

眼动交互的实时线性算法构造和实现

程 成, 杜菁菁, 蓝飞翔

(北京理工大学计算机科学技术学院, 智能信息技术北京市重点实验室, 北京 100081)

摘 要: 眼动跟踪技术作为一种新兴的交互手段, 具有自然、高效和易于意图理解等优点. 我们设计实现了一个桌面眼动跟踪系统作为虚拟装配系统的一个通道, 通过摄像机实时采集视频并经过实时的图像处理判断用户对虚拟环境中对象的注视焦点. 系统利用了瞳孔-角膜反射向量法理论, 设计和实现了一个高效的实时眼动跟踪算法, 即线性逼近预测算法. 实验证明这一方法很好的解决了眼动通道的实时性问题.

关键词: 眼动跟踪; 人机交互; 线性逼近预测算法; 瞳孔-角膜反射向量法

中图分类号: TP334.2+2 **文献标识码:** A **文章编号:** 0372-2112(2009)4A-012-04

Research and Realization on Real-time Linear Prediction Algorithm for Desktop Eye-Tracking

CHENG Cheng, DU Jing-jing, LAN Fei-xiang

(Beijing Key Lab of Intelligent Information, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China)

Abstract: Eye-tracking as a new method of human computer interaction has the merits of naturalness, efficiency and easy understanding. The authors have designed a Desktop Eye-Tracking System which used the technique of image processing to obtain the user's gaze focus in virtual environment scene. Based on the theory of pupil center corneal reflection, the authors improved and realized a highly efficient linear prediction algorithm for real-time eye gaze image processing. Experiments proved that this algorithm can well support real-time eye-tracking as an interaction modality for desktop virtual environment.

Key words: eye-tracking; human computer interaction; linear prediction algorithm; pupil center corneal reflection

1 引言

Dalhousie 大学的 Ritchie Argue 等人在 2004 年利用基于广泛使用的瞳孔-角膜反射向量法, 并采用普通设备设计了一个眼动跟踪系统, 只是仅计算了注视点位置, 没有讨论实时眼动问题^[1]. 现今的眼动跟踪技术正在将图像处理技术和各种实时技术相结合来解决实时的难题. Nevada 大学的 Zhiwei Zhu 在进行实时图像处理时采用椭圆拟合算法然后再利用几何特征进行排除并用支持向量机来进行瞳孔确认^[2], 这一复杂过程使得准确度得到保证但实时性大减; 德国的 Frank Wallhoff 等人利用人工神经网络建立了瞳孔识别模型用于实时图像处理^[3], 但他们的复杂和不稳定的实时眼部检测算法对系统的鲁棒性造成了很大的影响, 带来了大量数据处理和噪声干扰等问题. 同时, NTT 公司实验室实现了一个眼动系统, 由视线跟踪和眼睛定位系统组成, 由两台计算机分别进行计算^[4]. 这在很大程度上缓解了实时处理操作. 视线跟踪系统如果能够采用实时预测的优化就会

更加高效. 常用的预测运动目标位置的方法有线性预测法^[5]、统计预测法^[6]和卡尔曼滤波器预测法^[7]. 西北工业大学的郑江滨等人提出了利用扩展的 Kalman 滤波器和目标加速度位移方程^[8]来预测下一时刻目标位置, 收到了良好效果. 但对于眼动这一搜索范围小、实时性与精度要求高的特殊系统, 这些方法并不适合, 我们提出了一种改进的实时线性预测算法.

2 基本原理和系统设计

本系统利用瞳孔-角膜反射向量法来实现瞳孔位置到屏幕注视点的映射. 首先, 通过提取出眼球图像内的瞳孔及角膜对红外光的反光点位置. 接着, 根据观察者头部位置对映射函数进行标定. 最后, 把瞳孔-反光点向量代入映射函数求出注视点位置. 系统设计模块主要包括照明系统和摄像系统. 照明系统选择 850 纳米的红外光源照射人的眼睛, 来达到检测瞳孔的目的. 摄像系统我们选用松下 NV-GS158GK 作为实验用摄像机. 整个光电路部分由红外滤光片、红外光源、监视器、摄像机和一

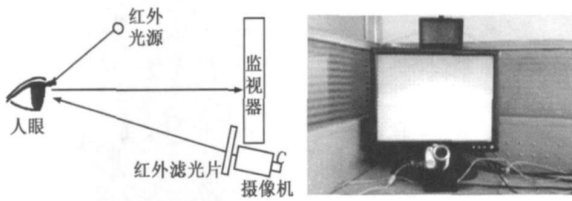


图1 摄像系统光路图



图2 系统硬件设备图

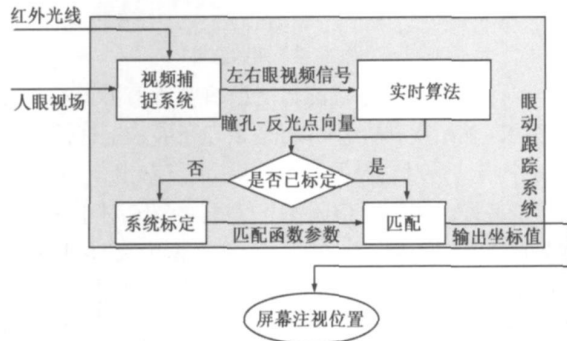


图3 系统原理图

些调节机构组成。眼动捕捉中的光路如图 1 所示; 图 2 是由上述硬件设备组成的眼动跟踪原型系统。

系统实现原理如图 3 所示。视频捕捉负责从摄像机中读取视频信息并按系统设定间隔抓取一帧图像, 实时算法从实时图像中计算出瞳孔-反光点向量, 通过系统标定针对头部位置进行映射函数参数的标定, 通过匹配在屏幕中反映出观察者注视点位置。

3 眼动实时图像处理算法

本文提出了一种基于线性预测的快速眼动检测与跟踪的实时算法。它利用线性预测的目标运动估计算法, 根据前若干帧中所获取的目标点位置来预测当前帧中的目标点位置, 这样可以充分利用前面图像处理的结果, 减少处理数据量, 缩短处理的时间, 且能够在一定程度上减少场景中其它物体对跟踪的影响。

在眼动跟踪系统中, 我们每从视频中抓取一张视频图像就会调用一次实时算法, 因此, 跟踪的结果准确性与实时性是由这一关键算法决定的。在本设计中, 采用最佳线性逼近预测算法来估计下一帧图像中瞳孔出现的位置, 并以这个位置为中心确定一个搜索范围, 在这个范围内采用瞳孔-反光点的检测算法来检测瞳孔中心位置。

设原函数 $f(t)$ 在 N 个顺序时刻的测量值为 $f(t_i)$ ($i = 1, 2, 3, \dots, N$), 且 $f(t)$ 可以用:

$$Y = a_0 + a_1 t \quad (1)$$

作为最佳线性逼近, 上式可写为:

$$Y = \begin{bmatrix} 1 & t \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \quad (2)$$

测量值与逼近值之间的误差为:

$$\Delta \varepsilon_i = f(t_i) - a_0 - a_1 t_i \quad (3)$$

对 N 点估计的误差平方和为:

$$\sum_{i=1}^N \Delta \varepsilon_i^2 = \sum_{i=1}^N [f(t_i) - a_0 - a_1 t_i]^2 \quad (4)$$

最佳逼近即是使上式取最小值。经最小二乘法计算后可得:

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \frac{\sum_{i=1}^N t_i^2 \sum_{i=1}^N f(t_i) t_i - \sum_{i=1}^N t_i \sum_{i=1}^N f(t_i) t_i}{D} \\ - \frac{\sum_{i=1}^N t_i \sum_{i=1}^N f(t_i) t_i - N \sum_{i=1}^N f(t_i) t_i}{D} \end{bmatrix} \quad (5)$$

$$\text{其中: } D = N \sum_{i=1}^N t_i^2 - \left(\sum_{i=1}^N t_i \right)^2$$

由以上线性逼近可以很方便地得到预测器表达式。如用三阶预测器, 即用过去三帧的位置去预测下一帧的位置问题。按上述表示方法有 $t_1 = 1, t_2 = 2, t_3 = 3$ 及 $f(t_1) = f(k-2), f(t_2) = f(k-1), f(t_3) = f(k)$ 。对于 $k+1$ 帧而言, 相当于 $t = 4$ 的情况, 故求取 $k+1$ 帧的位置为:

$$f(k+1) = a_0 + a_1 t_{k+1} = a_0 + 4a_1 \quad (6)$$

将上述 $t_1 \sim t_3$ 及 $f(t_1) \sim f(t_3)$ 的值代入式(5)可求得 a_0, a_1 的值, 代入式(6)可求得:

$$f(k+1) = \frac{1}{3} [4f(k) + f(k-1) - 2f(k-2)] \quad (7)$$

其中 $f(k), f(k-1), f(k-2)$ 为第 $k+1$ 帧, 即当前帧的前三帧瞳孔中心坐标实际测量值, $f(k+1)$ 为当前帧的坐标预测值。这样, 可以利用前三帧获取的瞳孔中心位置信息来估计下一帧, 即当前帧的瞳孔中心坐标。

系统中利用红外光的照射, 使得虹膜与瞳孔的边缘处像素的灰度值差异较明显, 然后利用基于高斯、金字塔分解的数字图像滤波算法对图像进行去噪处理, 最后通过检测瞳孔的轮廓, 使用椭圆拟合算法获取瞳孔中心。根据所获取的瞳孔中心, 在一个限定范围的区域内提取出值最大的像素块, 该像素块的平均坐标即作为角膜反光点的坐标。

图 4 所示为利用算法检测到的瞳孔中心和角膜反光点坐标, 图 5 显示的是线性逼近预测算法在检测到瞳孔后, 预测的反光点检测范围。

根据以上实时算法的两个过程, 我们改进和实现了一个完整的实时算法, 具体算法流程为(假设搜索范

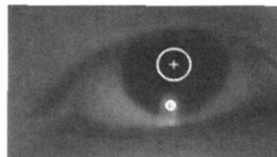


图4 检测到的瞳孔中心及角膜反光点

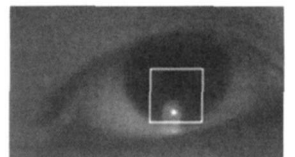


图5 根据瞳孔中心位置确定的反光点检测范围

围为: $\Delta P_{k+1} = K$, K 是常数):

(1) 从视频序列中抓取当前帧的一张瞳孔图像, 然后根据之前三帧图像中获取的瞳孔中心位置信息, 利用式(7), 估算出当前帧中瞳孔中心位置范围; 若之前不够三帧图像 就以设定好的参数值来作为已获取的瞳孔中心位置信息;

(2) 由公式得出当前帧估算的瞳孔位置信息, 即确定搜索范围 $f(k+1) + \Delta P_{k+1}$;

(3) 在这一范围内检测瞳孔, 成功则进入(4), 否则将这一范围扩大 ΔK 继续搜索, 成功则进入(4), 否则提示查找失败, 放弃当前帧的检测, 转(1);

(4) 获取当前帧中瞳孔中心位置 P_{k+1} ;





(5) 在搜索范围 $P_{k+1} + M$ (M 为搜寻反光点时确定的范围) 内提取反光点 G_{k+1} , 成功则完成这一帧图像的瞳孔、反光点检测, 转(1); 否则将范围扩大 ΔM 继续搜索, 成功转(1), 否则, 放弃当前帧图像开始下一帧图像的实时算法检测。

实验表明, 该算法减少了需要处理的图像区域面积, 计算量相比处理整幅人眼图像来说要少得多, 能够满足系统的实时性的要求, 且在一定程度上又能排除其他干扰物对眼动跟踪的影响, 从而保证匹配的可靠性。

4 眼动系统实验和结论

本设计使用的显示屏是 17 英寸液晶显示屏, 摄像机置于屏幕的正下方, 红外发光仪置于显示器的正上方, 这样正好可以照射人脸。同时, 本设计使用的计算机配置为 Intel 的奔腾 4, CPU 频率为 3.0GHz, 512MB 内存。摄像机的视频分辨率为 $720 * 576$ 像素, 帧率 15Hz。

由于映射函数的参数对于系统的精度具有很大的影响, 因此需要验证系统标定过程中标定点个数对于系统精度的影响。因此, 本实验分别设计了多种样式的标定点, 包括 4 个、5 个、9 个以及 13 个。其形状和位置如下表:

数目	13	9	5	4
位置和形状				

实验中在屏幕上按上表 1 标记 9 个点。用户眼睛依次从左上角第 1 个点开始盯视每个点 1~3 秒钟, 直至第 9 个点为止。然后, 根据这 9 个点的坐标值, 计算出映射函数的参数, 进而可以根据人眼瞳孔-角膜反光点向量计算出人眼注视点的位置。系统跟踪结果显示如图 7 所示。

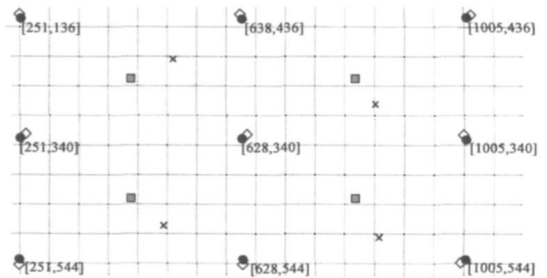


图 7 系统跟踪结果

图 7 中, 9 个圆点是系统的标定点, 9 个菱形标识系统对它们的匹配结果; 4 个小方格是预先在屏幕上标注了的点, 4 个小叉是对它们的匹配结果。可以看出, 对 9 个标定点的匹配结果较之中间 4 个点更为精确。对于这样的结果可以解释为: 当人的眼睛盯视屏幕中央时, 瞳孔中心的位置和角膜反光点的位置十分接近。事实上, 在用户注视中间这 4 个点时, 瞳孔中心和角膜反光点的位置相差不到 5 个像素, 因此, 即使是一个较小的计算误差也会导致相当大的匹配误差。

在不同视频频率条件下, 作者设计了两个不同的测试过程, 来评测眼动跟踪系统精度所受到的影响。让实验者在进行完系统标定过程后, 坐在原来的位置不动, 再对他进行眼动跟踪, 记录下该实验者对屏幕上以 $4 * 4$ 格式标注的点进行跟踪的结果数据。它们的平均精度如表 2 所示:

表 2 平均精度实验结果表

实验结果	平均精度(单位: 像素)	
	X	Y
计算机/视频格式		
奔腾 IV/15Hz	15.8	20.3
奔腾 IV/25Hz	26.5	29.7

图像处理过程中会出现两种情况: 一是利用实时图像处理算法成功获取人眼图像的感兴趣区域(ROI), 那么只需要对这一区域进行图像处理, 从而大大减少了图像处理的数据; 另一个是获取 ROI 失败的情况, 例如, 用户在眨眼的环境下, 则需要对整幅图片进行处理来获取瞳孔。实验结果如下表 3 所示:

表 3 成功或失败获取感兴趣区域情况下图像处理所用时间

结果	处理时间(单位: ms)	
	获取 ROI 成功	获取 ROI 失败
计算机/视频格式		
奔腾 IV/15Hz 720* 576	24	89
奔腾 IV/25Hz 320* 240	23	32

从表 3 可以看出, 对于成功获取 ROI 的一帧图片来说, 即线性预测算法成功时, 处理时间比不使用该预测算法少了很多。在分辨率高的情况下, 处理时间的增幅要比在分辨率低的情况下要大。该算法成功的发挥了

作用, 并且预测范围的确定还降低了噪声的干扰。

本系统采用瞳孔-角膜反射向量法, 获得眼睛的注视信息并显示在屏幕上。在实时图像的处理上采用线性逼近眼动预测算法, 一定程度上减少了不必要的图像处理内容, 并且避免了一些噪声的干扰, 实现了一种可行高效的眼动跟踪方法。目前系统在跟踪精度、头部运动等方面还需要解决若干实际问题。同时, 我们正在把眼动加入现有的虚拟装配原型系统中进行实验, 在后面的装配中实现通过眼动获取用户意图, 实现包括眼动和双手的多通道虚拟环境装配系统。

参考文献:

- [1] Ritchie Argue, Matthew Boardman, Jonathan Doyle, Glenn Hickey. Building a Low-Cost to Track Eye Movement [EB/OL]. <http://serv1.is.psu.edu:8080/viewdoc/summary?doi=10.1.1.92.2962>, 2004.
- [2] Zhiwei Zhu, Kikuo Fujimura, Qiang Ji. Real-time detection and tracking under various light conditions [J]. Computer Vision and Image Understanding, 2002, 98 (1) : 79– 85.
- [3] Frank Wallhoff, Markus Ablabmeier, Gerhard Rigoll. Multimodal face detection, head orientation and eye gaze tracking [A]. International Conference on Multisensor Fusion and Integration for Intelligent Systems [C]. Heidelberg, Germany: 2006. 13– 18.
- [4] Takehiko Ohno, Naoki Mukawa. A free-head, simple calibration, gaze tracking system that enables gaze-based interaction [A]. Proceedings of the 2004 Symposium on Eye Tracking Research & Applications [C]. San Antonio, Texas: ACM Press, 2004. 115– 122.
- [5] Manjula A Iyer, Morgan M Harris, et al. A performance comparison of piecewise linear estimation methods [A]. Proceedings of the 2008 Spring Simulation Multiconference [C], Ottawa, Canada: ACM Press, 2008. 273– 278.
- [6] Boyd J E, Meloche J, et al. Statistical tracking in video traffic surveillance [A]. The Proceedings of the 17th IEEE International Conference on Computer Vision [C]. Kerkyra, Greece: IEEE, 1999. 163– 168.
- [7] Haritaoglu I, Harwood D, et al. Active outdoor surveillance [A]. Proceedings International Conference on Image Analysis and Processing [C]. NW Washington, DC USA: IEEE Computer Society, 1999. 1096.
- [8] 郑江滨等. 视频监控中的运动目标跟踪算法 [J]. 系统工程与电子技术, 2007, 29 (11) : 1991– 1993.

作者简介:



程 成 男, 1966 年 11 月出生于辽宁营口。2003 年博士毕业于中科院软件所, 现为北京理工大学计算机学院副教授。从事人机交互技术研究。

E-mail: guoqiangcheng@vip.sina.com



杜菁菁 男, 1984 年 2 月出生于河南焦作。2007 年硕士就读于北京理工大学计算机学院, 现为北京理工大学计算机学院硕士研究生。从事图像处理技术研究。

E-mail: tadalon@126.com



蓝飞翔 男, 1984 年 9 月出生于江西高安。北京理工大学计算机学院硕士研究生。从事图像处理技术研究。

E-mail: flan99@gmail.com