

基于极限学习机(ELM)的视线落点估计方法

朱 博, 张天侠

(东北大学 机械工程与自动化学院, 辽宁 沈阳 110819)

摘 要: 基于极限学习机(ELM)所具有的训练速度快、适合多分类的特点,提出一种新的单摄像机视线追踪系统视线落点估计方法。在初始标定阶段,将多视线参数作为ELM输入,将视线在屏幕上的落点区域作为输出,将非线性多项式作为激活函数,通过初始标定获取ELM训练数据,建立视线特征参数和视线屏幕落点之间的映射模型。实验结果表明,通过对不同角度分布的视线落点进行估计和改变隐层单元数量进行训练,基于ELM的视线落点估计方法无论视线落点精度还是稳定性均优于传统的非线性多项式拟合方法。

关 键 词: 视线追踪; 极限学习机; 视线落点估计; 映射模型; 多项式模型

中图分类号: TP 391.41

文献标志码: A

文章编号: 1005-3026(2013)03-0335-04

Gaze Point Estimation Method Based on ELM in Gaze Tracking System

ZHU Bo, ZHANG Tian-xia

(School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110819, China.
Corresponding author: ZHANG Tian-xia, E-mail: txzhang@me.neu.edu.cn)

Abstract: A novel method of gaze point estimation used in single camera system, which has faster training speed and is suitable for multiple classifications, was proposed based on ELM. In the initial calibration phase, multi gaze parameters served as ELM input, while the gaze point area was ELM output. The nonlinear polynomial was used as activation function. ELM training data was obtained through the initial calibration, and then the mapping model between the line of sight parameters and the gaze point was established. Through estimation of gaze point distribution of different angles and changing the number of hidden neurons, it was found that the accuracy and stability of gaze point obtained by ELM method are better than those obtained by traditional nonlinear polynomial model.

Key words: gaze tracking; ELM(extreme learning machine); gaze point estimation; mapping model; polynomial model

视线追踪技术是利用机械、光学、电子等各种检测手段获取受试者当前“视觉注意”方向的技术^[1-3],目前基于单摄像机的视线追踪系统视线落点估计方法都是采用非线性多项式作为视线参数到视线落点的映射模型^[4-6]。这种方法的精度和稳定性不高。本文利用极限学习机(extreme learning machine, ELM)所具有的适合多分类、速度快和泛化能力强的特点,提出了一种基于ELM的视线追踪视线落点估计方法,将提取的视线特征参数作为输入,视线在屏幕上的落点作为输出,将非线性多

项式作为激活函数,通过初始标定获取ELM训练数据,建立视线特征参数和视线屏幕落点之间的映射模型。实验结果表明,基于ELM的视线落点估计方法在标定点处的视线落点精度和稳定性均优于传统的非线性多项式拟合,适合于需要稳定判断视线落点的应用场合,为单摄像机视线追踪系统从视线参数到视线落点的映射提供了一种新思路。

本文的视线追踪系统是单摄像机无光源系统,由普通摄像机、主机和显示屏组成。在人因分析、辅助驾驶等领域有着广泛的应用。本文的视线

收稿日期: 2012-09-11

基金项目: 交通运输部公路科学研究所运输车辆运行安全技术实验室开放研究课题。

作者简介: 朱 博(1970-),男,辽宁阜新,东北大学博士研究生; 张天侠(1945-),男,辽宁营口人,东北大学教授,博士生导师。

追踪系统算法首先检测平面视线参数,然后通过视线参数到视线屏幕落点之间的映射模型,估计视线的屏幕落点坐标.因此视线追踪算法分为 4 个部分:人脸和人眼定位、人眼特征参数检测、视线参数到视线屏幕落点的映射模型、头动情况下的视线落点补偿.本文工作就是提出了一种基于 ELM 的新的视线落点映射模型.

1 极限学习机(ELM)

极限学习机^[4]是单隐层前馈神经网络(SLFN)中一种很有效的学习算法.ELM 的输入权值,隐含层的阈值都是随机选取,已知隐含层的输出矩阵,它的输出权值利用广义逆矩阵而得.

SLFN 有 N 个隐层结点, n 个输入, m 个输出.假设有 M 个任意且独立的样本 (x_i, t_i) , 其中 $x_i \in R_n, t_i \in R_m$, 含有 N 个隐层神经元及激活函数为 $g(x)$ 的标准 SLFN 数学建模如下:

$$\sum_{i=1}^N \beta_i g(x_j) = \sum_{i=1}^N \beta_i g(w_i \cdot x_j + b_i), \quad j = 1, \dots, M. \quad (1)$$

其中: $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ 是连接第 i 层输入结点与隐层结点的权值向量; $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ 是连接第 i 个隐层结点与输出结点的权值向量; b_i 是第 i 层隐层结点的阈值; (w_i, x_i) 表示 w_i 和 x_i 的内积,输出结点为线性结点.含有 N 个隐层神经元及激活函数为 $g(x)$ 的标准 SLFN 可以零误差的逼近这 M 个样本,它的意思也就是存在 β_i, w_i 和 b_i , 使得

$$\sum_{i=1}^N \beta_i g(x_j) = \sum_{i=1}^N \beta_i g(w_i \cdot x_j + b_i) = t_j, \quad j = 1, \dots, M. \quad (2)$$

以上 N 个等式可以写成:

$$H\beta = T. \quad (3)$$

其中: $H(w_1, \dots, w_N, b_1, \dots, b_N, x_1, \dots, x_M) =$

$$\begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_N \cdot x_1 + b_N) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_M + b_1) & \dots & g(w_N \cdot x_M + b_N) \end{bmatrix}_{M \times N};$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_N^T \end{bmatrix}_{N \times m}; T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}.$$

根据文献[5], H 是隐层的输出矩阵, H 的第 i 列是输入为 $x_1, x_2, \dots, x_M$ 的第 i 层隐层结点的输出.如果激活函数是无限可微的,可以证明所需要的 $N < M$. 随机选取连接权值和阈值,输出权值的解为: $\beta = H^+ T$. H^+ 是 H 的 Moore-Penrose 广义逆.

2 基于极限学习机的视线屏幕落点估计方法

目前实现视线估计的方法多数是采用多项式模型估计出视线在屏幕上的落点位置,即通过标定(注视屏幕上 9 个不同的位置)可以求出二次多项式模型,再根据求出的多项式模型可以估计出不同的视线落点.该方法准确性和稳定性不高.本文基于 ELM 适合于多分类、训练速度快和训练样本本大的特点,提出一种基于 ELM 的视线落点估计方法.在初始标定过程中,用户注视屏幕上不同位置时,内眼角在人脸的绝对位置是不变的,而虹膜中心到内眼角之间的相对位置发生了有规律的变化,如图 1 所示.因此可以通过 ELM 训练大量虹膜与内眼角之间的参数关系来判断视线落点的位置.

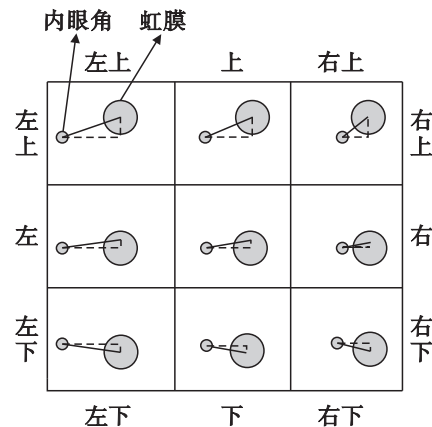


图 1 内眼角与虹膜相对位置
Fig. 1 The relative position of the eye inner corner and iris center

采用 ELM 方法实现视点估计的过程: 首先采集分布在图像平面 9 个不同位置的样本,如图 2 所示,即提取出用户注视屏幕上已知固定点的特征参数,其中水平边缘为 80 像素值,垂直边缘为 100 像素值,水平边距 = (屏幕宽 - 水平边缘 $\times$ 2) / 2,垂直边距 = (屏幕高 - 垂直边缘 $\times$ 2) / 2; 然后对采集到的样本参数进行 ELM 训练,训练出 9 种不同的类型,对应屏幕上的不同位置,判断出用户注视屏幕上视线落点位置.

为了提高 ELM 分类的准确性,本文中采用两种方法: 1) 采用多输入方法,即增加 ELM 网络输入层的输入个数,本文中 ELM 的输入参数包括虹膜中心到内眼角向量、虹膜的椭圆长短轴、椭圆度、虹膜中心到内眼角向量与水平面之间的夹角; 每一种分类代表一种视线落点位置,然后验证分类的准确性; 2) 增加隐层神经元的个数,但是

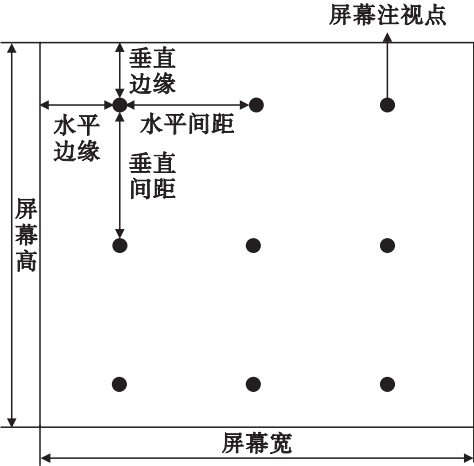


图 2 屏幕注视点分布图
Fig. 2 The distribution map of the gaze point on screen

随着隐层神经元个数的增多,ELM 的分类准确率逐渐增大,同时也增长了训练时间. 因此为同时满足准确率和训练时间的要求,尽量选择两者折衷情况.

图 3 为本文提出的采用多输入的 ELM 网络结构图,其中 ELM 的输入参数为 6 个,隐层神经元的个数 $N = 200$. 其中 $x, y, a_{\text{major}}, a_{\text{minor}}, a_{\text{major}}/a_{\text{minor}}, \theta$ 分别表示虹膜中心到内眼角向量坐标,虹膜的椭圆长短轴,椭圆度,虹膜中心到内眼角向量与水平面之间的夹角, w_{ij} 表示第 i 输入对应到第 j 个隐层神经元的权值, N 为 200, b_j 表示第 j 个隐层神经元的阈值, $g(x)$ 是激活函数, β_{jk} 表示第 j 个神经元到第 k 个输出的权值($i = 1, 2, \dots, 6, j = 1, 2, \dots, 200$).

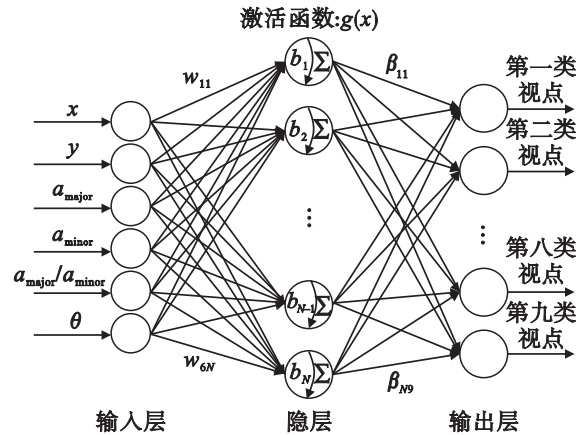


图 3 多输入 ELM 网络结构图
Fig. 3 Diagram of the ELM network structure with multiple input

3 实验结果及分析

3.1 采用多项式映射模型和 ELM 视点估计的对比实验结果

在头部静止和头部轻微移动情况下,本文实

验分别采用多项式映射模型方法和 ELM 方法进行视线落点估计,分别记录下注视屏幕上已知位置的 9 个点的视线落点,实验中对于每一个注视点分别采集 200 个样本,屏幕已知注视点的分布如图 2 所示. 在头部静止情况下采用多项式模型估计视线落点实验结果如图 4 所示. 针对同样的屏幕注视点,采用 ELM 方法进行视线落点估计. 隐层神经元数量取 200,采集的训练样本数为 1 440,训练时间为 0. 234 4 s,此时视线落点估计的准确性为 97. 22% .

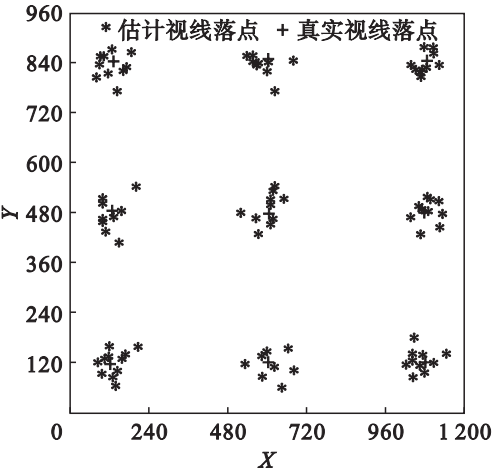


图 4 采用多项式方法的屏幕注视点分布图
Fig. 4 Gaze points on screen based on polynomial model

3.2 从视线落点分布的不同角度训练视线落点实验结果

本文通过对每行每一列及主副对角线数据分别训练并记录数据来证明采用 ELM 方法估计视点分布的准确性,如表 1 ~ 表 4 所示.

表 1 分别训练注视屏幕上每一行数据结果(100 神经元)
Table 1 Training data of staring at each row of calibration points (100 neurons)

位置	训练样本数	训练数据的准确性	实验验证的准确性
第一行	240	0. 985 7	0. 985 0
第二行	240	0. 975 0	0. 974 5
第三行	240	0. 987 5	0. 987 4

表 2 分别训练注视屏幕上每一列数据结果(100 神经元)
Table 2 Training data of staring at each column of calibration points (100 neurons)

位置	训练样本数	训练数据的准确性	实验验证的准确性
第一列	240	0. 941 7	0. 941 4
第二列	240	0. 929 2	0. 929 0
第三列	240	0. 887 5	0. 887 1

表3 训练主对角方向的三组数据的结果(100个神经元)
Table 3 Training data of staring at main diagonal calibration points (100 neurons)

位置	训练样本数	训练数据的准确性	实验验证的准确性
主对角方向 3组数据	240	0.966 7	0.966 1

表4 训练副对角方向的三组数据的结果(100个神经元)
Table 4 Training data of staring at secondary diagonal calibration points (100 neurons)

位置	训练样本数	训练数据的准确性	实验验证的准确性
副对角方向 3组数据	240	0.979 2	0.978 7

3.3 增加 ELM 的分类数量的实验结果

当对样本集进行9种分类时,屏幕上的分类区域相对比较大,所以实验中采用 ELM 多分类的特点来增加分类数目,进一步划分注视区域.该方法是:首先注视屏幕上均匀分布的16个位置;然后分别对屏幕上的16个不同区域进行特征参数提取;最后将采集到的16个区域中的平面视线参数进行 ELM 训练.隐层神经元数为200个,训练样本数为1440个,此时训练时间为0.956 3 s,视线屏幕落点的准确率为82.81%.

综合上述实验结果表明,采用 ELM 方法可以判断出人眼在屏幕上的注视位置,即实现视线落点的近似估计.当采用 ELM 训练1440个样本进行9种分类结果时,即采集的数据里面包含注视9种位置的数据信息时,从不同角度表明 ELM 方法的分类准确性基本上在90%以上.当使用者的注视点位置横向变化和纵向变化时,ELM 分类的准确率基本上都在97%以上.当增加训练分类数量为16,从不同角度表明采用 ELM 方法进行视点位置分类的准确性基本在80%左右.同时实验也验证了采用不同的隐层神经元数量会产生不同的分类准确性,为了提高分类准确性,一般采用较多的神经元数,但训练时间也会增多,如表5所示.因此本文采用准确性和训练时间折衷的方法,

表5 采用不同神经元数训练的实验结果
Table 5 Experimental results of network training by using different number of neurons

变量	100 神经元	200 神经元	300 神经元	400 神经元
训练样本数	1 440	1 440	1 440	1 440
训练数据的准确性	0.654 2	0.972 2	0.988 9	0.995 8
训练数据时间	0.109 4	0.234 4	0.453 1	0.843 8

即在分类结果为9种时一般选用隐层神经元的数量为200个,同时采集的样本数最好大于1000.由实验结果还可以看出,由于 ELM 适合处理多分类问题,将屏幕划分成若干区域,基于 ELM 模型的视线落点估计比常用的基于非线性多项式的视线落点估计更精确,并且视线落点很稳定.

4 结 语

本文基于极限学习机 ELM 提出了一种单摄像机视线追踪系统从视线参数到视线屏幕落点的估计方法.利用 ELM 的快速性和适于多分类的特点,将视线参数作为输入,视线的屏幕落点作为输出,通过初始标定数据训练 ELM,得到视线参数到视线屏幕落点之间的映射关系.该方法主要适合于需要稳定判别视线落点的应用场合.本文方法同样适合于其他类型的视线追踪系统,如双环形主动红外光源视线追踪系统.ELM 模型虽然可以将屏幕分类为不同的区域进行映射,但分辨率(即将屏幕区域分成多少 ELM 可以映射小区域)同样依赖于视线参数的可分性.因此在今后的工作中将在这方面继续研究.

参考文献:

- [1] Duchowski T. Eye tracking methodology: theory and practice [M]. New York: Springer-Verlag, 2003.
- [2] 赵其杰,张国栋,程德富,等.视线跟踪系统光源位置与视线分辨参数关系研究[J].系统仿真学报,2011,23(10):2239-2247.
(Zhao Qi-jie, Zhang Guo-dong, Cheng De-fu, et al. Relationship between light position and gaze identified parameters in infrared TV based eye gaze tracking system [J]. Journal of System Simulation, 2011, 23(10): 2239-2247.)
- [3] Sirohey S, Rosenfeld A, Duric Z. A method of detecting and tracking irises and eyelids in video [J]. Pattern Recognition, 2002, 35: 1389-1401.
- [4] Morimoto C H, Mimica M R M. Eye gaze tracking techniques for interactive applications [J]. Computer Vision and Image Understanding, 2005, 98: 4-24.
- [5] Shao G J, Che M, Zhang B Y, et al. A novel simple 2D model of eye gaze estimation [C] // 2010 Second International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC 2010). Nanjing: IEEE Computer Society Press, 2010: 300-304.
- [6] Zhu Z W, Ji Q, Kristin P B. Nonlinear eye gaze mapping function estimation via support vector regression [C] // Proceedings of the 18th International Conference on Pattern Recognition (ICPR 2006). Hongkong: IEEE Computer Society Press, 2006: 3443-3448.
- [7] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications [J]. Neurocomputing, 2006, 70: 489-501.
- [8] Huang G B, Zhou H M, Ding X J, et al. Extreme learning machine for regression and multiclass classification [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2012, 42(2): 513-529.