系（ 专业 ） 自动化

系 (专业) 主任 张爱民

批 准 日 期 2018-03-08

**毕业设计(论文)任务书**

电子与信息工程 学院 自动化 系(专业) 46 班 学生 付天爽

毕业设计(论文)课题 基于相关熵的注视点估计算法研究

毕业设计(论文)工作自 2017 年 11 月 1 日起至 2018 年 6 月 13 日止

毕业设计(论文)进行地点： 电信学院人机所

课题的背景、意义及培养目标

人眼注视估计主要是检测人眼注视的位置的过程，注视估计在各种各样的学科中使用眼球追踪技术，包括认知科学、心理学（特别是心理语言学、视觉世界的范式），人机交互（HCI）、市场研究和医学研究（神经诊断），语言阅读，音乐阅读，人类活动识别，感知的广告，和运动的发挥。

本毕业设计桌面为应用环境，运用基于相关熵的误差函数实现对人眼注视点进行估计。本毕业设计引导学生掌握回归分析的一般思路，并通过优化算法对参数进行准确估计，最终实现人眼注视点位置的估计。让学生理解掌握一般在机器学习算法，并进行算法实现与评价，希望使学生初步具有分析问题和解决问题的能力。

设计(论文)的原始数据与资料

人眼图像提取、对齐算法程序

相关熵分析相关算法程序

课题的主要任务

1. 实现一种基于相关熵的人眼注视点估计算法；

2. 实现一种核参数选择算法；

3. 实现桌面环境注视点估计演示程序；

4. 翻译指定英文文献、撰写毕业设计论文。

课题的基本要求(工程设计类题应有技术经济分析要求)

1. 掌握注视点估计的一般原理；

2. 掌握一般优化方法；

3. 对比多种估计方法。

完成任务后提交的书面材料要求(图纸规格、数量，论文字数，外文翻译字数等)

1. 规范的毕业设计论文

2. 相关程序源代码及说明文档

3. 相关外文论文翻译

主要参考文献

1. Feng Lu, et al., "Adaptive Linear Regressionfor Appearance-Based Gaze Estimation", IEEE Trans. on PAMI, 2015.

2. Erion Hasanbelliu, et al., "Information Theoretic Shape Matching", IEEE Trans. on PAMI, 2014.

指导教师 张雪涛

接受设计（论文）任务日期 2018-03-08

（注：由指导教师填写） 学生签名：

**西 安 交 通 大 学**

**毕业设计(论文)考核评议书**

电子与信息工程 学院 自动化 系(专业) 46 班级

指导教师对学生 付天爽 所完成的课题为 基于相关熵的注视点估计算法研究

的毕业设计(论文)进行的情况，完成的质量及评分的意见： 论文针对桌面环境下的注视点估计问题展开研究，采用相关熵构建了一种高精度鲁棒的注视点估计算法，并对相关熵核参数的选择方法进行的专门研究，作者在实际数据集上验证了方法的有效性。论文写作认真，较好地完成任务规定的工作。在毕设工作期间，态度积极主动，善于探索问题和解决问题，有效地锻炼了动手能力。

指导教师建议成绩： 良好

指导教师 张雪涛

2018 年 6 月 11 日

**毕业设计(论文)评审意见书**

评审意见：论文针对桌面环境注视点估计存在的问题，实现了小样本下基于相关熵的注视位置估计算法，并对这一算法的核参数提出了基于优化算法的选择调节方法。论文写作认真，表达清晰，有一定工作量，达到本科毕业设计的要求，同意参加本科毕业设计答辩。

评阅人建议成绩： 良好

评阅人 杜少毅 职称 副教授

2018 年 6 月 9 日

**毕业设计(论文)答辩结果**

电子与信息工程学院 **院**

自动化 **系(专业)**

毕业设计(论文)答辩组对学生 付天爽 所完成的课题为 基于相关熵的注视点估计算法研究

的毕业设计(论文)经过答辩,其意见为 该论文研究了基于相关熵的注视点估计。论文论述清楚，条理清晰，图表丰富，答辩讲述清晰明确，回答问题合理正确，是一篇良好的毕业设计论文，达到了本科毕业设计要求。答辩委员一致同意，通过其毕业答辩。

并确定成绩为 良好

毕业设计(论文)答辩组负责人 梅魁志

答辩组成员 孙宏滨 梅魁志

兰旭光 任鹏举

魏平

2018 年 6 月 13 日

**论文题目：基于相关熵的注视点估计算法研究**

**学生姓名：付天爽**

**指导教师：张雪涛**

摘 要

注视点估计在当今诸多研究和实际应用领域有着巨大的潜在价值。人眼注视位置的估计可以应用在包括人机交互，认知与心理研究，市场研究，医疗辅助等行业。注视点的估计技术在当前存在各种各样的限制与不足，尤其是在不同的实际条件和应用场景中，其表现还有较大改进空间，这也使得相关的方法研究受到了较多的关注。当下对注视点估计的方法主要有基于模型的方法和基于表观特征的方法两种。前者需要特定设备对人眼建模，而后者只需要普通相机。

随着图像处理和机器学习水平的发展，基于表观特征的方法获得越来越多的关注，本文的相关内容也是在这类方法基础上，在典型的桌面环境（用户与屏幕交互）进行的。传统的估计算法训练样本很多，但不能很好地克服噪声，且参数需要人工干预来选择。本文针对桌面环境注视点估计存在的问题，使用MATLAB编写相关程序，实现了少量样本下一种误差更小的，基于相关熵作为误差函数的注视位置估计算法，并对这一算法的核参数提出了基于优化算法的选择调节方法。

本文的主要工作内容包括：

1、实现了从用户图像采集处理到基于相关熵的注视位置估计的全套算法MATLAB程序；该方法相比于传统回归方法，能很好地适应非高斯噪声，以获得更好的估计结果，所需样本数量也相对较少。本文通过将计算结果的角度误差与自适应回归方法对比，验证了这一结论。

2、针对上述算法，设计并实现了使用衰减法和模拟退火法寻优的核参数自适应选择方法，得到了其合理的取值范围。该成果使本文的研究避免了人工设定参数和多次比较选择，简化了实验的流程，并有效提高了程序运行的精度。

**关 键 词**：注视点估计；相关熵；核参数；模拟退火

**Title:** **Appearance Based Eye Gaze Estimation via Correntropy Algorithm**

**Name: Tianshaung Fu**

**Supervisor: Xuetao Zhang**

ABSTRACT

Gaze estimation has great potential value in many researches and applications. The estimation of human eye gaze can be applied in such industries as human-computer interaction, cognitive and psychological research, market research, and medical assistance. There are various limitations and deficiencies in this technology, especially under different conditions and scenarios. Since its performance still needs to be improved, relevant researches are getting increasingly popular.There are two methods for gaze estimation: the model-based method and the appearance-based method. The former requires a specific device to model the human eye, while the latter only requires an ordinary camera.

With the development of image processing and machine learning, more attention has been paid to appearance-based methods. This paper is also based on this type of method and is performed in a typical desktop environment (user-screen interaction). Traditional estimation algorithm trains many samples but does not overcome noise well, and the parameters need manual intervention to select. Aiming at the problems in the gaze estimation of desktop environment, we used MATLAB to program a gaze estimation algorithm in this paper. It is based on correntropy, and we received a smaller error using fewer training samples. We also propose optimization methods for the self-adaptive selection of kernel bandwidth in this algorithm.

The main achievements of this article include:

1. A complete MATLAB program, from the user's image acquisition processing to the correntropy-based gaze position estimation, is implemented; this method can well adapt to non-Gaussian noise and obtain better estimation results compared to the traditional regression methods. It requires fewer samples as well. We verified this conclusion by comparing the angle error with ALR method.

2. For the algorithm above, we designed and implemented an adaptive selection method of kernel bandwidth using attenuation method and simulated annealing method, both obtained a reasonable range of values. This makes the research in this paper avoid manual setting and multiple comparison of parameters, simplifies the experimental process, and effectively improves the accuracy of program operation.

**KEY WORDS**: Eye gaze estimation; Correntropy; Kernel bandwidth; Simulated annealing

目 录

[1 绪论 1](#_Toc492980684)

[1.1研究背景和意义 1](#_Toc492980690)

[1.2国内外研究现状 1](#_Toc492980690)

[1.3研究内容和组织结构 3](#_Toc492980690)

[2 基于表观特征的注视点估计方法 4](#_Toc492980692)

[2.1 注视点估计的基本原理 4](#_Toc492980693)

[2.2 传统注视点估计方法 6](#_Toc492980693)

[2.3 本章小结 7](#_Toc492980693)

[3基于相关熵的桌面环境注视点估计 8](#_Toc492980695)

[3.1 问题描述 8](#_Toc492980696)

[3.2 人眼图像处理和降维 8](#_Toc492980696)

[3.3 基于相关熵进行注视点估计 10](#_Toc492980696)

[3.3.1 相关熵定义 10](#_Toc492980697)

[3.3.2 基于相关熵的样本重构 11](#_Toc492980697)

[3.3.3 程序实现 13](#_Toc492980697)

[3.4 实验结果及分析 14](#_Toc492980699)

[3.5 本章小结 15](#_Toc492980693)

[4 基于优化算法的核参数选择 16](#_Toc492980698)

[4.1 核带宽介绍 16](#_Toc492980699)

[4.2 模拟退火算法简介 17](#_Toc492980693)

[4.3 参数选择和优化 17](#_Toc492980699)

[4.3.1 参数选择方法 18](#_Toc492980700)

[4.3.2 程序实现 20](#_Toc492980700)

[4.4 实验结果及分析 21](#_Toc492980699)

[4.5 本章小结 21](#_Toc492980693)

[5 结论与展望 22](#_Toc492980701)

[5.1结论 22](#_Toc492980702)

[5.2展望 22](#_Toc492980702)

[致 谢 24](#_Toc492980731)

[参考文献 25](#_Toc492980731)

# 绪论

人眼的注视点估计技术在人机交互（Human-Computer-Interface，HCI）[1]，市场研究，医学研究[2]和人类心理认知[3]等领域，都有着巨大的潜在应用价值。尤其是随着技术进步，科技和商业公司对用户体验的高效便捷性的追求越来越极致，这使得眼动控制成为了当前热门研究领域之一。由于所需设备简单，实现起来也相对容易，因此本文主要关心的也是基于表观特征的框架，在桌面环境下的注视点估计。在这种用户与屏幕的交互场景中，一般有着较稳定的头部姿态和较小的环境因素干扰。在此基础上，本文主要解决基于相关熵的注视点估计算法的原理和实现，及其核参数优化选择问题。本章介绍了当前该领域的一些研究背景和成果，分析了遇到的问题和挑战，还对本文的组织框架和推进线索进行了梳理。

## 研究背景和意义

计算机软件和辅助系统，心理学与认知科学，医学诊断治疗，残疾人辅助，客户行为分析……鉴于注视点估计的这些应用潜力，国内外在这一技术的不同实践领域已经有了不少的研究成果。事实上早在上世纪八十年代，就已经出现了使用特定仪器捕捉人眼注视点的人机交互方式[1]。但也正是因为应用场合的广泛与多变，使注视估计的具体实现方案和流程各有长短，各类方法也在竞争互补中共同进步。

注视点估计实现方案主要分基于模型和基于表观特征两大类。随着机器学习技术的发展，近期的注视点估计相关研究大都开始基于表观特征，这也使得普通用户能够逐渐接触和使用眼动控制的便捷交互方式。虽然基于表观特征的估计方法只需要普通的摄像头就能实现，但当前的注视点估计技术在不使用红外光的条件下，还很难做到对各种环境和条件都很好地适应。如何在桌面环境实现小样本高精度的注视点估计，是该领域的研究重点，也是本文的主要工作之一。

在桌面环境实现少量样本快速标定和高精度注视点估计，将可能使实时眼动控制成为新的基础交互逻辑。这在计算机软硬件商，信息产业，以及商业、医疗等行业都将带来新的突破和机遇。

## 国内外研究现状

早期的注视点估计方案一般是基于模型的，即对人眼进行三维建模，然后利用一些小尺度的特点（如近红外反射亮斑，瞳孔中心，虹膜轮廓等）与注视位置建立映射[4,5,6,7]，从而进行估计，如图1-1所示。这些方法一般需要高分辨率相机或者近红外光源等特殊设备，才能对人眼进行拟合，可见它们并不具备易操作性和普遍性。



图1-1 基于模型的方法利用近红外反射斑或者椭圆拟合

近年来，图像识别技术的发展使得基于表观特征（人眼图像）直接进行估计的方法开始盛行。这类方法在桌面环境下，可从普通摄像头获取的眼部图像提取特征，直接回归分析得到注视位置，如图1-2所示。回归的部分可以使用支持向量回归[8]或者是神经网络[9]，用大量的训练样本获得较好的估计精度。新近提出的方法则是引入了线性插值[10]，通过证明人眼特征空间和注视点平面空间的相似性，将人眼测试集和训练集的权重关系共享到注视点空间，借助对注视位置的加权重构来代表估计。此类方法通过对回归方式和插值过程的方法优化，可以进一步减少训练样本。比如使用最近邻选择（K-Nearest-Neighbors，KNN）算法[10]，Williams提出的标记数据的半监督高斯回归[11]，或是陆峰等人使用的自适应线性回归（Adaptive Linear Regression，ALR）方法[12]。近来也有新型的误差函数被用于相关研究。本文的方法则是使用了相关熵这一概念作为相关性的误差函数来重构样本。

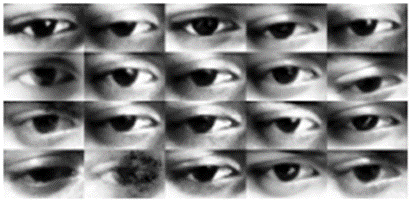


图1-2 基于表观特征的方法需要一定训练样本

依靠较少的训练样本实现较为精确的注视点估计，将会使得该应用在操作性方面有很大提升。那么该项技术将有望应用于各类交互场景，实现更便捷的控制，特别是针对常见的浏览和选择行为，作出相关分析反馈。另外，相关的成果还能与残疾人辅助，游戏娱乐，自动驾驶等热门技术领域相结合，有巨大的科学和商业价值。因此，这方面的探索有重要的实际意义。

综上，现在有许多方法来实现注视点估计，尤其是基于表观的方法得到了较大推进，但要在较少训练样本的前提下实现较高精度的估计仍然面临困难。这也是本文所做工作尝试要解决的东西。该技术要较好地适应不同应用场合和多变环境下的具体需求，还需要更为持久深入的研究。

## 研究内容和组织结构

上文提到，基于表观特征的注视点估计，重点在于人眼特征空间和注视点空间能够共享测试集的插值权重[10]。因此求解权重向量的过程，就是注视点估计的过程。本文的研究主要针对桌面环境，即保证较小头部姿态变化和环境干扰时，实现了一种基于相关熵的注视点估计算法，并且基于优化算法实现了该算法的核参数自动选择。通过验证，该方法不仅精度相对较好，所需训练样本少，实现起来也方便。基于这几点，本文主要内容安排如下：

第一部分是基于相关熵的注视点估计算法实现。主要包括以下几个流程的介绍：由于此类方法对眼部图像提取精度要求较高，需要先准确识别提取人脸眼部区域；然后对人眼特征降维，验证共享权重的理论依据，即人眼空间与注视空间相似性。最后，本文引入了相关熵作为误差函数，求解权值从而实现了注视位置的估计。

第二部分是算法的参数优化。使用基于相关熵的方法进行估计涉及到核带宽参数，其取值很大程度决定了算法的性能表现。之前的研究中，大多采用了人工设定来选择该参数的具体取值，需要多次比较结果。而本文设计实现了该参数的自动寻优方法，使得算法运行时可以自适应获取最优结果，无需人工调整。

本文共分五个章节，涉及方面分别如下：

第一章：绪论。主要介绍了注视点估计的研究背景及其实际意义，简单回顾了该领域研究的发展和国内外现状，以及当前方法各自的优缺点。概括了本文的研究重点以及内容安排。

第二章：基于表观特征的注视点估计方法。主要介绍了本文涉及的注视点估计问题的理论基础和背景知识，以及相关领域的研究。包括基于表观特征的注视点估计的原理，基本框架，以及几个典型方案的介绍和分析。

第三章：基于相关熵的桌面环境注视点估计。主要介绍了本文用相关熵作为误差函数的原因及其特点，以及桌面环境下基于相关熵实现注视点估计的具体流程、方案和优势。介绍了对该部分基于MATLAB的程序实现，并将实验的结果与一般方法做了对比和分析。

第四章：基于优化算法的核参数选择。本章介绍了核参数的特点以及对实验结果的影响，并设计实现了使之自适应选择取值的几种优化算法。同时介绍了这些方法的程序实现和实验结果，验证了自动参数优化的优势。

第五章：结论与展望。对本文的工作进行了回顾、总结和评价，得出一般性的结论。对仍然存在的问题和今后该领域的工作提出了展望。

# 基于表观特征的注视点估计方法

本章主要对注视点估计相关的概念，原理，方法进行简单的介绍，以作为后文实验过程的理论基础，并提供支撑和比对。其中包括基于表观特征实现注视点估计的基本框架，以及当前研究的一些典型方法的简介。

## 注视点估计的基本原理

本文所提的基于相关熵的方法，也属于基于表观特征的注视点估计。所谓基于表观特征，是指整个算法只需要借助普通相机拍摄的用户图像，实现简单。具体来说，这类注视点估计方法都是对拍摄的人眼图像进行特征提取，再利用人眼特征空间和注视点空间的相似性[10]，共享人眼训练样本对测试样本的插值权重，再把权重向量映射到注视平面对屏幕坐标位置加权，完成注视位置的估计。

显然在这个过程中，人眼的训练图像样本要与屏幕上相应的标定坐标一一对应。即在估计算法之前对于用户存在一个标定阶段，采集注视屏幕特定位置的眼部照片。算法的输入是图片样本（训练集和测试集）和坐标（标定位置），最终的估计结果也体现为屏幕上的一个（一组）坐标。在实验中，误差取真实坐标与上述计算结果的距离。

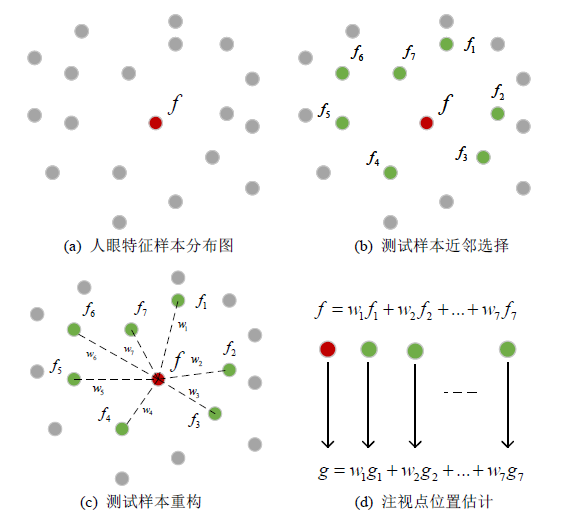


图2-1 注视点估计基本原理示意

在注视点估计的实际操作中，是先由图像提取得到两组特征，再分别输入特征进行估计的。这里不妨假定已经得到了训练集和测试样本的特征向量和。如图2-1（a）（b）（c）的过程所示，将人眼测试样本用邻近的训练样本集{,,……, }的加权来表示，由于两边都已知，这个插值将能够得到一组权值向量。用公式表示如下：

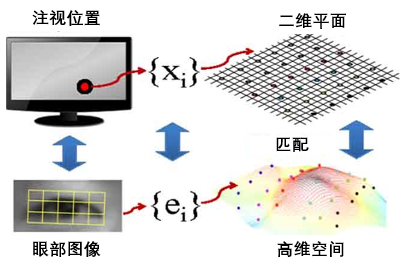
（2-1）

且满足约束条件：

在人眼特征空间和注视点空间的相似性得到验证（在第三章详细描述）的情况下，就可以在两个空间共享权值向量。用同样的加权映射到注视平面，利用现有坐标位置再次插值，对注视的位置进行估计，如图2-1（d），即：

（2-2）

图2-2进一步阐述和展示了这一基本框架。从普通相机拍摄的眼部图像就可以直接处理，提取出相应的特征。接着可以看到，处理后的人眼图像特征所表现出的周期性与注视点平面对应的分布具有相似性。这个直观的结果可以通过对人眼特征的降维操作进一步验证，该部分将在第三章涉及。



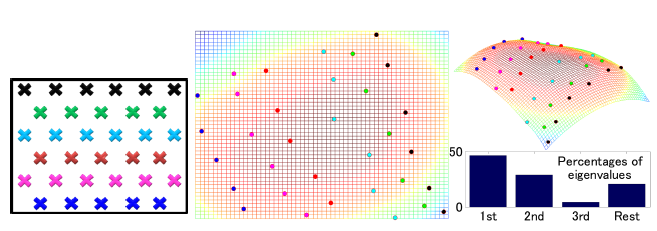


图2-2 注视点空间与人眼特征空间中，标定点呈现类似的周期性

至此，此类注视点估计问题的核心就完全落在了如何求解两边所共享的权值上。这就是注视点估计的理论基础和总体框架。而在具体实现中，重要的是如何得到这样一组最优解，能使加权的误差最小。对此，实验中可以采用领域约束，自适应线性回归(ALR)[12]，或者本文提出的相关熵等方法来求解。当然，也可利用神经网络和机器学习等方式，从人眼图像直接采取回归操作进行估计，但这不是本文研究范畴。

## 传统注视点估计方法

下面简单介绍基于上述框架的一些典型估计方法，在绪论中已经提及过。依据共享权重的原理，这些方法的关键点都表现在如何求解最优的插值权重上。它们的基本思路和优缺点为本文使用的基于相关熵的方法提供了很好的借鉴和比较价值。

要得到一组重构后误差最小的权值向量，最简单直接的方法无非是直接使用回归，比如基于最小均方误差（MSE）法。在采用大量训练样本的且噪声影响较小情况下，这不失为一种解决途径，当然它的低鲁棒性和粗糙性也显而易见。基于MSE，Tan等人采取了增加最近邻约束（KNN）的方法[10]，仅使用离测试点距离最近的K个训练样本来进行加权插值。

此外，也有从样本特征直接进行回归操作的方法。其中一种是引入支持向量机（SVM）[8]，将特征投射至更高维空间，从而寻找一种对应输出的映射关系，一般称之为核函数。也可使用高斯过程回归（GPR），基于机器学习，利用样本的分布来预测函数空间。对于注视点估计这种二维的问题，不妨构造两个一维的GPR代表空间分布，进而使用贝叶斯推理。此外，经典的神经网络模型也能用在这里。

最近提出的一种典型的方法是本文已经多次提到的，由陆峰等人提出的自适应线性回归（ALR）[12]方法。这个方法的主要优势在于保证结果的同时，能够减少训练的样本数。由于人眼球的运动只存在两个自由度，注视屏幕也是二维平面，所以该研究通过寻找一个转移矩阵来完成两个空间的匹配，表示如下：

***AE***=***X*** （2-3）

其中***A***是一个转移矩阵，用于匹配两个空间的特征向量；***E***和***X***分别是通过插值得到的人眼样本特征向量和注视位置向量。正如上文所阐述的，式（2-3）成立的基础是***E***和***X***是由同一组权值计算得到的，由于在这里最终估计位置是未知的，求解该转移矩阵的本质还是为了求解权值向量来表示，即：

（2-4）

该方法基本上将注视点估计的重点建立在一个回归问题上，而不是对人脸特征的分类。相比之前的方法，采用线性回归不仅能保证准确性，还大大减少了样本数。

上述这些经典的方法从不同角度出发，在基本框架下不断进行优化改进。这些工作为本文的进一步探索打下了坚实的基础，提供了许多借鉴，也是我们用以比较反思的对象。

## 本章小结

本章对基于表观特征的注视点估计方法的基本原理和框架进行了阐述。重点介绍了权重共享映射的原理和理论基础，以及估计问题的核心，为引入相关熵方法作了准备工作。另外，以自适应线性回归（ALR）为例，介绍了该框架下的传统方法，为本文工作提供了参考和比较的依据。

# 基于相关熵的桌面环境注视点估计

## 问题描述

本文所研究的注视点估计问题，是指在桌面环境下，基于表观特征的注视点估计。这种条件使得我们可以不考虑头部姿态的变化和光照等环境变量带来的影响。在人眼特征空间和注视点空间相似性的基础上，本文利用了相关熵作损失函数能够适应非高斯噪声的特点，求解共享权值从而完成注视位置估计[13]。该方法不仅提高了估计的精度，同时所需训练样本数相对较少。在实验中，主要包括使用MATLAB程序实现的图像采集处理，降维比较，权值求解几个部分，如图3-1所示。

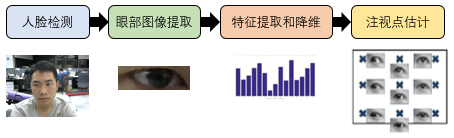


图 3-1 注视点估计实验流程

## 人眼图像处理和降维

基于表观特征的注视点估计只需要普通相机采集用户图像，就能提取特征进行注视点估计。显然，采集的图像需要尽量保持规范，尤其是用户脸部的变化应当尽可能小，否则使用图像提取的特征会造成较大的误差。事实上，不少研究也探讨了如何克服微小头部变化的影响，本文暂不涉及。

为了准确地提取眼部图像，需要先识别图像中的人脸，消除背景的干扰。目前有许多相对成熟的人脸识别方法，本实验使用MATLAB的Face Parts Detection 工具，对采集的样本进行了批量的脸部识别和裁剪。该工具的优势在于能定位五官区域，为对齐眼部图像作了更好的准备。这类识别一般采用的是Ada-Boost算法[14]，即用多个弱分类器的结果投票加权，来获得一个较好的分类器。在粗提取得来的眼部区域，再使用边缘滤波裁剪出眼部图像。边缘滤波Edge Filter使用sobel算子检测边缘，这种基于灰度梯度的方法能够较好的对齐眼睛边缘和眼角。在某些情况下，也可以把精度提高到像素级图像提取，以获得更好的样本。

接着，为之前得到的图像提取HoG（Histogram of Oriented Gradient，方向梯度直方图）特征。具体提取方法是将转成灰度图的输入分成小的区域，也就是所谓的细胞单元（Cell）。然后在一个较大区域（Block）内，获取细胞单元里各个像素点的梯度方向或边缘方向密度分布来构成直方图（图3-2），该直方图代表了相应角度范围内的各像素点的权重投影。最后把这些直方图组合起来，就可以定量描述图片特征，也就是一个特征向量。

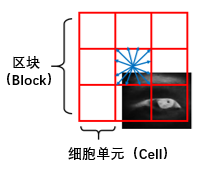


图3-2 HoG特征的提取示意

下一步是对得到的特征向量使用转换矩阵进行降维，从而能得到人眼特征空间和注视点空间的相似性。人眼特征空间的维度较高，包含了许多信息，考虑到平面位置和眼部运动自由度都只有两个，所以与注视平面的匹配还是以二维信息为主。至于高维特征还是否保留，可以取决于具体实验对细节信息丢失的敏感程度。

总之，由于这两者的结构本身还是具有比较大的差别，所以必须对人眼图像先降维处理后，才可能支持插值的权值在两边共享。即对于降维之后的人眼特征投影空间，其空间结构的衡量应当与注视点平面类似[13]。本文在该部分使用了KL散度降维[15]，这里所谓的KL散度是对两个分布差异性的一种描述。其数学表示为：

（3-1）

其中P、Q为两个概率分布为和)的变量，等式左边的 即为P关于Q的KL散度。其取值越小，则P与Q的分布越接近彼此。

在这里通过一个半正定的转换矩阵C来投影出人眼特征空间。特别地，本文以KL散度作为度量，通过对训练样本的学习，寻找一个合适的半正定矩阵（PSD），来使投影后的人眼特征空间尽可能接近注视点空间，也就是取得最小的KL散度值。该过程可用公式表示如下：

s.t. A （3-2）

其中， 是注视点空间的条件分布， 是投影后的人眼特征空间的条件分布，它们都是通过各自空间内的距离测度来计算的，可以根据实际训练和学习的情况具体地定义。

从下图（图3-3）可以看出，对人眼特征空间采取降维的操作以后，在投影的平面表现出了和注视点空间相似的特点和周期性。这样的结果验证了上一章所论述的注视点估计的基本原理，也是为后面的估计部分作了理论上的保证和铺垫。在这一基本框架下，本文才得以引入相关熵的概念完成注视点的估计。

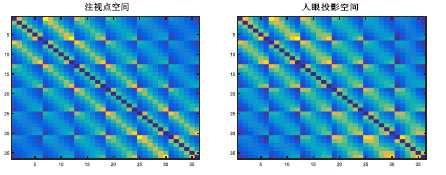


图3-3 注视点空间（左）和投影后人眼特征空间（右）的相似性

## 基于相关熵进行注视点估计

由于其本身反映相关性的特点，相关熵可以在机器学习等领域中作为一种相关函数使用。在第二章所介绍的框架下，本文的注视点估计过程中同样需要进行插值和映射，所以本文使用了相关熵这一概念，来作为求解权值向量时的误差函数。其主要优势在于能够比传统的MSE方法更好地适应噪声[16]，尤其是训练样本常出现的非高斯的头尾部噪声。

### 相关熵定义

基于信息论的模型近来在机器学习，特别是模式识别领域受到越来越多的重视。而“熵”的概念在信息论学习中，一般表示随机变量的出现的期望[17]。在本文中，相关熵作为一种相关函数[18]，被用作一种新的损失函数的计算方式，它被用作计算两个随机变量的相似性，在本文指的就是人眼图像样本间的相似性。

假定两个随机变量X和Y，那么它们的相关熵公式表示如下：

（3-3）

其中，E[ ]—表示函数期望值；—表示一个将其投射到高维空间的核函数；—表示X和Y的联合概率密度分布。

在实际中，由于两个随机变量的分布未知，因此本文实际上已经提供样本X和Y，那么相关熵可以表示为如下公式：

（3-4）

其中，=，采用高斯核函数[14]。

基于相关熵准则的误差函数，相比于传统的最小均方误差方法（Mean-Squared-Error，MSE），能更好地适应并克服本实验训练样本中不同分布的噪声[16]。因为MSE以平方形式增长，对头尾快速衰减的样本能够较好处理，却难以适应非线性或者头尾部的噪声；相比之下相关熵的函数衰减主要取决于核函数，加上它对高阶矩也进行量化，所以有它处理非高斯非线性分布的优越性，如图3-4所示。

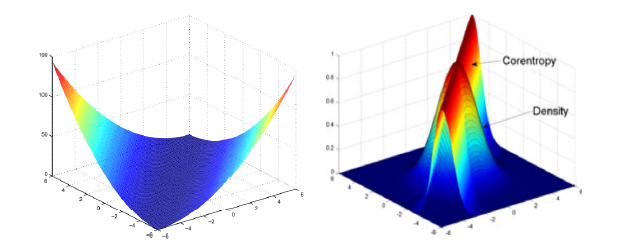


图3-4 MSE（左）与相关熵（右）的损失函数示意

### 基于相关熵的样本重构

按照框架，基于表观特征的方法本质上是通过共享权重实现人眼特征到注视点平面的映射[10]。本文所说的使用相关熵，属于求解权重的一种方法，是针对其中的一个步骤进行的优化。方法的原型，或者说本质仍然是进行映射和加权。这类注视点估计的映射原理都是如图3-5所示。上文的工作已经完成了从图像到二维特征的跳跃，而下文将借助相关熵求解插值权重，将坐标加权反映出对注视位置的估计值。

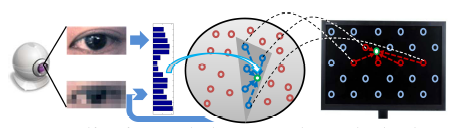


图3-5 共享来的权重用于从已知坐标估计实际注视点

正如第二章所描述的那样，通过对定位裁剪处理好的人眼图像进行HoG特征提取，可以获得人眼的训练集{}，同样地，对于新输入的一个测试样本，此处也可通过训练集加权来插值表示，由于所有样本（特征向量）都是已知的，于是便能借助（3-5）解出权重：

（3-5）

根据前一节的论证，凭借降维投影所证实的相似性，便可以将从人眼图像求得的权值共享用到注视点空间，从而用屏幕上邻近的参考标定点坐标（对应人眼训练样本）来加权，进而得出估计的注视点位置，即：

（3-6）

要获得最优的估计值，需要求解最优的权值向量。具体来说，在使用式（3-5）计算求解权重向量时，本文提出使用相关熵作为评价函数，在最小误差的条件下对测试集的人眼特征进行线性插值[13]，这一重构过程采用了如下公式：

J(w)= （3-7）

其中， 是人眼特征的训练样本集； 是人眼测试样本， 是高斯核的带宽，即为用训练集插值表示测试集的加权向量。

如果需要进一步减少需要的训练样本数量，在具体实验的插值过程中，可以参考KNN方法，仅选择与测试样本最邻近的K个训练集进行加权。这样在实际使用中并不会过多影响估计精度，但只会用到很少的样本。改进后带有局部约束的公式可以表示如下：

J(w)= - （3-8）

式中 通过测试样本与每个训练样本的距离，进行了局部邻域约束。公式（3-7）和（3-8）阐述了本实验是如何利用相关熵来衡量训练集和测试集的相似性的，并以此为依据求解权重 ，用于注视点的估计。

具体到实验中，在开始的标定阶段，用户所注视的屏幕上就分别显示4，9，16，25，36个标定点，其分布要具有代表性和规律性，如图3-6所示。这些点的坐标对应了用户注视它们时的人眼图像，作为训练集输入算法。当测试集被插值表示出以后，也是依靠这些已知的坐标位置，用加权才能表示出注视位置。

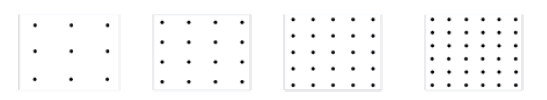


图 3-6 标定阶段屏幕上显示的不同数目的注视位置

在求出权值以后，通过对应的标定点坐标加权估计当前注视位置就相对容易了，如图3-7。所以实验的重点步骤，尤其是在MATLAB程序的编写过程中，还是集中在求解权重的过程上。

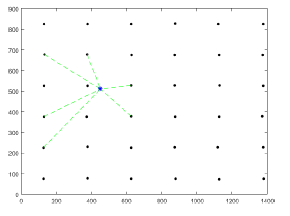


图3-7 利用加权完成的注视点估计结果

从上图也能看出，实际操作一般只需要邻近的少量样本就能恢复出实际的注视位置，不需要特别多的训练样本。所以公式（3-8）使用邻域约束有一定的合理性。当然，这也多少是因为本文研究限定于桌面环境下，相对来说环境和头动的影响小，所以样本比较准确和固定。

### 程序实现

本实验中，在MATLAB编程实现的部分包括人眼图像的提取处理，图片特征提取，降维，以及基于相关熵的权值的计算。大部分功能都以单独的函数实现，以供主程序依次调用。该程序的流程图如图3-8所示。

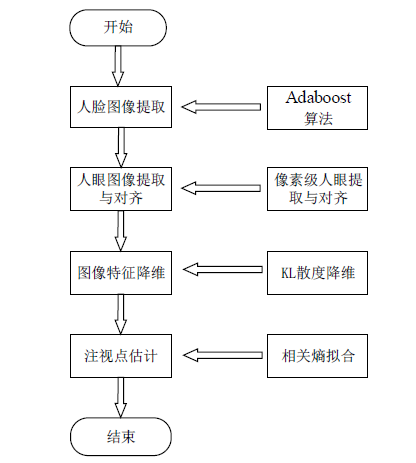


图 3-8 注视点估计程序的流程图示意

在程序实际运行中，是将事先提取处理好的人眼图片按照先训练样本，再测试样本的顺序导入主程序进行HoG特征提取。同时导入的还有训练和测试时相应的注视坐标。这样分模块独立工作有利于缩短实验时间，避免了在调试后面算法的过程中无谓地重复图像部分的工作。

在预设好的两个空数组中，通过相关程序，能够提取得到两个特征向量，即训练样本总特征，和测试样本总特征。之后再根据公式（3-6），调用相关熵的函数，进行最佳权值的计算。具体地，输入，以及核带宽参数的值，能够在相关熵子函数的返回值中得到理论上最优的权值向量。这个值即可用于在注视平面根据已有的标定点（输入坐标）来估计测试样本对应的用户的注视点。

本实验的误差是将加权估计出的坐标与输入的坐标进行距离比较，求出误差均值和标准差。本文具体使用的是基于距离的角度误差，即：

（3-9）

其中，是算法估计的坐标，是真实注视坐标，n是测试集数目。在本文后续的部分，也都是根据这一指标来对我们的算法以及参数选择进行分析和评价。

在后期的优化中，还可以在程序中用图像反馈核带宽参数的取值和实验误差的关系，有利于参数选择工作进一步提高结果的精确度。本文第四章中，将会详细地阐述该部分的内容和实验过程。另外由于进行HoG特征提取时的特定要求，图片的长和宽需要是Block大小的整数倍，因而第一步裁剪所得到的眼部图片往往需要再次调整尺寸。这也给程序的运行时间带来了不小的增加。

## 实验结果及分析

本文使用基于相关熵的注视点估计算法，所需样本数少，并能够更好地克服噪声，进而获得更小的角度误差，达到较理想的估计精度。

按3.2到3.3所述，在MATLAB上编写相关的算法程序，对专门采集的用户图像进行处理，估计注视坐标。现将本文的算法多次实验运行结果的平均角度误差罗列如下，并与第二章介绍的ALR方法作比较，如表3-1所示。当标定数目偏少时，比如采用9点或者16点，本文方法的误差很小，这是因为克服孤立噪声的优势充分体现，而标定点增加后，选择邻近样本带来的损失也相对增大。

表3-1 基于相关熵的估计结果以及与自适应线性回归的比较

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 平均角度误差 | 9点 | 16点 | 25点 | 36点 |
| 自适应线性回归ALR | 0.32 | 0.21 | 0.25 | 0.30 |
| 基于相关熵CER | 0.12 | 0.13 | 0.11 | 0.20 |

为了更直观体现注视点估计的效果及其对比，图3-9显示了在屏幕上两种方法的估计位置与真实位置。这是仅随机选取单一测试样本，用36点标定的情况下进行的注视点估计。图中红色点是真实注视点，蓝色点是本文方法估计点，绿色点是作为对比的自适应线性回归（ALR）方法所得到的估计点。

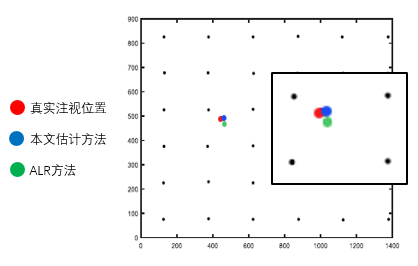


图 3-9 屏幕显示的注视位置比较（右侧为局部放大）

由此可见，使用本文的方法能够很好地改善一般情况下注视点估计的精度，这主要得益于损失函数对样本噪声的克服能力提高，以及对插值训练样本的合理邻近选取。要指出的是，在这里参与比较的两种方法都是按照经验直接指定核参数（均取=0.5）。在本文第四章，介绍了基于优化算法的核带宽自适应选择方法，还能进一步提高算法的精度。

总的来说，在样本数量合适且标定结果较好的情况下，本文使用的基于相关熵的注视点估计相比其他方法能更好地控制精度。但对用户头部姿态的较高要求也带来了一定的限制，尤其是训练样本偏少时的优势不明显。

## 本章小结

本章按照流程介绍了桌面环境下基于相关熵的注视点估计算法的实现，详细阐释了相关熵的具体定义及其优势，以及算法的细节。另外还分析了该实验过程所包括的图像处理，特征提取和降维，插值求解权重几个部分所运用的技术。最后，将本文算法的实验误差与传统方法进行对比，验证了其优势。

# 基于优化算法的核参数选择

在本文第三章已经完成了注视点估计的任务。由于在实验过程中，核参数（Kernel Bandwidth）发挥着重要的作用，其取值合适与否在一定程度上决定了实验结果的好坏，所以本章主要解决核带宽这一参数的选择问题，通过优化方法实现了在实验中自动选择最优参数，而不是多次由人工指定，降低了注视点估计的误差。

## 核带宽介绍

核带宽，即kernel bandwidth，在使用相关熵的权重线性重构公式中出现。核函数很大程度上决定了相关熵计算相似性的表现结果，故带宽参数在一定程度上控制着算法结果的好坏。许多研究中将核带宽称作“窗口”参数，因为该参数取值类似于观察相似性的窗户—评判尺度[19]。其取值越大，则观察尺度越大，有利于算法的收敛，但也会相应增大误差，且会因为更接近MSE的方法，从而使相关熵的优势无从体现；其取值越小，虽然对误差的包容小从而提高了部分精度，却也必然带来算法收敛时间甚至收敛性上的表现下降，如图4-1所示。

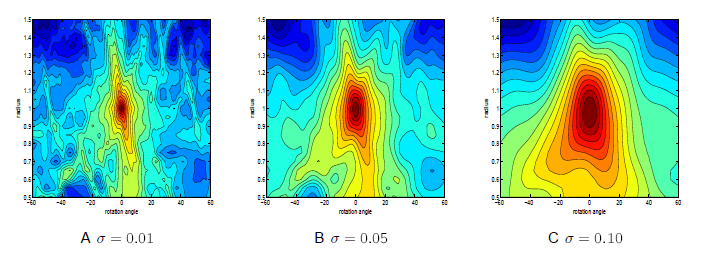


图4-1 不同核参数取值下的算法结果平面表现出不同特征

为了体现核带宽参数的重要性，有必要在这里再次列出第三章中，本文基于相关熵来求解权值向量的公式：

J(w)= - （3-8）

在上述公式中，除了要求解的权值向量，现有的测试样本和训练集，以及已知的约束条件之外，只剩下一个未确定的参数，也就是核带宽。一个好的参数取值能有效降低误差，所以其取值在很大程度决定了求解得到的w效果如何。

上文提到核参数本质上是一个窗，取值决定了该函数度量误差的尺度大小，会改变算法程序的各项性能。而即便是在其合适的取值范围内，我们得到的也会有不同的误差均值和标准差，并且有着不同的收敛速度。所以，参数的选择本质上也是一个决策和优化的问题。

在注视点估计的研究中，核参数的选择常常采取先人工指定，再多次运行比较选择的方法。早期提出了一些公式来估计核带宽的取值，Erion对图像处理问题中核参数的研究，采用了设置初值再衰减的方法[19]获得较好的取值。本文提出了估计过程中基于优化算法的自动参数选择，以简化实验流程并提高精度。具体地，本文基于MATLAB采取了衰减法和模拟退火法，在估计过程中对核带宽自适应寻优。

## 模拟退火算法简介

模拟退火算法（Simulated Annealing Algorithm），简称为SA，是一种基于概率的优化方法，最早是由N. Metropolis等在1953年提出。其原理和思想来源于物理学中固体的退火冷却过程。本质上，它基于蒙特卡洛迭代求解策略，结合了按时间（即温度）变化的概率突跳来避免局部最优解，最终趋向于全局最优解。SA是一种在生产调度，工程设计，机器学习，数据处理等领域应用极为广泛的通用启发式全局优化算法。

在固体的退火过程中，高温时，固体内部的分子运动相对活跃和无序；随着温度逐渐冷却下降，能量减小，分子运动也逐渐减弱趋于有序和稳定。模拟退火算法利用这一特性，将寻优目标函数与固体的能量E作了类比，用温度T来模拟控制参数，在降温过程中，通过“产生新解-得到目标函数差值-按概率接受新解”的逻辑循环迭代，最终在完成降温时趋向于全局最优解。

该算法执行中仅仅由温度的冷却过程来控制，针对任意目标函数的优化问题，它只需要设置好初始温度T，同一温度下的迭代次数i，降温的速率以及停止条件，就能最终求得最优解。算法的基本思想和执行过程可以概括如下：

1. 算法初始化：设置初始温度T，初始解S，每个温度下迭代次数i；
2. 产生一个新的解S1，并计算目标函数的差 ；
3. 若，则接受S1作为新解；否则按一定概率（该概率常取为exp(-ΔT/T)）接受S1作为新解；
4. 按一定速率降低温度，重复执行2到4步直到满足停止条件得到最优解。

已经证明，模拟退火算法寻优与初始值和初始解的设定无关。理论上，其最终会以概率1渐进收敛于全局最优解。

在本文第四章，正是基于模拟退火算法，并且结合本文研究对象的特点设计了针对性的接受概率和降温过程，对注视点估计算法中的核带宽参数进行寻优，使之能自适应调节取值，不仅大大简化实验的操作过程，也有效地减小了注视点估计的误差。

## 参数选择和优化

为了使该参数能够在计算过程中取到合适的值，许多时候大家都会先人工给其设定一个值，再根据程序实际运行的情况，不断重复调整，多次跑程序去找到一个较为理想的配置。显然，这会增大实验的工作量，而且这样依靠试验比较和经验来设定的参数值也并不一定就是最佳的。在之前的计算机视觉相关研究中，曾经提出了核带宽取值的经验公式，或者其他人为调节的办法[19,20]。本文则是在程序中直接实现了基于优化算法的自适应参数选择。虽然将不可避免的令程序运行时间大幅增加，但减少了人工干预的大量精力和时间，并且得到的是同等条件下误差较小的估计结果。

### 参数选择方法

在图像处理领域，曾经提出过关于核带宽取值的经验公式，比较著名的例如Silvermans Rule[20]，用公式表示如下：

（4-1）

其中指核带宽，指样本的均值，R为经验指定的常数，n为样本数量。

但是，这样根据经验的公式，在不同的实际应用中显然难以保证准确和易用。尤其是在具体的MATLAB程序里难以灵活设置和调节。

佛罗里达大学(University of Florida)的Erion Hasanbelliu[16]在其论文中也解释了核带宽的作用和取值范围。另外，他还对不同参数取值的实际效果做了比较，并使用了设置步长的衰减以及指数衰减的方法，来得到一个在图像匹配算法中较为理想的参数值。

选择参数的目标是令误差相对更小。第二章已经介绍过，由于核带宽本身的特性，其取值越小，可观察到的平均误差也显然减小了，并且会带来收敛性的困难。因此在参数仅需要微小调节的较小范围内，使用误差的均值这一“剩余价值”显然不能衡量核带宽的优劣。本文在这种情况下，舍弃了追求继续降低无谓的误差均值，而是选择利用其标准差，追求增加估计的稳定性。

据此，本文在实现参数选择的优化程序中，也首先使用了设置初值和步长（等间距），或者指定衰减函数（一般采用指数函数）这类启发性的寻优方法，并且获得了预期的较好效果。该类方法总能够保证获得一个最佳的取值，效果如图4-2和图4-3所示。但这样的选择过程需要一定的运行时间，并人为指定初值以及一些过程参数，这显然是需要先验知识的。因此，这类方法具有较好的启发性，但推广性略显不足。

图 4-2 使用指定初值和步长的方法衰减获得最优核参数，纵轴的Std指标准差



图 4-3 指定初值使用指数函数衰减，获得最优核参数

在这方面，推广性更好的方法显然是使用模拟退火（Simulated Annealing，SA）算法[21] ，或者类似的遗传算法（Genetic Algorithm，GA）等具有全局性优化特点的优化方法。具体到本实验中，本文采取了模拟退火算法对核带宽参数进行自适应的寻优计算。理论上，该过程与注视点估计可以同时完成，不再需要在其运行前后的人工干预配置。另外，这类全局优化算法配置简洁，相比之前的衰减法更具推广性，无需相关知识指定步长和初始值之类的有关参数本身的信息。

如图4-4的模拟退火寻优结果图所示，在实验中，不论是通过衰减法还是模拟退火法，都能够自适应地得到类似的核带宽最优取值区间以及相应的最小标准差。该取值建议的准确性既能从之前提及的经验和公式上得到验证，同时也可以从误差的相应减小得到支持。具体地，使用模拟退火算法寻优得到的核带宽取值，相比实验者的指定，能将相关熵估计误差降低5%到12%。在图4-4中可见，本文特意取了两个不同的初始解空间和降温过程，但还是得到了几乎一模一样的结果，这验证了该方法的稳定性，说明这样的选择过程可以移植应用于许多类似的算法和实验中。



图4-4 用模拟退火法在不同范围内对核带宽取值寻优的结果

左图的寻优范围相对较小（初始解取1），右图寻优范围较大（初始解取10）。

当然，考虑到模拟退火算法运行需要较多的时间（比衰减更长），实际情况下也可以人为加以设置，分开完成两部分的工作，以提高效率。比如对温度和接受函数进行一定的定向调整，那么自动寻优的效率还能在此基础上大大增加。

总而言之，根据上述两类方法的程序，实验者可以从反馈的图形直观地看出实验中合理核带宽取值区域。当然在不同的条件，核参数最佳范围也会有相应波动，但大都在0.15~0.55区间。本文的结果基本吻合了之前的各类经验数据，也能够有效降低注视点估计实验的误差。

### 程序实现

上一节讲到，在核参数自适应调节选择的部分，本文先后使用了衰减法和模拟退火法两种自动寻优算法，前者结合了部分人工设置更具效率和启发性，后者则具备全局优化性能和更好的推广性。在MATLAB中，分别实现了这两个功能的函数，在注视点估计的主程序中直接调用，从而使参数选择在估计算法中完全自动进行，有效降低了算法的误差，也简化了实验流程。

不论采用上述哪种方式，参数调节的程序实现基本思路都是以注视点估计的误差标准差作为目标函数，对核带宽参数进行寻优求解。算法在估计过程中，用不同的核带宽取值带入公式（3-8）进行计算，根据误差的变化不断自适应地调整，最终找到能使误差最小的取值。

在寻优算法设计的过程中，主要面临的问题包括如何设置初始值和停止条件，搜索步长和速度，以及怎样克服局部极值。模拟退火算法的特性已经保证了其对以上问题的克服的通用性，这一点已经在本文第二章介绍过。但是若要采用衰减的方法，以上几个问题的解决还是需要经过人工的调节和设计，才能使函数相对快速地获得理想的结果。

具体地，在类似于“爬山法”的贪婪搜索过程中，实验者需要设置一定的“惯性”和“延迟”，来保证其停止点不会位于局部极值区域。本文采取的实现方式是通过设置固定的停机阈值和延迟，避免了算法在局部极值区域就跳出。而这一阈值的设定，需要在调试过程中人为设置。这也正是该方法的局限性所在：虽然在设定好算法的各个参数后其寻优速度快，效果明显，但缺少了推广性和通用性。

根据章节4.2的介绍，模拟退火算法的程序实现过程相对来说不再需要过多专门额外的设计和干预。实验者只需要控制好温度下降速率，选用合适的接受概率，使得降温过程能够合理还原寻优过程即可。在本实验中由于误差函数的差值相对都较小，故对传统的Metropolis接受概率进行了修正，首先“放大”误差，再使之在降温过程中合理地逐步减小概率，并且减少了同一温度下的迭代次数，适当加快降温速度增加效率。

如上节所述，不论采取哪种方式，都能得到合理的最优值以降低实验标准差。在具体操作中，根据效率和准确度的要求，可以灵活取舍设置该部分的程序和参数。本文认为，这类方法在涉及到调参的类似问题中都有着普遍的适用性，是把优化问题放在本文研究课题下的一次很好的实践。

## 实验结果及分析

本文优化了第三章所述程序代码，使得在运行注视点估计程序的同时，用上述方法进行核参数自适应的寻优。根据估计误差，得到了每组样本核带宽最佳取值的参考结果。标定数目为36点时，参数迭代的过程在图4-2，4-3和4-4中体现。

使用上述几种调节方法，均能使本实验估计当中的角度误差的标准差达到最小。此时的核参数取值可能会在一定的小范围内（约0.01）存在偏差。由于这里的带宽最佳取值常常会表现为一个平坦区间，而不是一个峰值点，所以这里可以认为所得最优取值的结果相差不大。下表展示了不同标定点数目下参数寻优的情况：

表4-1 注视点估计实验中核参数选择结果比较

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 选择方法 | 9点 | | 16点 | | 25点 | | 36点 | |
|  | 最优参数 | 目标函数 | 最优参数 | 目标函数 | 最优参数 | 目标函数 | 最优参数 | 目标函数 |
| 人工指定 | 0.5 | 0.0593 | 0.5 | 0.0592 | 0.5 | 0.0578 | 0.5 | 0.0997 |
| 步长衰减 | 0.1850 | 0.0453 | 0.1625 | 0.0551 | 0.2425 | 0.0477 | 0.5475 | 0.0929 |
| 指数衰减 | 0.1858 | 0.0453 | 0.1615 | 0.0546 | 0.2410 | 0.0464 | 0.5475 | 0.0929 |
| 模拟退火 | 0.1852 | 0.0453 | 0.1620 | 0.0546 | 0.2396 | 0.0464 | 0.5556 | 0.0929 |

由表4-1可见，在调整核带宽后得到的最佳的标准差比初始设定的情况减小了约10%。此外本文在实际实验中用了两类共三种不同的寻优算法，且针对模拟退火算法也取了不同的初始解和降温速率，都能得到相近的结果。所以参数选择的正确性是可以得到充分的交叉支持的。

本文提供的基于优化算法的核参数选择，主要价值是在减少误差的同时，免去了人工选择反复挑选的麻烦。考虑到如今计算能力和资源不再是限制，因此拿运行时间来换取更高的自适应性，在许多类似的课题中也有应用的必要性和潜力。但实验者是否做出“为一两个参数的优化而大大增加运行时间”的交换，还需要结合实际情况。

## 本章小结

本章介绍了核带宽的性质，并在此基础上提出了实验中核带宽参数的自动优化选择方法。该部分使用了多种优化方式在估计过程中对参数进行寻优，并分析它们各自的优势和不足。最后，总结了核参数优化选择对于注视点估计的意义和实际效果。

# 结论与展望

综上所述，本文主要提出了在桌面环境下，基于相关熵的一种表观特征注视点估计的方法，验证了其相较于典型传统方法的优越性。另外还实现了基于优化算法的核带宽参数自适应选择，借此提高精度的同时简化了实验操作，进一步改善了本文提出的估计方法的效率。

## 结论

在之前的研究的基础上，本文的工作成果主要可以概括为以下结论：

首先，实现了基于相关熵的桌面环境注视点估计。其原理是利用注视点空间和人眼特征空间的相似性，通过特征提取和共享权重，由训练图像样本和坐标来估计当前注视位置坐标。本文所提出的基于相关熵的桌面环境下注视点估计，较其他方法而言主要有两点优势：第一是利用相关熵作为损失函数的特性，比大部分方法采用MSE，或者采用最近邻的方法能够更好克服非高斯的噪声，其准确性更好；第二是相对来说需要的训练样本数不多，实现方案也简单。经验证，该方法的估计角度误差平均在0.2以下，要优于ALR等之前的方法。

其次，实现来基于优化算法对核带宽参数的自适应优化。本文分别实现了使用衰减法和模拟退火法的参数自适应选择。两者的优势分别在于启发性和推广性，但它们都能够有效地在估计过程中自动对核带宽进行稳定的全局寻优。实现核参数的自动选择，意味着注视点估计的实验流程得到了简化，改变了过去多次人为测试来指定参数的情况。此外，本文基于标准差的评价函数寻优，在估计中可以自动地将估计角度的标准差降低约11.5%。

两者结合起来看，本文在MATLAB下实现了完整的注视点估计及其参数选择程序，能够较准确地完成用户人眼提取处理，注视位置估计，以及参数寻优等全过程。经过对程序的调整改进，还可以根据实际情况灵活调节各类子函数及其参数，从而实现不同需求下的稳定和高效输出。在此基础上，将可以实现桌面程序，在有一定延迟和误差的情况下，基本做到对用户的注视位置进行估计反馈。

## 展望

当前，受益于机器学习的发展，各类基于表观特征的注视点估计方法研究热度不减，国内外的研究提出了许多好的方法和改进。得益于研究领域的整体成果，本文也在此基础上，在该领域进行了一定有益的探索与尝试。但要认识到目前的方法都还普遍存在局限性和不足，尚需要为将来的研究继续做好铺垫，找好方向和动力。

如上一节所述，本文的方法引入了相关熵作为误差函数，充分利用其特点，在训练样本不多的情况下有效地克服了噪声影响，减小了估计误差，并且实现了核参数的自适应调节。但不论是这种实现方案，还是其他的方法，依然存在许多明显的不足。比如，虽然各类方法大都经过了比较繁琐的标定和训练过程，目前却仍然难以较好地处理用户头部姿势和眼睛区域变化较大的情况。更不要说户外，近视，复杂光照等常见情况尚需克服。这些限制，无疑对于使用者和使用场景提出了相当高的要求，显然在实际应用中将是一个很大的困难。

另外，从对图片的处理到权值的计算，都会造成估计的延迟，难以保证实时性，这就给注视点估计的交互型应用造成了很大的不便，在眼动控制和行为分析等场景下，既没有条件先标定再估计，也很难接受现有这种延迟的存在。

还需要再次指出的是，本文选择参数的过程，虽然通过基于优化的方法实现了自适应的调节，但都需要较长的运行时间，尤其是采用模拟退火法所需运行时间较长，还需具体分析是否要用运行时间来换取实验操作的方便。

因此，将来的工作主要应包括对算法稳定性和速度的提升，令使用条件更具有包容性。除此之外，本实验在代码结构上应当更加规范化，模块化。只有在此基础上继续加以努力，推动其走向成熟和普适，才能够为该技术的广泛应用打下坚实的基础。

当然，每一次的探索和付出，都为注视点估计技术在实际生活中的推广使用做出了贡献。相信在不久的将来，这项科技的成熟会带给我们一种更加智能和便捷的生活方式，广泛而深入地造福社会。

# 致 谢

四年的时间似乎很长，等到最后回头，才发现一切其实都过得非常快。随着毕业设计工作进入尾声，我在交大本科阶段的学习也即将画上句号。此时，心里除了种种回忆带来的百感交集，更多的是对一路上帮助我的人最真诚的感谢与敬意。在这篇论文的结束之处，我想要表达对他们发自内心的感恩。

首先，我想特别感谢我的导师张雪涛老师。张老师不仅是我的指导老师，也是我最后一年的班主任。整个大四学年，除了学习研究上的耐心指导，张老师还在我转班，实习，以及出国申请过程中提供了许多重要帮助，一路陪伴和激励我前行。从零开始的毕设工作中，张老师每周都和我详细地讨论，督促我的同时也为我指明了方向，培养锻炼了我的能力。一起相处的时间里，张老师认真负责，亲切随和的态度让我十分尊敬。无论今后走到哪里，我都会对张老师心存真挚的感恩与敬意。

另外，感谢杨奔学长，在毕设过程中热心提供了许多关键的指导，在我遇到困难时帮助我，也教会了我如何面对并解决问题。在实验项目的数据采集和图像处理部分，杨奔学长也付出颇多，是我的学习榜样。在此祝你学有所成。

感谢我的舍友孙崇尧、唐昊，以及其他许多关心我的同学和朋友。大家朝夕相处，共同奋斗也互相鼓励，给毕设的这段时光留下许多有趣的回忆。分别之际，祝各位前程似锦。

感谢西安交大和我的全体老师们，在这里我收获了知识与技能，严谨的习惯，奋斗的精神，当然还有一段难忘的大学岁月。这些伴随一生的成长源自于各位交大人的奉献，我也会感恩并祝福母校。

最后，要感谢我的家人，一直在背后默默关心支持我，让我时刻感到爱与温暖。

这一路走来，需要感谢的人还有很多，实难一一列举。总之，我定会铭记于心，继续努力，不辜负大家的帮助与支持。再次感谢张老师以及关心我的各位！

# 参考文献

[1] Strandvall T. Eye Tracking in Human-Computer Interaction and Usability Research[M]// Human-Computer Interaction – INTERACT 2009. Springer Berlin Heidelberg, 2009:35-52.

[2] Holzman PS, Proctor LR, Levy DL, et al. Eye-tracking dysfunctions in schizophrenic patients and their relatives[J]. Archives of general psychiatry, 1974, 31(2): 143-151.

[3] Borji A, Itti L. State-of-the-Art in Visual Attention Modeling[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(1):185-207.

[4] Yoo DH, Chung MJ. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion[J]. Computer Vision and Image Understanding, 2005, 98(1): 25-51.

[5] Zhu Z, Ji Q. Eye gaze tracking under natural head movements[C]//Computer Vision and Pattern Recognition, IEEE Computer Society Conference on. IEEE, 2005, 1: 918-923.

[6] Hennessey C, Noureddin B, Lawrence P. A single camera eye-gaze tracking system with free head motion[C]// Eye Tracking Research & Application Symposium, Etra 2006, San Diego, California,Usa, March. DBLP, 2006:87-94.

[7] Williams O, Blake A, Cipolla R. Sparse and Semi-supervised Visual Mapping with the S^3GP[C]//Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on.IEEE, 2006, 1: 230-237.

[8] Martinez F, Carbone A, Pissaloux E. Gaze estimation using local features and non-linear regression[C]// IEEE International Conference on Image Processing. IEEE, 2013:1961-1964.

[9] Baluja S, Pomerleau D. Non-Intrusive Gaze Tracking Using Artificial Neural Networks.[C]//Advances in Neural Information Processing Systems. 1994:753-760.

[10] Tan KH, Kriegman DJ, Ahuja N. Appearance-based Eye Gaze Estimation[C]// IEEE Workshop on Applications of Computer Vision. IEEE Computer Society, 2002:191.

[11] Sugano Y, Matsushita Y, Sato Y, et al. Appearance-based gaze estimation with online calibration from mouse operations[J]. IEEE Transactions on Human-Machine Systems, 2015, 45(6): 750-760.

[12] Lu F, Sugano Y, Okabe T, et al. Adaptive linear regression for appearance-based gaze estimation[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(10): 2033-2046.

[13] 李中常．基于表观特征的人眼注视点估计算法研究[D]．西安交通大学，2017．

[14] Ma S, Du T. Improved Adaboost Face Detection[C]// International Conference on Measuring Technology and Mechatronics Automation. IEEE, 2010:434-437.

[15] Highleyman WH. Linear decision functions, with application to pattern recognition[J]. Proceedings of the IRE, 1962, 50(6): 1501-1514.

[16] Erdogmus D. Information theoretic learning: Renyi’s entropy and its applications to adaptive system training[D]. University of Florida, 2002.

[17] Stephan R. Sain. The Nature of Statistical Learning Theory[J]. Technometrics, 1996, 8(4):1564.

[18] Liu W, Pokharel PP, Príncipe JC. Correntropy: Properties and applications in non-Gaussian signal processing[J]. IEEE Transactions on Signal Processing, 2007, 55(11): 5286-5298.

[19] Erion H. Information theoretic similarity measures for shape matching[D]. University of Florida, 2012.

[20] Badong Chen, Lei Xing, Xing Wang, et al. Robust learning with kernel mean p-power error loss[C]// IEEE Transactions on Cybernetics. IEEE, 2016, PP(99):1-13.

[21] 严刚峰，赵宪生．基于模拟退火-遗传算法的控制控制参数寻优研究[J]．四川大学学报，2003，40（5）：874-877．