赣南师范学院教科系，江西赣州 341000; 2. 福建师范大学教育科学与技术学院，福建福州 350007)

[摘 要] 人们在观察外部世界时眼睛总是与其它人体活动自然协调地工作，并且眼动所需的认知负荷很低，人眼的注视包含着当前的任务状况以及人的内部状态等信息，因此眼注视是一种非常好的能使人机对话变得简便、自然的候选输入通道。本文介绍了视线跟踪的基本原理、方法，对视线跟踪作为人机交互通道所具有的特点进行了讨论，并指出了其在水机交互领域中的应用及前景。

[关键词] 人机交互；视线跟踪；多通道用户界面

[中图分类号] TP311 [文献标识码] A [文章编号] 1008—178X(2005)02—0130-03

目前用户界面所使用的任何人机交互技术几乎都有视觉参与。早期的视线跟踪技术首先应用于心理学研究（如阅读研究），后被用于人机交互。眼动在人的视觉信息加工过程中，起着重要的作用。它有三种主要形式：跳动（Saccades）、注视（Fixations）和平滑尾随跟踪（Smooth Pursuit）。在正常的视觉观察过程中，眼动表现为在被观察目标上一系列的停留及在这停留点之间的飞速跳动，这些停留一般至少持续 100ms 以上，称为注视。绝大多数信息只有在注视时才能获得并进行加工。注视点间的飞速跳跃称为眼跳动。眼跳动是一种联合眼动（即双眼同时移动），其大小为 1 到 40 度视角，持续时间为 30 到 120ms，最高运动速度为 400～600 度/秒。

视线跟踪精度与在测量时对用户的限制和干扰就是一对矛盾。在多通道用户界面设计中，减少这种限制和干扰是非常重要的，视线作为交互装置最直接的用处就是代替鼠标器作为一种指点装置。

1. 视线跟踪的基本原理、方法

视线跟踪技术及装置有强迫式（intrusiveness）与非强迫式（non-intrusiveness）、穿戴式与非穿戴式、接触式（如 Eyeglass-mounted）与非接触式（Remote）之分；其精度从 0.1°至 1°或 2°不等，制造成本也有巨大差异。在价格、精度与方便性等因素之间作出权衡是一件困难的事情，例如视线跟踪精度与对用户的限制和干扰就是一对尖锐的矛盾。有关视觉输入的人机界面研究主要涉及两个方面：一是视线跟踪原理和技术的研究；二是在使用这种交互方式后，人机界面的设计技术和原理的研究。

眼睛能平滑地追踪运动速度为 1 到 30 度/秒的目标，这种缓慢、联合追踪眼动通常称为平滑尾随跟踪。平滑尾随跟踪必须有一个缓慢移动的目标，在没有目标的情况下，一般不能执行这种眼动。在水机交互中，主要表现为跳动和注视两种形式。而主要的视线跟踪技术方法如表 1 所示。

视线追踪的基本工作原理是利用图象处理技术，使用能锁定眼睛的特殊摄像机。通过摄入从人的眼角膜和瞳孔反射的红外线连续地记录视线变化，从而达到记录分析视线追踪过程的目的。在水机交互中对视线追踪的基本要求是：（1）要保证一定的精度，满足使用要求；（2）对用户基本无干扰；（3）定位校正简单；（4）可作为计算机的标准外设。

2. 视线跟踪技术在水机交互通道中的特点

人机交互通道中大部分具有一定形状的物体或现象，可以通过多种途径使用户产生真实感很强的视觉感知。CRT 显示器、大屏幕投影、多方位电子墙、立体眼镜、头盔显示器（HMD）等是 VR 系统中常见的显示设备。不同的头盔显示器具有不同的显示技术，根据光学图像被提供的方式，头盔显示设备可分为投影式和直视式。

能增强视线跟踪真实感的立体显示技术，可以使用户的左、右眼看到有视差的两幅平面图像，并在大脑中将它们合成并产生立体视觉感知。头盔显示器、立体眼镜是两种常见的立体显示设备。目前，基于激光全息计算的立体显示技术、用激光束直接在视网膜上成像的显示技术正在研究之中。

[收稿日期] 2005—02—29

[基金项目] 全国教育科学“十五”规划重点课题“我国高校心理健康教育方式网络现代化的系统软件开发与研究”阶段成果之一（DCA030090）。

[作者简介] 廖卫华（1977—），男，江西赣州人，赣南师范学院教科系讲师，从事现代教育技术研究。

表 1 六种主要的视线跟踪技术

视觉追踪法	应用场合	技术特点	测量参照系
眼电图（EOG）	眼动力学	高宽带 精度低 对人干扰大	头
虹膜—巩膜边缘	眼动力学注视点	高带宽 垂直精度低 对人干扰大 头具误差大	头具
角膜反射	眼动力学注视点	高带宽 头具误差大	头具
瞳孔—角膜反射向量	注视点	准确 头具误差小 对人无干扰 低宽带	头具或室内
双浦肯野象	眼动力学 网模图象稳定 注视点	高精度 高带宽 对人干扰大	室内
接触镜	眼动力学 微小的眼动	精度最高 高带宽 对人干扰大 不舒适	头

从视线跟踪装置得到的原始数据必须经过进一步的处理才能用于人机交互。数据处理的目的是从中滤除噪声（filter noise），识别定位（recognize fixations）及局部校准与补偿（compensate for calibration errors）等，最重要的是提取出用于人机交互所必需的眼睛定位坐标。但是由于眼动存在固有的抖动（jitter motion）以及眼睛眨动所造成的数据中断，即使在定位这段数据段内，仍然存在许多干扰信号，这导致提取有意（intentional）眼动数据的困难，解决此问题的办法之一是利用眼动的某种先验模型加以弥补。将视线应用于人机交互必须克服的另一个固有的困难是避免所谓的“米达斯接触（Midas Touch）”问题。如果鼠标器光标总是随着用户的视线移动，可能会引起他的厌烦，因为用户可能希望能随便看着什么而不必非“意味着”什么。在理想情况下，应当在用户希望发出控制时，界面及时地处理其视线输入，而在相反的情况下则忽略其视线的移动。然而，这两种情况一般不可能区分。

目前，美国 Texas A&M 大学使用装有红外发光二极管和光电管的眼镜，根据进入光电管的光的强弱来决定眼睛的位置。ASL（Applied Science Lab）也已有较成熟的视线追踪系统。

3 视线跟踪技术在人机交互领域中的应用及前景

随着计算机科学技术与产业的飞速发展，计算机日益进入普通用户的工作和生活中。面对急剧增长的普通用户，人机交互的自然性愈来愈为系统设计所重视，计算机系统的设计必须力图使用户不需要专门的计算机技能，而只是利用自己的自然技能就能够使用计算机，虚拟现实正致力于这方面的研究。另一方面，多媒体技术和网络技术的迅猛发展使计算机到用户的通讯带宽不断提高，而目前的人机交互由于受传统专业化的窄带宽输入界面的限制，使得用户到计算机的通讯带宽仍停滞不前，严重阻碍了人向计算机的信息交流。此外，传统人机交互的串行性和精确性在许多场合不必要地增加了用户的作业负荷，降低了交互效率，破坏了自然性。为了赋予人机交互灵活性、健壮性、高带宽性和类似于人与人交互中的自然性，人机交互必须打破常规，走多通道的道路。

现在多通道界面的研究采用视线跟踪、语音识别、唇读和手势输入等新的交互技术，允许用户通过多个通道自然的并行和协作来进行人机交互，并可通过整合互补通道的模糊输入来捕捉用户交互意向。多通道界面的特点和目标给设计带来了一些特殊问题，主要存在于三方面：

（1）多通道整合。互相独立地利用多个通道并不是真正意义上的多通道界面，并不能有效地提高人机交互的效率。多通道交互需要从多个并行、协作和互补的通道的非精确输入获得用户想传达的任务信息，这就是整合问题——实现多通道的首要前提。

（2）软件结构。多通道协同操作使得软件结构必须具有支持强语义处理的能力，传统 UIMS 所采用的 Seeheim 模型，在这方面有较大的局限性。

（3）用户任务分析模型和界面描述方法。这样的模型和方法必须适合多通道界面的特点。从用户角度而言，要能够表达用户利用多通道协作来完成交互任务时的特点，应能很好地分析和描述交互过程中人的活动，体现以人为中心

的设计宗旨;从系统角度而言,要能够处理多种交互手段的并行协作使用与系统处理的关系。

视线跟踪技术还处于起步阶段。视线跟踪技术主要是解决眼睛运动特性的检测问题,目前主要的检测方法有接触镜法、电磁线圈法、红外光电反射法、红外电视法等。其中红外电视法具有操作方便,对人无干扰,可移动,非接触等优点。基于红外电视法的眼睛盯视人机交互技术是通过眼睛盯视激活对话框,从而实现对外部设备的控制,阐述了眼睛盯视技术的原理和视线方向的判定方法,提出了对于使用过程中头部微小变化时视线方向的修正方法,降低了对使用者的使用要求。眼睛盯视技术的应用领域对正常人来说,通过对鼠标和键盘操作,就能实现与计算机间的交互,但是对某些瘫痪病人或四肢麻痹,又不能说话的人来说,如此简单的任务却无法完成。有关资料统计显示:全国至少有50万的人口存在不同程度的肢体瘫痪,生活不能自理。那么如果他们能用眼睛来代替手操作,以后再加上机电控制技术情况就不一样了,就完全可以增加他们的独立能力,提高生活质量。另外,通过眼睛盯视对外部设备进行控制可以实现多任务操作,比如在军事上,飞行员如果发现了目标,在手动操作应付不过来的时候,可以通过眼睛瞄准的同时,用眼睛来控制火控系统的发射,这样可以使飞行员既为驾驶员又为武器操纵员。同时,在飞行加速度环境下,飞行员的头部和手部活动受到极大的限制,但眼睛却还可以自由转动,这对于增加战斗力非常有益。随着研究工作的深入,这项技术必将服务于医学、军事及教育等各个方面。

总之,尽管视线跟踪技术的应用存在着诱人的前景,但由于各方面技术的不成熟性,目前还未达到实用化阶段。成功的、有效益的演示性项目还很少。在多通道用户界面基础上,要进行进一步的原型探索分析,必须对人机交互设备的性能加以改善,同时注意有关标准的建立,以便把视线跟踪技术更好地与现有技术结合起来。

[参 考 文 献]

[1] 冯成志, 沈模卫. 视线跟踪技术及其在人机交互中的应用 [J]. 浙江大学学报 (理科版), 2002, 3: 1.
[2] 李茂贞, 戴国忠, 董士海. 多通道界面模型与关键技术 [J]. 计算机科学, 1997, 24 (1): 1—4. LI Mao-zhen, DAI Guo-zhong, DONG Shi-hai. Multimodal interface model and key technology [J]. Computer Science, 1997, 24 (1): 1—4.
[3] STIEFELHAGEN R, FINKE M, YANG J, et al. From gaze to focus of attention [J]. Lecture Notes in Computer Science, 1999, 1614: 761—768.
[4] Green, M. (1985). The University of Alberta user interface management system [J]. Proc. SIGGRAPH' 85, 205—213.
[5] STIEFELHAGEN R, YANG J, WAIBEL A. A model-based gaze tracking system [J]. International Journal on Artificial Intelligence Tools, 1997, 6 (2): 193—209.
[6] STIEFELHAGEN R, YANG J. Gaze tracking for multimodal human-computer interaction [A]. In Proceedings of International Conf. on Acoustics, Speech, and Signal Processing [C]. Germany: Munich, 1997.

Research on Visual Trace Technology in Human-Computer Interaction

Liao Wei-Hua, Yan You-Wei, Xu Zi-Qing
(Gannan Teacher's College, Garzhou 341000, China)

Abstract: When people observe the outside world, their eyes always coordinate naturally with the other physical movements with very low cognitive load. The gaze of human eyes contains much information such as the condition of the present task and the inner state of humans, so the gaze of human eyes is a superior alternative input channel which enables the dialogue between humans and computers to be simpler and more natural. The essay introduces the basic principles and methods of the visual trace, dicusses the characteristics of the visual trace as the channel of Human-Computer Interaction, and points out the application as well as the prospect of the technology in the field of Human-Computer Interaction.

Key words: Human-Computer Interaction Visual trace Multi-channel interface of users