

# 北京交通大学

## 硕士专业学位论文

智能沙盘中的自动配置与光斑控制研究

Research on automatic configuration and spot control in intelligent  
sand table

作者：赵映南

导师：李纯喜

北京交通大学

2021年6月

## 学位论文版权使用授权书

本学位论文作者完全了解北京交通大学有关保留、使用学位论文的规定。特授权北京交通大学可以将学位论文的全部或部分内容编入有关数据库进行检索，提供阅览服务，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。同意学校向国家有关部门或机构送交论文的复印件和磁盘。学校可以为存在馆际合作关系的兄弟高校用户提供文献传递服务和交换服务。

（保密的学位论文在解密后适用本授权说明）

学位论文作者签名：赵映南

导师签名：李四

签字日期：2021年6月1日

签字日期：2021年6月1日

学校代码：10004

密级：公开

# 北京交通大学

## 硕士专业学位论文

智能沙盘中的自动配置与光斑控制研究

Research on automatic configuration and spot control in intelligent sand table

作者姓名：赵映南

学 号：19125078

导师姓名：李纯喜

职 称：副教授

工程硕士专业领域：信息系统

学位级别：硕士

北京交通大学

2021年6月

## 致谢

本论文是在我的导师李纯喜副教授的悉心指导下完成的。在本文的完成过程中，从研究方向的确立到论文的选题再到论文的最终完成，李老师给予了我极大的帮助和指导。从确立论文思想，到疏通论文结构，再到修改论文细节，李老师始终耐心细致、不遗余力地指导着我，帮助我不断完善毕业论文。李老师严谨的教学态度、渊博的知识、和善的待人态度不仅是我学习的榜样，而且会在以后的工作生活中始终激励我不断进取。在此，我由衷地感谢李老师对我在学习、生活、研究等各方面无私的帮助。

同时我要感谢实验室的所有老师，感谢赵永祥老师、郑宏云老师、郭宇春老师和其他老师在我研究生期间对我的帮助和指导。老师们不仅为我指明了研究方向，还在论文完成过程中为我提出了许多宝贵的意见，帮助我解决了研究中遇到的困难。在此我向实验室的各位老师表达真挚的感谢。

在研究生期间和学位论文撰写期间，还要感谢实验室的高志朋师兄、张虎信师兄、张琛玥师姐、李小乐师姐、李文雯、张庆贺、穆彦龙等同学对我的支持和帮助，尤其是李文雯同学，在我遇到困难时，不遗余力的帮助我。感谢张雪等同学对我生活和学习的帮助，感谢你们陪我度过这段难忘的学习生涯。在每次遇到困难时，通过与他们的沟通给予了我很多建议和启发，帮助我走出困境。在此向他们表达我的感谢之意。

最后，特别感谢我的家人和其他亲人朋友，正是他们的热情鼓励和默默奉献，才使我顺利完成学业，成为社会的有用之才。

## 摘要

目前沙盘系统已广泛应用于各个领域。近年来出现的新型智能沙盘系统采用物理沙盘和电子沙盘相结合的方式进行信息的展示,物理沙盘是指人们按照一定比例用泥沙等材料制作的沙盘模型,主要通过激光笔进行简单的交互,电子沙盘是融合了多媒体技术和物理沙盘的新型沙盘系统,其布置过程虽然简单,但是显示信息量较少,也无法实现图形信息的立体化显示。但是,目前的智能沙盘系统还存在以下问题:(1)激光笔在沙盘或者屏幕指示时的目标光斑很小,当观众与物理沙盘以及屏幕之间的距离较远时,难以看到光斑。(2)电子屏幕的交互性很差,要么只能播放既定编排的内容,要么需要演示者使用激光笔/鼠标的操控设备控制屏幕显示内容,操控不方便。(3)物理沙盘和电子沙盘操控的操控功能没有兼容于一个设备,给演示带来不便。

本文基于现有实体沙盘设计了一个基于图像识别的智能沙盘系统,以解决物理沙盘和电子沙盘的操控问题;实现了基于红外激光笔的电子沙盘鼠标操控核心功能,解决了电子沙盘的灵活交互问题。论文讨论了所设计的系统的可行性,实验结果证实了实现的核心功能的可用性,为后续实现完整的智能沙盘系统奠定了基础。本文的主要工作如下:

(1) 设计了基于现有实体沙盘的智能沙盘系统,该系统包括屏幕自动配置模块、光斑识别模块、坐标转换与光标追踪模块,实现了屏幕的自动化配置并利用激光光斑实现了相对便捷的人机交互。

(2) 实现了一种基于图像识别的屏幕自动配置功能,以匹配电子显示屏角点坐标和摄像头坐标。该自动配置功能不严格约束摄像头与屏幕的空间摆放位置,而是通过分析摄像头图片与屏幕显示图片的特征点对应关系自动校准摄像头对应的屏幕区域。

(3) 实现了基于红外激光笔的虚拟鼠标操控功能,以实现灵活方便的大屏幕远程操控。该操控功能技术流程包括:红外光斑提取、光斑的坐标转换(从摄像头坐标到屏幕坐标)、虚拟鼠标操控。

**关键词:** 智能沙盘系统; 屏幕自动化配置; 光斑识别; 坐标转换

## ABSTRACT

At present, sand table system has been widely used in various fields. In recent years, the new intelligent sand table system uses the combination of physical sand table and electronic sand table to display information. Physical sand table refers to the sand table model made of sand and other materials according to a certain proportion. It mainly uses laser pen for simple interaction. Electronic sand table is a new sand table system that integrates multimedia technology and physical sand table, Although the layout process is simple, the amount of display information is less, and the three-dimensional display of graphic information can not be realized. However, the current intelligent sand table system still has the following problems: (1) the target spot of the laser pen on the sand table or screen is very small, when the distance between the audience and the physical sand table or screen is far, it is difficult to see the spot.(2) The interactivity of the electronic screen is very poor, either it can only play the scheduled content, or it needs the demonstrator to use the laser pen / mouse control equipment to control the screen display content, which is inconvenient to control.(3) The control functions of physical sand table and electronic sand table are not compatible with one device, which brings inconvenience to the demonstration.

Based on the existing solid sand table, an intelligent sand table system based on image recognition is designed to solve the problem of physical sand table and electronic sand table; The core function of electronic Sandtray mouse control based on infrared laser pen is realized, and the flexible interaction problem of electronic sand table is solved. The feasibility of the system is discussed in this paper. The experimental results confirm the availability of the core functions, which lays the foundation for the realization of the complete intelligent sand tray system. The main work of this paper is as follows:

(1) A set of intelligent sand table system based on the combination of entity sand table and large screen display is designed. The system includes screen automatic configuration module, spot recognition module, coordinate conversion and cursor tracking module. The automatic configuration of screen is realized, and the relatively convenient human-computer interaction is realized by using laser spot.

(2) A screen automatic configuration function based on image recognition is realized to match the coordinates of the corner points and cameras of the electronic display. The automatic configuration function does not strictly restrict the space placement position of camera and screen, but automatically calibrates the screen area corresponding to camera

by analyzing the corresponding relationship between camera picture and feature point of screen display picture.

(3) The virtual mouse control function based on infrared laser pen is realized to realize the flexible and convenient remote control of large screen. The technical flow of the control function includes: infrared spot extraction, coordinate conversion of spot (from camera coordinates to screen coordinates), virtual mouse control.

**KEYWORDS:** Intelligent sand table system; Screen automation configuration; Spot recognition; coordinate transformation

## 目录

摘要 .....	iii
ABSTRACT.....	iv
1 引言 .....	1
1.1 研究背景与意义 .....	1
1.2 行业发展现状 .....	2
1.3 本文研究内容 .....	3
1.4 论文结构安排 .....	4
2 相关研究 .....	5
2.1 基于交互的屏幕演示系统 .....	5
2.2 人机交互 .....	7
2.2.1 概述 .....	7
2.2.2 激光笔现有产品 .....	7
2.2.3 激光光斑识别 .....	8
2.3 视频采集技术 .....	9
2.3.1 视频采集卡 .....	9
2.3.2 数字视频软件开发包 .....	9
2.4 屏幕校准 .....	11
2.4.1 概述 .....	11
2.4.2 特征点提取 .....	11
2.4.3 特征点匹配 .....	13
2.4.4 误匹配点的剔除 .....	14
2.5 本章小结 .....	16
3 系统总体设计 .....	18
3.1 研究背景 .....	18
3.2 需求分析 .....	19
3.3 系统设计 .....	20
3.3.1 总体思路 .....	20
3.3.2 硬件架构 .....	22
3.3.3 软件架构 .....	26
3.4 本章小结 .....	27
4 软件实现 .....	28
4.1 概述 .....	28
4.2 屏幕校准 .....	30
4.2.1 概述 .....	30
4.2.2 特征点提取 .....	31
4.2.3 特征点匹配 .....	33
4.2.4 错误匹配剔除 .....	34
4.3 光斑识别 .....	37
4.3.1 视频采集 .....	37
4.3.2 激光光斑定位 .....	37

4.4	坐标转换与鼠标追踪 .....	39
4.4.1	坐标转换 .....	39
4.4.2	鼠标追踪 .....	40
4.5	本章小结 .....	41
5	系统评估 .....	42
5.1	评估方法 .....	42
5.1.1	软硬件环境 .....	42
5.1.2	评估标准 .....	42
5.2	评估结果分析 .....	43
5.3	本章小结 .....	45
6	结论与展望 .....	46
6.1	本文总结 .....	46
6.2	工作展望 .....	46
	参考文献 .....	47
	作者简历及攻读硕士/博士学位期间取得的研究成果 .....	50
	独创性声明 .....	51
	学位论文数据集 .....	52

# 1 引言

## 1.1 研究背景与意义

智能沙盘系统在人们日常生活中的应用领域非常广泛，比如军事、旅游、教育以及房地产等不可或缺各个方面<sup>[1]</sup>。沙盘是以一定比例及方式体现了实体以及实体某些特征要素的模型。随着时代的进步和经济社会的发展，人们的生活需求变得越来越高，对于智能沙盘系统的要求也在不断增加，沙盘的制作工艺和表现形式也在不断向更加智能化的方向进步。目前，智能沙盘系统逐步成为了人们学习、工作和宣传的一种高效且方便的工具。

智能沙盘系统最初的表现形式为实体沙盘。实体沙盘实际上就是人们按照一定的参考用泥沙等常见材料制作而成的模型<sup>[2]</sup>，该模型是对地形图、航空图片或者实地的地形进行一定比例的缩放。由于沙盘能够生动形象的模拟地形并为用户进行交流互动提供便捷条件<sup>[3]</sup>，因此传统的实体沙盘成为了从古代开始人们进行决策演示的重要手段和有力工具<sup>[4]</sup>。因其广泛的应用范围，沙盘成为了人们生活中离不开的重要工具。但是传统实体沙盘也存在缺点，在对实际的应用场景进行精细且逼真的还原过程十分复杂繁琐，而且会消耗大量的时间。还有不可忽视的一点是实体沙盘的携带运输十分不方便，这些缺点进一步限制了实体沙盘的应用范围。

随着时代的进步，计算机技术进入了千家万户，实体沙盘中加入声、光、电等元素变成了电子沙盘，继续在方方面面影响着人们的生活方式<sup>[5]</sup>。沙盘的应用范围逐步扩大到了各行各业，主要的应用方向分为宏观和微观两个方向。宏观方向主要用于旅游业、房地产业以及城市规划等<sup>[6]</sup>，比如某些城市或者区域的电子沙盘能够清晰细致并且立体的展示山水风景，并基于这种立体化的进行一些类似水电管道的规划。微观方向主要应用于某一具体区域的平面显示，比如应用于一栋建筑物发生火灾或其他突发事件时的辅助应急指挥<sup>[7]</sup>。虽然电子沙盘的显示效果得到了大大的提升，但还是难以满足当代社会信息化的需求，对于各种应用场景的适应能力也越来越低。因此，对于智能沙盘的研究有着重要的实际意义和应用价值<sup>[8]</sup>。

随着人机交互技术的兴起，电子沙盘也在朝着更加智能化的方向发展。人机交互技术起初是通过鼠标以及键盘来实现，这些传统设备在空间上限制了演示人员的活动范围。这种限制使得演示者与听众们之间缺乏交互拓展的灵动性、有效性以及多样性<sup>[9]</sup>。屏幕是数据信息可视化主要载体之一，大型的屏幕所能够传递的信息量不可估计，这种形式在带来极大便利的同时也存在着许多弊端，第一难以准确的对演者的真实意图进行准确高效的传达，尤其是在需要多人共同研讨的场景<sup>[10]</sup>；

第二是只能以二维的形式实现信息图形的展示，无法为听众带来多角度的视觉体验。

综上，设计并实现一个基于人机交互技术的智能沙盘演示系统，协助演示人员高质高效的完成信息的表达传递，为军事指挥、协同办公、应急决策、城市规划等各个方面提供帮助，具有深远的社会意义和应用价值，同时也是对人机交互技术等现代科技在沙盘演示领域这种传统应用上的一种探索，进一步实现沙盘演示系统的应用价值。

本文的主要研究工作是设计并实现一个将屏幕演示与实体沙盘联动的演示系统，将激光笔作为交互设备，通过对视频采集技术、激光光斑识别与追踪技术、屏幕矫正技术等各方面技术的研究，实现简单、自然、便捷的人机交互行为。

## 1.2 行业发展现状

人机交互系统自发展以来，一直以机器作为人机交互的中心，用户作为操作者需要对自己的演示行为做出调整以适应机器，然而在新一代的人机交互系统中，用户将代替机器成为人机交互系统的中心。用户的行为、语言、肢体动作，才是演示行为中最常见的交流方式。国内外的研究者致力于尝试通过不同的途径使计算机感知到人的行为、语言、肢体动作等，从而实现用户与计算机之间的交互。常见的交互方式有数据手套、运动传感器<sup>[11]</sup>或图像识别手势<sup>[12]</sup>等几种方式。

国外学者对于自然人机交互方面的研究起步较早。2000年，Visual Panel 系统被 Ying Wu 等人设计出来<sup>[13]</sup>，这个系统主要是利用摄像头进行信息的获取，并通过计算机对摄像头获取到的信息进行实时的指尖检测，当指尖同时处于同一位置进行了短暂的停留作为点击事件的标志。2008年 Nakamura 等人对投影环境下的使用的笔触式交互系统进行了开发。该系统需要用户穿戴上配套的 LED 灯，系统通过对 LED 灯的检测来进行位置和动作信息的定位，从而实现了用户与投影屏幕之间的交互。2009年 Mistry 等人发明了 SixthSense，这种可穿戴的移动设备通过对不同的颜色进行检测标记来确定动作进行人机交互<sup>[14]</sup>。也是在这一年 Kinect 摄像头被微软公司发布<sup>[15]</sup>，这种摄像头能够实现深度信息的检测，并利用散斑信息实现三维定位，对体感的游戏人物进行控制。2010年小型交互设备 Light Touch 由 Light Blue Optics 公司首次发布，实现了更为方便的人机交互。

国内对于自然人机交互的研究起步稍晚。各大高校也开始对于自然人机交互技术的研究。北京航空航天大学重点对混合现实中的人机交互技术进行了研究。北京理工大学通过摄像头、穿戴头盔以及一些传感器对人机交互系统进行了搭建。清华大学也有专门研究人机交互理论的实验室。2008年，浙江大学提出了一种基于

视觉的实时手势跟踪识别的人机交互系统<sup>[16]</sup>。在手势分割的阶段提出了肤色提取并融合运动信息的方法，在手势跟踪算法中提出了 CAMSHIFT 算法与 Kalman 滤波算法相结合确定搜索窗口以获得手的位置，最后在手势识别阶段利用基于 Hausdorff 距离的模板匹配方法实现了两种是手势识别。2010 年，一种基于模型和表观相结合的手势交互系统通过判断手势轮廓曲率实现了手指之间的实时检测<sup>[17]</sup>。2011 年利用触屏技术实现的交互式电子沙盘将静态模型与触摸屏交互功能进行结合，形成了一种可视化，可交互训练场景<sup>[18]</sup>。2013 年长安大学对基于红外摄像机和投影仪的人机交互系统进行了研究实现<sup>[19]</sup>。2016 年，张玉军实现了基于手势交互的三维电子沙盘<sup>[20]</sup>。2017 年哈尔滨工业大学实现了人机交互示教系统，该系统主要基于计算机视觉实现<sup>[21]</sup>。通过空间中多个标志物的位置对示教器的姿态进行了计算，同时也对单手跟踪检测和姿势进行了计算。2019 年，长春市一家科技公司实现了将基于视觉的人机交互技术应用到了机器人的控制之中。

上述人机交互技术普遍存在缺点，利用手势或指尖识别对屏幕实现人机交互，将用户的活动范围限制在了屏幕附近，并没有完全打破传统鼠标对于用户活动范围的制约，其他利用计算机视觉实现的人机交互技术无法做到实体沙盘与屏幕演示系统之间的联动。

### 1.3 本文研究内容

通过 1.2 节的论述，智能沙盘和人机交互技术都在不断的发展中，但是这两个应用领域还是存在着许多不足。随着智能沙盘系统的应用场景不断增多，不同的应用场景的需求也不相同，对智能沙盘系统的开发设计造成了很大困难。目前的智能沙盘系统存在的主要问题有：

(1) 当用户与物理沙盘以及屏幕之间的距离较远，并使用激光笔在沙盘或者屏幕上指示时，会存在目标光斑很小，人眼难以看到光斑具体位置的问题。

(2) 目前系统中大屏幕部分显示信息需要提前进行设置，同时播放的视频只能按照预设时间线进行播放，无法根据现场实际情况做出调整，也无法根据演示者的意图实现交互功能。

(3) 虽然一些屏幕显示系统能够实现部分交互功能，但是进行交互行为的操作十分不便，演示人员需要同时手持教鞭和鼠标才能同时实现对物理沙盘和大屏幕的交互行为。

为解决以上问题，本文基于现有实体沙盘设计了一个基于图像识别的智能沙盘系统，以解决物理沙盘和电子沙盘的操控问题；实现了基于红外激光笔的电子沙盘鼠标操控核心功能，解决了电子沙盘的灵活交互问题。论文讨论了所设计的系统

的可行性,实验结果证实了实现的核心功能的可用性,为后续实现完整的智能沙盘系统奠定了基础。本文的主要工作如下:

(1) 设计了基于现有实体沙盘的智能沙盘系统,该系统包括屏幕自动配置模块、光斑识别模块、坐标转换与光标追踪模块,实现了屏幕的自动化配置并利用激光光斑实现了相对便捷的人机交互。

(2) 实现了一种基于图像识别的屏幕自动配置功能,以匹配电子显示屏角点坐标和摄像头坐标。该自动配置功能不严格约束摄像头与屏幕的空间摆放位置,而是通过分析摄像头图片与屏幕显示图片的特征点对应关系自动校准摄像头对应的屏幕区域。

(3) 实现了基于红外激光笔的虚拟鼠标操控功能,以实现灵活方便的大屏幕远程操控。该操控功能技术流程包括:红外光斑提取、光斑的坐标转换(从摄像头坐标到屏幕坐标)、虚拟鼠标操控。

## 1.4 论文结构安排

本论文主要通过六个章节进行论述,各个章节的主要内容为:

第一章为本论文的研究背景与研究意义,并对该行业的发展现状进行了充分调研,结合研究现状与主要应用背景确定了本论文的主要工作内容。

第二章对屏幕演示系统的相关技术进行了介绍,对该智能沙盘研究实现过程中涉及到的具体技术进行了简要介绍。

第三章是整个系统的具体设计,通过对背景介绍和需求分析,进行了智能沙盘系统的设计,首先对智能沙盘系统的硬件进行了型号选择,根据硬件选型带来的问题和系统需求进行了系统软件部分设计。

第四章对整个系统进行了软件部分的实现,对软件设计部分根据所需实现功能进行了模块化的划分,并介绍了具体实现步骤。

第五章为系统的评估部分。在准确率和时延两方面评估了系统的工作情况。

第六章为结论部分。对本文的工作内容进行了总结以及对未来工作的展望。

## 2 相关研究

基于交互的屏幕演示系统主要涉及自然人机交互技术。本文对自然人机交互技术进行研究，选择激光笔作为交互设备设计并建立了一个基于激光光斑的智能沙盘演示系统。本章将对激光笔交互技术、光斑跟踪技术以及屏幕校准技术做一个大致的介绍。

### 2.1 基于交互的屏幕演示系统

目前，信息技术正在逐步走向智能化，但是基于交互的屏幕演示系统并没有跟上智能化的脚步。为了实现人机交互、提升演讲效果、提高听众的参与度。本文对大屏幕演示系统进行了研究。

当前社会的各个领域，不论是教育、科研还是企业办公，为了追求工作的高效性和沟通的高质量，用户需要进行便捷的信息共享都离不开协同办公环境的支持。信息的含义是构成一定含义的一组数据，信息的概念处于不断地变化和发展之中，其内涵也在不断地扩展和外延，渗透到了人类社会和科学技术的众多领域，并且与材料、能源一起被列为现代社会和科技发展的三大支柱。在逻辑上，信息的载体是数值、文字、语音或者图形图像。在日常的办公生活中，从简单的文本信息到流畅的视频信息，用户渴望从大量复杂的信息中获取高质量的信息，从而在激烈的竞争环境或者紧急环境中协助用户做出正确的分析和判断，所以在诸如医疗、教育、气象、城市规划等领域都需要进行海量信息的演示，这些场景下的交互都需要一个性能较高的演示系统。

信息交互的双向性是一个十分重要的因素。为了实现双向的信息交互，目前大部分屏幕演示系统的人机交互主要是通过一些传统的设备实现，比如，鼠标、键盘等。当在大型的演示环境下进行公开的信息交流，传统的交互设备将演示者的活动范围限制在计算机设备附近。后来，蓝牙技术带来了无线鼠标以及无线键盘，这些无线设备的出现稍微缓解了有线设备对演示者活动范围的限制。但是，无线的交互设备还是需要有一个固定的操作台，此外，复杂的无线环境对其准确率存在较大影响，这样一来，演示者很难专注的面向听众进行信息输出，听众对演示者的表达意图也得不到清晰的把握，这种情况的出现违背了演示系统的设计初衷。此外，传统的交互设备无法对演示屏幕进行点击等基本操作，缺乏了交互的灵活性。为了适应千变万化的信息时代，更好的满足用户需求，设计交互方式更为灵动多样的演示系统具有实际意义和应用价值。

目前市面上的大屏幕演示系统按照安装的类型来分,可以分为壁挂式、滑道式以及壁橱式。根据厂商之间的技术手段不同,各种交互式大屏幕显示产品也存在不同优势,这些优势都离不开的关键技术是对笔的定位,也就是当演示者在显示图形的屏幕上写字、绘图时,获取笔的相对位置信息。但是基于电阻膜、电容以及表面声波的交互式屏幕演示系统虽然能满足显示系统所要求的显示精度,但是当大屏幕的尺寸不断增大时,各方面的成本都会随之提高,成本问题很大程度上影响了屏幕演示系统的应用范围。

基于遥控交互设备的多媒体演示系统一般由三个部分组成,分别是:计算机、多媒体演示大屏和交互设备。选用不同的交互媒介需要使用不同的技术,通过设计专门的应用程序,交互设备能够代替鼠标在多媒体演示平台上进行操作,比如,点击运行应用程序,或选中、编辑、保存文件等操作。此外,可交互的多媒体演示系统还应具有醒目的提示功能,方便听众可以跟上演示者的思路。总的来说,基于遥控交互设备的演示系统需要满足以下要求:

### (1) 灵活交互

演示系统需要实现演示大屏与计算机之间的灵活交互。主要利用一些交互设备及设计自动化程序来实现实体鼠标可完成的功能,比如,单击、双击以及对文件的拖拽等。该功能可以帮助演示者脱离传统的鼠标进行演说,在可以近距离与听众沟通互动的同时不失去对电脑的控制,从而最大限度地提高使用演讲者的演讲效果。

### (2) 用户友好

随着用户的不断增加,用户对于各种系统的使用水平参差不齐。演示系统的使用应该简单便捷,其中控制功能与操作方法的展现应清晰明了、易读易懂;系统的操作简单方便、直接有效;在操作过程中,要求用户做出的判断或要求用户记忆的东西应尽量减少。演示系统应利用系统的自动与智能来最大限度地使用户感到轻松。

### (3) 可拓展性

该功能主要用于描述演示系统与操作系统之间的交互,主要利用一些交互工具及自动化程序设计来完成鼠标的功能,比如单击、双击以及对文件的拖拽等。当对系统新增功能时,对现有系统功能影响较少,即不需要对现有功能做任何改动或者改动很少。

## 2.2 人机交互

### 2.2.1 概述

人机交互 (Human-Computer Interaction; HCI) 简单来说是指人与计算机之间的交互<sup>[22]</sup>。人机交互的相关研究主要是交互式系统的设计, 交互系统的实现与设计主要根据用户的需求进行<sup>[23]</sup>。其中, 用户、系统和系统的实现功能是人机交互关注的三个焦点问题。人机交互技术从出现发展至今, 已发展成一个重要的工程研究领域<sup>[24]</sup>, 引发了国内外的研究学者对其进行科学研究的兴趣。通过对人机交互技术的研究和相关系统的开发实现对产品设计、服务设计、城市规划等诸多实用领域带来了积极的影响, 各种便捷交互手段的实现极大的提高并改善了用户体验。

二十世纪七八十年代, 开始出现人机交互的概念<sup>[25]</sup>。人机交互的研究重心逐渐从机器层面转移到用户层面。从起初的鼠标键盘等传统的交互方式逐渐发展到通过肢体、手势、语音等自然方式交互。交互界面也从难以理解的专业文字向易于理解的图形交互界面过度。除了生活中常见的计算机和投影仪, 按键激光笔、电子白板以及互动投影系统等相关产品都是人机交互性不断增强的体现。越来越多的学者对通过激光笔实现人机交互的交互方式有广泛研究, 主要是因为激光笔成本低, 且激光笔发射光斑具有亮度高、发散角小等突出优势<sup>[26]</sup>。

### 2.2.2 激光笔现有产品

在屏幕演示系统中使用激光笔作为大屏幕演示系统的辅助设备交互, 让演示者在演示时能够实现与听众更好的互动, 最大限度地发挥眼神接触和肢体语言的作用, 持续吸引听众的注意力, 提高演讲效果, 使演示更加生动、完美, 彻底的解决了以往演示场合需要使用鼠标的不便。在本文中, 我们对激光笔产品进行了深入研究, 并对一些产品进行简单的介绍。

#### (1) 具有激光笔功能的笔式手机

近年来, 移动通信技术不断发展, 手机基本上属于人手一台的设备。但是对于一些特定职业的人群, 比如商场里的销售人员等, 他们需要有一种同时具备通信又具有类似激光笔功能的设备。这种将激光笔与手机相结合的设备具有很大的优越性和实用性。但是该设备只能在实体沙盘上起到指示的作用, 而无法对屏幕演示系统进行点击等操作。

#### (2) 语音激光笔

这项专利设计的激光笔首先对演示者的声音进行收集, 然后将演示者的声音传输至计算机, 计算机通过处理将声音信息转化为控制指令, 使这种设备在具有普

通激光笔指示功能的基础上,也能够通过语音信息进行交互,还具备麦克风的扩音功能,但是这种设备也存在缺点,第一我们可以通过更便捷的方式实现语音控制的控制指令,第二嘈杂的环境会对语音识别产生严重的影响,使控制指令得不到准确的响应。最重要的一点是目前的手持激光设备并没有对语音数据进行编码的能力,使需要传输的数据量十分庞大。

### (3) 光点鼠标激光笔

这项专利主要用于教学场景,为演示者提供了远程教学的功能,作为鼠标使用时,设备的底部具有类似鼠标的光电移动定位模块,能够为用户提供精确的鼠标操作。这种设备的不光使激光笔具有教学指示的功能,又可以当作鼠标或者手写笔使用,在教学时为来回切换教学用具省去了很多麻烦。这种设备存在的缺点是无法实现点击实物时的醒目化提示功能。

### (4) 基于轴陀螺仪的多功能无线鼠标笔

该产品使用无线传输技术作为电脑的无线外部设备,利用 6 轴陀螺仪实现了体感信息的采集,并利用自主设计的算法处理数据后实现了光标移动控制从而对电脑进行操控。其应用场景非常广泛,主要包括办公以及演讲等场合。这类产品保留了传统笔的功能,在此基础上还具备无线鼠标的功能。不仅能够简单实现在屏幕上的标记,也能通过配套软件实现在机房的电脑群中切换至内网远程端使用,更加方便了老师进行课程的指导与讲解。但是这种演示器受到其原理的限制,其光标的工作距离只有 30 米,无线环境复杂的背景下,实际的工作距离只有 10 米左右。而且体感存在一定的误差,准确度会受到一定的影响。这种设备同样具有无法对实物进行醒目化提示等功能。

综合分析上述产品优缺点,目前现有激光笔相关产品并不能做到使听众与演示者之间的信息交互流畅自然,本文对基于遥控的演示系统进行了深入研究,充分利用激光光斑的优点,与屏幕等演示信息载体相结合更为方便的实现信息的双向交互。

## 2.2.3 激光光斑识别

利用激光笔作为人机交互设备,需要利用计算机视觉技术对光斑进行识别。激光光斑识别属于目标检测技术,主要基于图像处理实现光斑的定位。激光笔产生的光斑是一种小目标,对于类似的小目标检测技术,目前已有的相关研究包括:光斑提取、特定形状识别。这些技术以亮度为基础,准确度较高。激光笔的应用场景的不同会导致光斑具体形状和光斑识别效率之间存在区别。现有的解决方案列举如下:

### (1) 小目标尺度自适应检测

这是计算机视觉领域的研究热点之一，使用的基本方法是正负 LOG 算子对连续帧图像的连续检测，从而确定真实目标亮点的位置，这种方式对光斑目标大小对检测结果产生的影响进行了充分的考虑，在真实性、准确性以及抗干扰能力等方面都有较好的表现。这种检测方法的思想非常值得借鉴。

### (2) 基于形状识别的检测

这类方法对于光斑的识别主要基于光斑的形状进行，这类方法对光斑的几何特征进行了充分的考虑，同样适用于其他特定形状的检测需求。比如在现有的计算机和图像采集设备中能够对于矩形的检测简单且高效。这种方法对形状进行了强调，所以并没有对运动这一概念进行深入的考虑。在大多数的应用场景中需要识别的光斑在连续帧中并不能表现出相同或者相似的固定形状，因此对形状进行识别的方法虽然值得借鉴，但并不是完全适用。

### (3) 基于频域峰值的检测

这种算法的检测精度非常高，尤其应用于光学中多个互相接近的亮点位置识别，比如卫星图中的行星位置。通过频域上的波峰对亮点的边缘位置进行定位，理论上能够实现对光斑边缘位置进行精确的判断，在抗干扰方面也有良好的表现。

## 2.3 视频采集技术

为了实现光斑的识别，需要对光斑信息进行采集，当前实现视频采集的方法主要有两种，第一种是利用视频采集卡生产厂商附带的软件开发包（SDK），第二种方法则是采用数字视频软件开发包。

### 2.3.1 视频采集卡

利用视频采集卡生产厂商附带的软件开发包（SDK）进行视频采集时，开发者需要采用开发包中采用的控件进行二次开发，这种方式最大的优点就是方便快捷操作简单。但是这种方式的缺点也十分明显，比如，SDK 必须与特定的设备相对应，而且 SDK 都是封装起来的，开发者无法进行一些底层的修改和操作，这样就对系统的开发和设计带来了许多不便。

### 2.3.2 数字视频软件开发包

利用数字视频软件开发包进行视频采集，这种方式对视频捕捉设备采集到的数据进行直接处理，这种方法相对于 SDK 来说，最大优点就是没有特定硬件设备的束缚。Windows 平台为实现数字视频的采集提供了技术支持，主要有 video for

windows (VFM) 技术和 DirectShow 技术。下面对这两种技术进行简单介绍。

### (1) VFM 技术

VFW 是 Microsoft 公司为开发 Windows 平台下的视频应用程序提供的软件工具包，提供了一系列应用程序编程接口 (API)，用户可以通过这种技术方便地实现视频捕获、视频编辑及视频播放等通用功能，还可利用回调函数开发更复杂的视频应用程序，其特点为播放视频时不需要专用的硬件设备，而且应用灵活，可以满足视频应用程序开发的需要。但是这种视频采集技术正在逐步趋于废弃。

### (2) Directshow 技术

Directshow 是 windows 平台上流媒体的体系结构<sup>[27]</sup>，这种体系结构在很大程度上简化了流媒体的格式变换以及捕获等工作。Directshow 的出现使应用程序不用再考虑繁琐的数据传输以及多种硬件的同步问题，从而简化了平台上多媒体应用程序开发任务。也因其具有模块化的结构，使之能够方便的对多种多样的数据源、文件格式和硬件设备进行操作。Directshow 解决的主要问题是快速处理流媒体的大量数据以及音频和视频的同步问题<sup>[28]</sup>，其原理图如图所示。

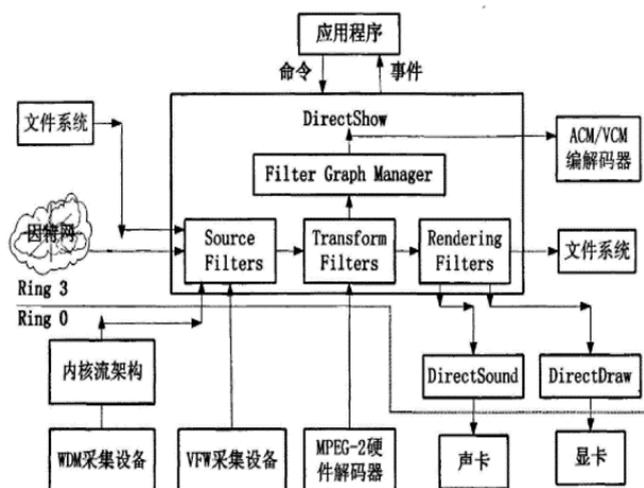


图 2-1 Directshow 原理图

Figure 2-1 schematic diagram of DirectShow

其中 Ring3 应用层上提供了一致的 COM 接口，实现了底层硬件设备和上层应用程序之间的分离。在 Directshow 中，总体应用框架和底层工作由 Directshow 完成，这样使多媒体应用的开发复杂度极大降低并在一定程度上提高了开发的效率。综上所述，Directshow 具有高效性、协调性、多样性以及可移植性等突出优点<sup>[29]</sup>。

## 2.4 屏幕校准

### 2.4.1 概述

由于受到摄像机摆放的空间位置限制，直接对摄像机采集到的视频信息进行识别并不能准确的描述光斑位置。本文对摄像头拍摄区域内的屏幕校准进行了具体研究。屏幕校准是指针对不同坐标系下拍摄到的一幅或多幅图像进行空间位置的匹配。摄像机采集到的图像相对于屏幕放映的图像，可能会产生一定程度的旋转、平移以及尺度上的变化，而且还会受到光照条件的影响；另外摄像头的摆放位置比较偏僻、投影变换失真、弧形投影屏幕等因素，都会造成采集的投影图像出现非线性畸变，因此屏幕校准算法应该对上述问题具有鲁棒性。下面将对屏幕配准需要的步骤进行详细介绍。

### 2.4.2 特征点提取

首先将获取大屏幕当前显示的图像作为训练图像，摄像头采集到的图像作为待配准图像。进行屏幕的自动化配准时，第一步要进行训练图像和待配准图像的特征点提取，特征点提取的质量将会直接影响到后续的屏幕配准工作。由于图像表达内容存在差异，图像特征分为全局特征以及局部特征两个方面<sup>[30]</sup>。全局特征对图像的整体特性进行了表达，一些常见的全局特征包括颜色特征、纹理特征、形状特征以及空间特征。局部特征指的是一种在各相互独立区域中经过筛选出来的特征<sup>[31]</sup>，只能作为图像局部特性的表达，并不能将一幅图像整体概括。全局特征和局部特征作为图片的表达各自拥有各自的优势，但是在图像配准的领域，局部图像特征具有相关度小以及及时环境复杂的情况下依旧具有较高的稳定性等优点，使其在图像配准领域占据绝对优势。图像的局部特征技术发展至今，出现了很多具有代表性的图像局部特征描述子，这些算术符可以适用于不同的用户需求，而且具有良好的性能，下面介绍几种应用广泛的局部特征描述符。

#### (1) 尺度不变特征转换 (Scale-invariant feature transform, SIFT)

首先于 1999 年被 David Lowe 发表在计算机视觉会议上，该方法通过对图像局部区域的侦测与描述从而生成了描述子的局部特征。SIFT 特征对图像的尺度、旋转变化均具有鲁棒性的局部特征描述，而且对光线、噪声以及视点变化也具有一定的稳定性<sup>[32]</sup>。SIFT 特征提取共包含四步：尺度空间极值检测、特征点定位、特征点方向赋值、特征点的描述<sup>[33]</sup>。

尺度空间极值检测：首先引入一个尺度因素的变量，通过改变这个尺度参数得到不同尺度下的图片，在这些尺度不同的图像上分别进行特征点的提取，使得计算

机对图像在不同尺度下有一个统一的认知。图像的尺度空间 $L(x, y, \sigma)$ 一般是通过图像和高斯函数卷积得来。

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (2-1)$$

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (2-2)$$

其中 $*$ 表示卷积运算,  $G(x, y, \sigma)$ 是指高斯函数,  $I(x, y)$ 指的是原图像,  $(x, y)$ 表示的像素点在图像矩阵中的空间坐标。 $\sigma$ 是迟钝因子, 尺度因子越大表示图像的模糊程度越高, 越小则表示图像的细节更加完善。尺度空间建立完善以后, 为了获得全局稳定的特征点, SIFT 采用了高斯差分空间将相邻的两个高斯空间的图像相减得到了高斯差分图像<sup>[34]</sup>。整个图像的局部特征是由高斯差分空间的局部极值点构成<sup>[35]</sup>, 在提取这种极值点时, 将每个点都和其周围的点做比较, 如果这个点保持和周围点不同, 那么可以判断这个点是一个局部极值点。

**特征点定位:** 在高斯差分空间中提取出的局部极值点还需要经过 Harris 角点检测算法和海森矩阵运算进行筛选, 对稳定特征点进行较高质量的保留。

**特征点方向赋值:** 为了保持旋转不变性需要对特征点的方向进行赋值。

**特征点描述:** 得到特征点的三个关键信息, 比如位置、尺度和方向, 最后使用一组向量对关键点进行描述, 使其对各种变化都具有稳定性。其生成步骤主要分为三个部分: 第一步调整主方向, 保持旋转稳定性; 第二步生成特征向量, 描述子需要在关键点尺度空间内计算 8 个方向的梯度信息; 第三步要进行归一化处理, 这一步消除了光照条件的变化对描述符稳定性的影响。

SIFT 特征对各种变化都能保持良好的鲁棒性, 在图像匹配等领域有着十分重要的作用。算法特征十分复杂, 整个算法耗时时间过长, 特征提取实时性不强。

## (2) 快速特征鲁棒特征算法 (Speed-Up Robust Features, SURF)

在 2006 年由 Bay 提出, 这种算法是对 SIFT 的一种改进算法, 利用积分图像来对图像进行卷积操作, 这种算法在计算速度方面有了很大的提升。SURF 相对于 SIFT 有以下几方面的提升:

第一方面 SURF 利用海森矩阵变换图像, 利用海森矩阵计算其矩阵行列式, 行列式值得正负号决定了极值点的判断结果。为了使特征点保持尺度不变性, 使用了盒状模糊滤波对海森矩阵进行滤波处理。

第二方面 SURF 保持图像大小不变, 改变盒状模糊滤波的大小来构建图像的尺度空间。该算法性能的提升主要通过同时对尺度空间中的多层图像进行操作实现。

第三方面 SURF 使用小波转换进行了关键点主方向的计算, 而不是使用直方图统计的方法。利用这种方法得到的向量大小比 SIFT 方法减小一半。所以效率也

比 SIFT 方法提高很多。

### (3) ORB 算法 (Oriented FAST and Rotated BRIEF, ORB)

这种算法主要是一种基于二进制描述子的特征提取算法，主要利用二进制描述子代替 SIFT 特征描述子从而大幅度的提高了算法的运行速度。ORB 算法主要分为两个步骤，第一步利用一种改进的 FAST (Features from Accelerated Segment Test, FAST) 检测算法获得带有旋转不变性的 ORB 特征点；第二步利用改进后的 BRIEF (Binary Robust Independent Elementary Features, BRIEF) 算法对特征点进行描述，其中 BTIEF 算法是一种基于二进制的高鲁棒性特征的特征点描述方法。

FAST 特征检测方法和传统的特征检测方法在运算速度方面有很大程度的提升，该算法的具体过程为：在图像中随机选取一个像素点，以这个像素点为圆心半径为  $R$  构建一个圆形邻域区域，该圆形邻域的原点为  $P$ ，将圆形外圈上的像素点与原点像素值做差，并将圆形外圈上的点做差值，求取绝对值并和阈值作比较，计算公式为：

$$|P_x - P_0| > \delta \quad (2-3)$$

其中  $P_x$  指圆圈外围的像素值， $P_0$  指中心像素点的像素值， $\delta$  表示阈值。在 FAST 算法中当一个点周围连续 12 个点的像素差满足条件时，则认为该像素点为 FAST 角点。而在 ORB 算法中只要一个像素点周围 9 个点的像素差满足条件即可，为了加快运算，从外围像素点中抽取编号为 1、5、9 和 13 四个像素点进行运算，这四个点中至少有三个点满足条件则继续与其他点作差值运算。使用这种方法对整幅图像进行遍历就能得到 FAST 角点，在此时利用 Harris 角点检测的思想根据一定规则选择合适的角点作为 ORB 特征点。为了使该点具有旋转不变性，将以  $P$  为圆心的圆形区域看成是一个有质量的物体，并把每个点的灰度值看成是该点的质量。用质心公式得到这个区域的质心  $Q$ ，将  $P$  到质心  $Q$  的方向设定为坐标系的横轴，垂直  $PQ$  连线且经过  $P$  点设为坐标系纵轴。此坐标系检测的特征点具有旋转不变性。

ORB 特征提取算法使用了 BRIEF 描述子对特征点进行描述标记，BRIEF 算法使用二进制编码的方式对提取到的特征点进行数学描述，这种方法对特征点周围的邻域信息封装到特征点描述子中。这种利用二进制的方式对特征点进行描述使得 ORB 特征点在匹配阶段相对于 SIFT 和 SURF 的运算用时很少，这种二进制的描述子占的存储空间也很少。

## 2.4.3 特征点匹配

图像特征的匹配阶段需要对两幅图像中相同的特征点进行寻找。将正确匹配

的点带入图像的几何变换模型中,这样就可以得到图像的变化矩阵,从而将待配准图像中的点映射到原始的屏幕图像中。对两幅图像是否匹配需要根据相似性度量来计算图片之间的相似程度来实现。相似性度量就是根据图像的特征编码描述符,对样本描述符之间的距离进行计算并排序。不同的相似性度量算法得到的匹配结果可能不同,常见的相似性度量准则<sup>[36]</sup>有欧几里得距离、曼哈顿距离、余弦相似度、汉明距离等。

欧几里得距离是对两个向量之间的物理层面距离进行计算,每个特征向量正常情况下各个分量的单位是不一致的<sup>[37]</sup>,因此需要对量纲进行统一标准,避免量纲不一致带来的麻烦。欧几里得距离的计算规则是,假设两个输入向量 $X = (x_1, x_2, \dots, x_n)^T$ 、 $Y = (y_1, y_2, \dots, y_n)^T$ ,这两个向量之间的欧几里得距离可以用如下公式计算:

$$D(x, y) = \|X - Y\| = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2} \quad (2-4)$$

余弦相似度是通过两个向量在向量空间形成的夹角来衡量相似性的。根据夹角对余弦值进行计算,余弦值越大说明两个向量之间越相似,反之则越不相似。当两个向量的方向相同时,夹角为 $0^\circ$ ,余弦值为1,当两个向量的方向完全相反时,夹角为 $180^\circ$ ,其余弦值为-1。公式为:

$$\cos \theta = \frac{X^T Y}{\|X\| \|Y\|} \quad (2-5)$$

汉明距离是使用在数据传输差错控制编码里面的,汉明距离是一个概念,它表示两个(相同长度)字符串对应位不同的数量,我们以 $d(x, y)$ 表 $x$ 与 $y$ 之间的汉明距离。对两个字符串进行异或运算,并统计结果为1的个数,那么1的个数就是汉明距离。

#### 2.4.4 误匹配点的剔除

在得到一系列匹配点对之后需要对匹配错误的点对进行剔除,这些点一般是由于匹配的误差过于大以至于不符合图像转换的关系。图像的基本变换模型有四种,分别是刚体变换、相似变换、仿射变换以及投影变换。这些基本变换模型的矩阵由不同图像基础变换组成的 $M$ 矩阵构成。

刚体变换是指图像中任意两点经过变换后距离保持不变,简单来说就是图像中物体的大小和形状都保持不变,这种变换主要由平移和旋转组成,其变换矩阵为:

$$M = \begin{bmatrix} \cos \theta & -\sin \theta & m_2 \\ \sin \theta & \cos \theta & m_5 \\ 0 & 0 & 1 \end{bmatrix} \quad (2-6)$$

点 $(x, y)$ 经过刚体变换后得到点 $(x', y')$ 的变换公式为:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} m_2 \\ m_5 \end{bmatrix} \quad (2-7)$$

其中 $\theta$ 指的是图像旋转角度， $m_2$ 和 $m_5$ 是平移量。

相似变换是指由一个图形变为另一个图像的过程中形状保持不变，但是图形的大小和形状可以改变，简单来说就是在刚体变换中加入尺度变换参数就变成了相似变换模型，其变换矩阵为：

$$M = \begin{bmatrix} \gamma \cos \theta & -\gamma \sin \theta & m_2 \\ \gamma \sin \theta & \gamma \cos \theta & m_5 \\ 0 & 0 & 1 \end{bmatrix} \quad (2-8)$$

点 $(x, y)$ 经过刚体变换后得到点 $(x', y')$ 的变换公式为：

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \gamma \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} m_2 \\ m_5 \end{bmatrix} \quad (2-9)$$

其中 $\gamma$ 代表尺度， $\theta$ 代表旋转， $m_2$ 和 $m_5$ 代表平移量。

仿射变换比相似变化的原子变化多了一部分错切变换，图像经过仿射变换后二维图像中的直线依然是直线，平行线依然是平行线。仿射变换的矩阵为：

$$M = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ 0 & 0 & 1 \end{bmatrix} \quad (2-10)$$

点 $(x, y)$ 经过刚体变换后得到点 $(x', y')$ 的变换公式为：

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} m_0 & m_1 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} m_2 \\ m_5 \end{bmatrix} = A \begin{bmatrix} x \\ y \end{bmatrix} + T \quad (2-11)$$

其中 A 矩阵表示旋转、缩放和错切等几个基本变换矩阵，并且 A 矩阵可以进行奇异值分解：

$$A = \gamma \begin{bmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} \quad (2-12)$$

其中 $\phi$ 是指相机的视角， $\psi$ 对应了相机的旋转角度。

点 $(x, y)$ 经过刚体变换后得到点 $(x', y')$ 的变换公式为：

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \gamma \begin{bmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} m_2 \\ m_5 \end{bmatrix} \quad (2-13)$$

投影变换主要用于描述相机在一定距离内任意的视角和位置对图像进行拍摄得到的结果。图像经过投影变换直线能继续保持直线但是平行线无法继续保持平行。投影变换的矩阵为：

$$M = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & 1 \end{bmatrix} \quad (2-14)$$

点 $(x, y)$ 经过刚体变换后得到点 $(x', y')$ 的变换公式为：

$$x' = \frac{m_0x + m_1y + m_2}{m_6x + m_7y + 1} \quad (2-15)$$

$$y' = \frac{m_3x + m_4y + m_5}{m_6x + m_7y + 1} \quad (2-16)$$

上述四种变换中，投影变换是比较一般的变化，可以通过投影变换进行其他三种变换的转化。当投影变换矩阵的最后一行变为 $[0 \ 0 \ 1]$ 时，退化为仿射变换。当仿射变换 M 矩阵左上角 $2 \times 2$ 矩阵是正交的，则仿射变换退化为相似变换。当相

似变换  $M$  矩阵左上角  $2 \times 2$  矩阵的行列式为 1，那么相似变换退化为刚体变换。

在图像配准技术中，根据图像之间的特点，经常采用的图像变换模型有相似变换、仿射变换以及投影变换等。其中相似变换模型的配准效率最高，这种模型适用于视角相同但是拍摄位置不同的两幅图像之间的配准，比如在获取全景图像时，如果拍摄的景象与相机之间的距离与景象自身的大小之间相差较大，那么可以将图像之间的变换模型看作是相似变换模型，这类模型典型的应用对象是卫星遥感图、航拍图等。投影变换的精度最高，但是计算复杂度也很高，这种变换可以反映在成像时平面和场景平面不平行带来的梯形失真和线性调频现象，这是其它变换无法做到的。仿射变换模型适用于大多数常见的图像配准情况，它可以对相似变换模型对图像计算带来的误差进行弥补，对错切变换带来的误差进行处理。不同的变换模型具有不同的适用范围和处理结果，需要根据实际的应用场景和需求进行具体的选择，合适的变换模型可以使配准结果的效率和计算速度方面都有较好的效果。

最小二乘法是进行误匹配点剔除的基础理论，这种方法是一种数学优化技术，主要利用是模型误差的平方和最小化来找出原始数据集中的最佳匹配模型。对于点  $(x_i, y_i)$  经过投影变换矩阵得到对应点  $(x'_i, y'_i)$ ，其中  $i = 0, 1, 2, \dots, N - 1$ ，由上述公式得到：

$$m_0x_i + m_1y_i + m_2 - m_6x_ix'_i - m_7y_ix'_i - x'_i = 0 \quad (2-17)$$

$$m_3x_i + m_4y_i + m_5 - m_6x_iy'_i - m_7y_iy'_i - y'_i = 0 \quad (2-18)$$

将上式写成矩阵形式，有  $AX = 0$ ，其中  $A$  是一个  $2N \times 9$  的矩阵， $X = [m_0, m_1, \dots, m_7]^T$ ，利用最小二乘法进行求解就是找出令  $\|AX\|$  最小的  $X$ ，简单来说就是求与  $A^T A$  的最小特征值相对应的特征向量。最小二乘法可以直接对矩阵进行计算求解，这种方法的实现较为简单。当数据集的整体误差较小时，大部分数据接近模型时，这种方法能够做到模型参数的有效求解。但是当数据集的整体误差较大时，最小二乘法无法保证模型求解的有效性。经过上述分析可知，最小二乘法在计算量和计算速度上具有显著优点，但是只有在待配准点集中误匹配点对所占比例较小时才能够较为准确的预测变换矩阵，实现误匹配点对的剔除。然而，当误匹配点集的比例稍大，就会使最小二乘法很难计算出准确的变换矩阵，最小二乘法很难满足问题的需求。

## 2.5 本章小结

本章主要介绍本文用到的相关技术，主要对屏幕演示系统的构成以及需求进行了介绍，对人机交互中的激光笔产品进行调研分析，并对激光光斑的识别技术进行了介绍。同时，本章介绍了两种视频采集方式，用于捕获激光光斑信息。最后，

本章介绍了屏幕校准方法，主要包括特征点提取、特征点匹配以及剔除误匹配点三个方面。

### 3 系统总体设计

本章设计了基于激光笔交互的智能沙盘系统进行了设计，主要分为四部分，第一部分为系统的研究背景，主要对目前现有的典型系统进行介绍并分析不足；第二部分对系统的需求进行了分析，根据目前沙盘系统中存在的缺点提出本文研究应满足的需求；第三部分根据系统的需求进行系统的整体设计，在完成硬件架构设计与硬件设备选型的基础上完成软件部分的流程设计；第四部分为本章小结。

#### 3.1 研究背景

沙盘是指以一定的比例及方式体现实体及实体某些特征要素的模型，已广泛应用于各个领域，例如城市规划、智能交通和军事部署等。目前的沙盘系统主要有两种表现形式，分别是物理沙盘系统和电子沙盘系统。物理沙盘系统一般是指人们按照一定比例用泥沙等材料制作的沙盘系统，虽然可以通过教鞭或者激光笔进行简单的交互，但是存在布置过程复杂、信息不能及时更新等缺点。电子沙盘利用多媒体技术等现代化手段实现了图形等形式信息的展示，其布置过程虽然简单，但是显示的信息量很少，无法实现图形信息的立体化显示，并且需要通过鼠标或者键盘这种传统的方式来实现人机交互，这限制了演示者的活动范围。近年来，沙盘系统的表现形式越来越多样化，人们开始考虑同时结合物理沙盘和电子沙盘进行信息的多维立体化展示，这种结合既能够生动立体的显示图形信息，布置方式相对简单，也缓解了信息更新不及时的问题。

但是，目前利用物理沙盘和电子沙盘相结合的方式信息进行展示的沙盘系统还存在以下问题：

(1) 当演示者与观众和物理沙盘以及屏幕之间的距离都较远时，并使用激光笔在沙盘或者屏幕上指示时，会存在目标光斑很小，人眼难以看到光斑具体位置的问题。

(2) 目前系统中大屏幕部分显示信息需要提前进行设置，同时播放的视频只能按照预设时间线进行播放，无法根据现场实际情况做出调整，也无法根据演示者的意图实现交互功能。

(3) 虽然一些屏幕显示系统能够实现部分交互功能，但是进行交互行为的操作十分不便，演示人员需要同时手持教鞭和鼠标才能同时实现对物理沙盘和大屏幕的交互行为。

本课题组其他成员完成了针对物理沙盘的醒目化设计工作，包括图像采集、

POI 域选择、光斑识别与检测等功能<sup>[38]</sup>。工作流程为：利用摄像头实时捕获激光笔发出的光斑，将带有光斑信息的视频流传输到计算机中。计算机主机接收到视频后，对视频帧进行处理并计算出光斑的坐标信息，将该坐标传递给 DMX512 设备，利用该设备对灯光进行一系列实时控制，使其照亮光斑区域，醒目化显示激光笔的指示，此外，该设备还可以实现指向跟踪、亮度控制、预定义的图案显示等功能。但是目前还存在以下问题：

(1) 配置方式复杂。当摄像头的拍摄区域无法实现与物理沙盘的完全对准，需要用户进行手动配置，用户参差不齐的操作水平为沙盘系统的初始化配置过程带来困难。

(2) 仅解决了物理沙盘的部分交互问题，仍然无法实现展示信息的及时更新，布置复杂的问题也没有得到解决。

## 3.2 需求分析

本节中，首先对现有沙盘系统的缺点和需要改进的问题进行了分析和总结，结合对现有智能沙盘系统的完成情况和仍存在的问题进行分析和总结。本文在对智能沙盘系统领域长期调研的基础上，在目前智能沙盘系统中已有实体沙盘部分的研究基础上进行了需求分析，总结如下：

在信息展示形式方面，目前智能沙盘系统已完成的部分只有实体沙盘部分，这部分虽然能够实现信息的三维展示，但是实体沙盘部分的信息更新不够及时，布置的方式复杂。

在醒目化指示方面，目前利用激光笔进行交互的沙盘系统都无法对激光光斑进行醒目化的标识，当进行远距离的交互时，演示者和观众都无法看清楚激光笔的具体指示位置。目前智能沙盘已有实体沙盘部分只能在实体沙盘的基础上利用灯光控制系统实现醒目化的提示。

在交互灵活性方面，虽然现在已有沙盘系统采用物理沙盘和电子沙盘相结合的方式对信息进行多维立体化展示，但是这种沙盘系统的交互需要演示者同时手持教鞭和激光笔两种交互设备才能实现。

在使用简单便捷方面，目前已有智能沙盘的实体沙盘部分，其摄像头摆放位置受到空间条件的约束，需要演示者手动进行系统的初始化配置，演示者的软件使用水平参差不齐，无法保证初始化配置的准确，而且也增加了系统在使用过程中的复杂程度。

在智能沙盘系统方面，本文计划在现有实体沙盘系统的基础上，对整个智能沙盘系统的设计与实现进行完善，在交互相对灵活的前提下实现利用物理沙盘与电

子沙盘相结合的方式对信息图像的立体化展示，并且对于演示者来说该系统使用的复杂程度不高。总的来说，本文将实现如下需求：

(1) 在信息展示形式与交互灵活性方面，在现有实体沙盘的基础上增加大屏幕显示设备作为信息展示的载体，在交互设备方面仅采用激光笔作为交互设备，减弱了演示者需同时使用两种交互工具带来的不便。

(2) 在智能沙盘系统的整体设计方面，对激光笔、摄像头等关键硬件设备进行选择，根据该系统所需实现的特殊需求对硬件设备进行设计，并对整个硬件结构进行设计，针对硬件的结构、硬件自身能够实现的功能以及硬件设备选型带来的问题实现软件部分的设计，最终硬件架构和软件架构组成了智能沙盘系统的整体结构。

(3) 在系统易用性方面，由于摄像头摆放位置受到空间条件的限制，该系统已完成的实体沙盘部分的系统初始化配置需演示者手动完成，系统初始化配置的完成程度受到了演示者操作水平的限制，而且配置的过程降低了整个系统的易用性，本文研究内容部分计划采用一种自动化的配置方式，该自动配置功能不严格约束摄像头与屏幕的空间摆放位置，自动化的实现智能沙盘系统的初始化配置。

(4) 在醒目化指示方面，目前现存的物理沙盘和实体沙盘没有考虑到这一问题，该智能沙盘系统已完成的实体沙盘部分利用灯光控制设备完成了激光光斑的醒目化指示，加入大屏幕显示设备之后也要考虑光斑在屏幕上的醒目化显示问题。

此外，在交互的灵活性方面，本系统前期在物理沙盘部分仅通过红外激光笔发射的激光光斑进行交互，本课题组其他成员正在实现通过语音对灯光设备进行控制的部分，该部分的设计与实现进一步提高了整个系统的灵活性。

### 3.3 系统设计

#### 3.3.1 总体思路

本节中，将对基于物理沙盘和电子沙盘相结合的智能沙盘系统进行具体的设计，首先对目前该系统的完成情况进行总结，并对课题组其他成员的工作进行简要介绍，在这些工作的基础上介绍智能沙盘系统的整体设计思路以及整体架构，并对硬件架构进行设计，根据该架构实现硬件设备的选择，最后根据系统需求、硬件设备具备的功能以及硬件设备选择带来的问题对系统软件部分进行设计。

在改进前，本智能沙盘系统只有实体沙盘部分，可实现图像采集、POI 域选择、光斑识别与检测等功能。工作流程为：利用摄像机实时捕获激光笔发出的光斑，将带有光斑信息的视频流传输到计算机中。计算机主机接收到视频后，对视频帧进行处理并计算出光斑的坐标信息，将该坐标传递给 DMX512 设备，利用该设备对灯

光进行一系列实时控制，使其照亮光斑区域，醒目化显示激光笔的指示，此外，该设备还可以实现指向跟踪、亮度控制、预定义的图案显示等功能。

本文为智能实体沙盘增加了基于遥控激光笔的屏幕显示功能，利用遥控激光笔在演示屏幕上发射激光光斑，可以在进行醒目化演示的同时将演示者从计算机前或者演讲台上解放出来，摆脱传统人机交互例如键鼠等带来的有线距离约束；此外，还可以为遥控激光笔添加按键将其转化为远程鼠标，可以实现传统鼠标的部分功能，例如指示、点击、拖拽等功能，实现了演讲者与演示屏幕之间的灵活交互。

此外，本小组的另外一名成员为智能沙盘系统增加了语音识别功能，该功能可以识别演示者说出的关键字指令，例如“蓝方攻”、“双方在”等，并根据指令控制灯光设备的操作，实现了演讲者与灯光设备之间的灵活交互。

本智能沙盘系统的总体示意图如图 3-1 所示。该系统由以下部分组成：实体沙盘、计算机主机、DMX512 灯光控制设备、灯光设备、摄像头、激光笔、话筒、多媒体演示屏组成。该智能沙盘系统通过遥控激光笔与语音识别功能实现了用户与灯光设备、用户与演示屏幕之间的灵活交互，其中红框内的部分为本文研究的主要内容。

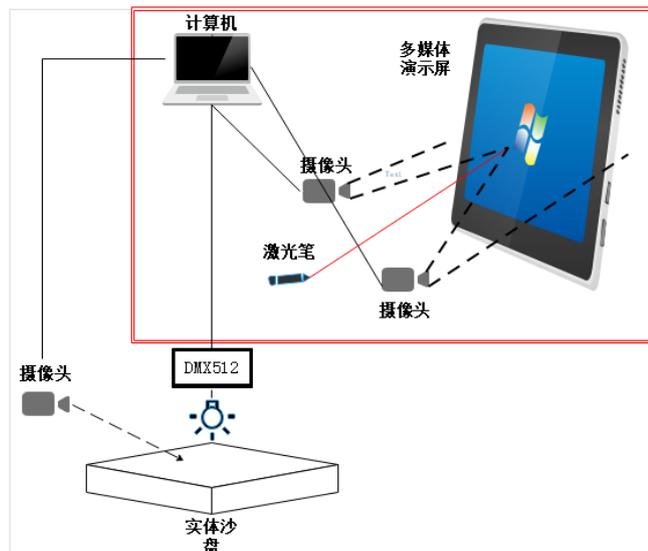


图 3-1 智能沙盘系统总体示意图

Figure 3-1 General diagram of intelligent sand table system

本节对智能沙盘系统进行了具体设计，图 3-1 为智能沙盘系统的整体结构图，其中涉及到的硬件以及硬件需要实现的功能如下：

激光笔：同时发射可见激光和红外线光斑、远程操控鼠标；

大型屏幕：显示图像信息；

实体沙盘：三维立体化演示实体；

摄像头：捕获屏幕、实体沙盘和红外线光信息。

计算机：处理摄像头输入的视频流，定位光斑位置，实现交互操作。

DMX512：基于 USB 接口以及 RS485 协议，实现对灯光设备的控制。

灯光设备：负责灯光照明，醒目化标识激光光斑，以及根据用户说出的指令完成相应操作。

其中图 3-1 中红框部分为本文主要研究内容和实现部分，本文研究范围主要涉及激光笔的选型、摄像头设计选择以及计算机软件处理流程的设计，其中软件部分的设计主要根据选定硬件设备已实现的功能和硬件选型带来的问题进行设计，主要实现的功能包括：屏幕的自动化校准、光斑的识别与定位、光斑坐标的转换和光标的追踪。

### 3.3.2 硬件架构

系统硬件的总体结构如图 3-2 所示。其中红框部分为本文主要研究内容，其中关键硬件包括激光笔、摄像头、滤光片。

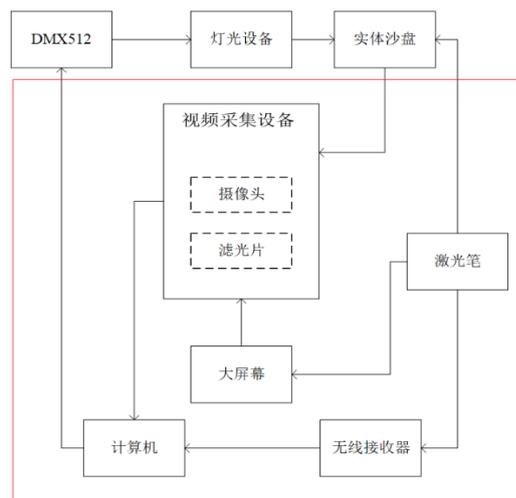


图 3-2 硬件总体结构

Figure 3-2 Overall hardware structure

#### (1) 激光笔选型

经过研究与分析，本文选用激光笔作为交互设备，综合图像处理、模式识别等技术来实现用户与演示屏幕之间的交互。在基于遥控激光笔的交互系统中，用户可以使用激光笔在屏幕上进行指示，并通过激光笔上的按键远程完成交互。这种方式摆脱了传统交互设备对交互距离的限制，在一定程度上提高了交互的灵活性和普适性。采用激光笔这种交互方式能够激发用户更多的肢体动作，有利于用户准确表达自己的意图。

用户使用激光笔在屏幕上进行指示时，激光笔会同时向屏幕的同一位置发射

可见激光光斑和红外线光斑。可见激光光斑用于屏幕的醒目化指示，红外线光斑会被带有滤光片的摄像头捕捉。摄像头将采集到的屏幕数据传输到计算机进行图像处理，通过分析红外线光斑的位置和接收到的激光笔按键指令来完成用户的交互请求，如鼠标点击等。

本系统采用的激光笔如图 3-3 所示，有如下优势：

- 1) 使用可见激光光斑在屏幕上进行指示，实现醒目化标识；
- 2) 使用红外线光斑可以避免环境中复杂可见光的干扰，实现计算机对红外线光斑的精准定位。
- 3) 操作简单、方便携带。



图 3-3 激光光斑发射器

Figure 3-3 laser spot transmitter

该激光光斑发射器与计算机之间的通信通过无线接收器实现，无线接收器如图 3-4 所示。

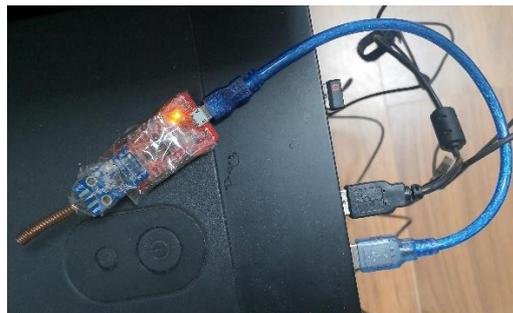


图 3-4 无线接收器

Figure 3-4 wireless receiver

该无线接收器由无线接收模块和串口转 USB 芯片 CH340 两部分组成：

- 1) 无线接收模块通过 2.4GHz 频率的无线信道来接收激光笔发出的红外线和按键动作的信息；
- 2) 串口转 USB 芯片对无线接收模块接收到的串口信息进行转换，便于计算机的读取与处理，提高无线接收器的普适性。

## (2) 摄像头选型及摆放

摄像头主要分数字摄像头和模拟摄像头两大类。模拟摄像头通过特定的视频捕捉卡将模拟视频信号转换为数字信号，并经过压缩后才能计算机上进行存储和运用。数字摄像头可以将捕捉到的影像直接转化为数字信号，然后通过串并口或者 USB 接口进而将其存储在计算机里。

本文研究的系统采用数字摄像头以及 USB 接口来实现对视频信息的捕捉与传输。具体设备如图 3-5 所示。



图 3-5 摄像头

Figure 3-5 camera

该设备利用 USB 数字摄像头来采集视频有如下优势：

1) 安装快捷、操作方便。数字摄像头无需特定的视频捕捉卡即可直接捕捉数字图像，便于计算的存储与处理；使用 USB 进行连接可以实现即插即用，无需关机再开机等动作。

2) 支持连接多种设备。USB 是一种通用连接器，因此选用 USB 接口来完成数据的传输可以提高摄像头的通用性和普适性，使得摄像头可以连接多种设备。

3) 支持较远的连接距离以及提供较快的传输速度。通过延长 USB 传输线的长度可以使得摄像头在较远处进行拍摄；此外，USB Version 2.0 能够支持 480Mbps 的高速传输速率，实现视频流的实时传输。

本文选用的摄像头焦距为 8mm，选用屏幕的大小是 5m×2.5m，计划摄像头距离屏幕的距离为 5m，受到摄像头的焦距限制，此距离下摄像头的覆盖范围为 3.5m×2.6m，为了完成整个屏幕的覆盖，需要使用两个摄像头对屏幕进行拍摄。

## (3) 滤光片应用

当使用摄像头捕获的视频信息进行激光光斑的识别与定位时，自然光线和屏幕本身的亮度都是不可避免地影响因素。为了减弱这种因素对光斑识别准确度地影响，本系统为每个摄像头配置一个滤光片，滤光片主要用于减弱可见光的影响，

可见光是指电磁波谱中人眼可以感知的部分，一般人眼可以感知到的光波域为 400nm（紫光）到 700nm（红光）之间，而波长 760nm 到 1mm 之间则为红外线。采用红外滤光片可以增强红外激光光斑与背景环境的区分度，提高亮点识别的精确度，减小算法的复杂度。使用拥有截至波长特性的滤光片作为摄像头的辅助设备，可以选择使特定范围波长的光波通过，而滤除其他不希望通过滤光片的波段。滤光片实物如图 3-6 所示，普通摄像头拍摄屏幕的效果如图 3-7 所示，加滤光片后摄像头拍摄效果如图 3-8 所示。通过这两张图，可以看出滤光片能够实现可见光的滤除，可以简化激光光斑的识别和定位过程。



图 3-6 滤光片实物图片

Figure 3-6 physical picture of filter



图 3-7 摄像头拍摄画面

Figure 3-7 camera image

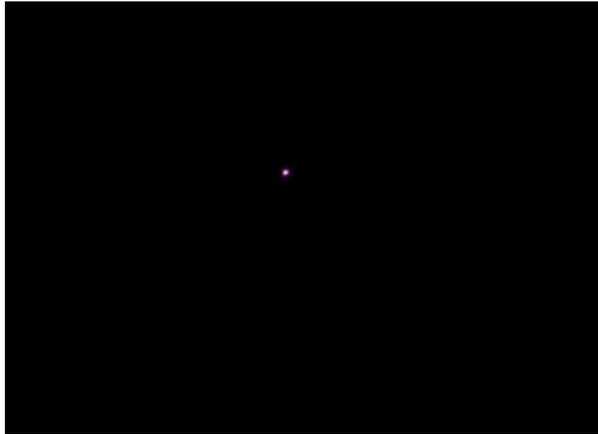


图 3-8 加滤光片后摄像头拍摄效果

Figure 3-8 camera shooting effect after adding filter

#### (4) 硬件设备选择带来的问题

为了实现本系统中的设计需求，根据本节中选择的特定硬件设备来完成用户与屏幕之间的交互，但是在满足要求的同时也会带来如下问题：

1) 激光光斑发射器的红外线不是一直发送，红外线存在闪烁频率，由于红外线闪烁频率可以进行用户的区别与标识，所以闪烁的情况不可避免，但是这种闪烁会对激光光斑的定位与识别带来一定难度。

2) 摄像头的摆放位置受到空间条件的限制，无法做到与屏幕完全重合，严重影响了光斑位置的追踪。

### 3.3.3 软件架构

根据系统设计需求以及硬件设备选择带来的问题，该智能沙盘系统中本文研究部分主要为演示者与大屏幕交互部分，其软件部分的主要工作流程如图 3-9 所示：

- (1) 初始化配置屏幕演示部分，校正由摄像头拍摄角度所引起的图像偏差。
- (2) 摄像头实时捕获激光光斑信息，将带有光斑信息的视频流传输至计算机进行处理，并根据校正结果计算出光斑位置在视频帧上的坐标信息。
- (3) 转化上述坐标，并得到光斑在演示屏幕上的对应坐标。
- (4) 利用演示屏幕上的光斑坐标进行鼠标操控，完成用户与演示屏幕之间的交互。

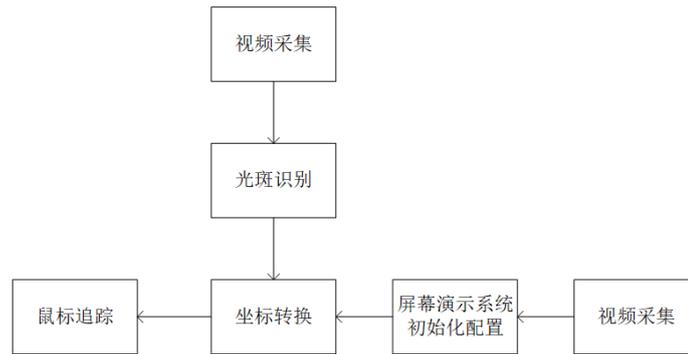


图 3-9 软件工作流程

Figure 3-9 software workflow

### 3.4 本章小结

本章主要介绍了智能沙盘系统的研究背景、应满足需求以及系统总体设计。本系统的研究目的是解决物理沙盘和电子沙盘的操控问题；实现基于红外激光笔的电子沙盘鼠标操控核心功能，解决电子沙盘的灵活交互问题。本系统设计部分首先进行了关键硬件设备的选择，其次根据系统的需求和硬件设备带来的问题进行软件部分的设计，软件部分主要实现屏幕自动化校准、光斑识别、坐标转换与光标追踪三个主要功能。

## 4 软件实现

上一章主要介绍了智能沙盘系统的整体设计、关键硬件的选择以及软件部分的简要设计，本章介绍了基于红外激光笔的电子沙盘鼠标操控核心功能的具体实现，主要包括五个部分。第一部分首先描述了系统软件部分需要解决的关键问题，并将软件部分划分为屏幕校准、光斑识别、坐标转换与光标追踪三个功能模块；第二部分主要介绍了屏幕的自动化校准实现步骤；第三部分介绍如何对红外激光笔发射的激光光斑进行识别与定位；第四部分介绍坐标的转换以及光标追踪的具体实现细节；第五部分为本章小结。

### 4.1 概述

#### (1) 关键问题

根据 3.2 部分对智能沙盘系统的需求进行分析以及 3.3.2 部分硬件选型后带来的不可避免地问题。该系统的软件部分面临以下几个关键问题。

1) 摄像头摆放位置无法实现与屏幕显示的完全对准，为了避免用户进行复杂的屏幕配准操作，本文设计系统需要实现自动化的屏幕配准功能。

2) 激光光斑不是一直出现在摄像头捕捉到的视频帧中，本文设计系统需要考虑准确快速的对激光光斑进行定位。

3) 为了实现灵活便捷的人机交互，本文设计系统需要将利用摄像头捕捉到的光斑通过屏幕配准获得的屏幕原点坐标进行坐标的转换，并将屏幕的鼠标定位到激光光斑在屏幕上的实际位置。

#### (2) 系统软件设计

根据上述问题描述，以及系统结构的设计，本文设计系统软件部分主要包含屏幕配准模块、光斑识别模块、坐标转换与光标追踪模块。各部分具体工作流程如图 3-10 所示。

1) 屏幕校准模块主要用于实现屏幕的自动化校准。原有实体沙盘部分也存在摄像头的视野范围无法对准实体沙盘的情况，本小组其他成员的解决办法是在系统初始化配置时利用红外激光笔在实体沙盘的角点进行打点并对该坐标的具体位置进行识别，利用该坐标值作为坐标转换的参考坐标。这种方法虽然能够实现系统的初始化配置，但是这种配置方式需要用户手动操作，用户的操作水平和理解能力参差不齐，其配置准确率和配置时延方面可能会受到影响。本文研究系统采用了一种自动化的屏幕配准方案，该模块首先需要获取屏幕显示的图像作为训练图像，摄

摄像头捕获的图像帧作为待配准图像，分别对这两张图像中的特征点进行提取，根据特征点之间的对应关系可以计算出特征点之间的变换矩阵，进而计算出屏幕角点的坐标，整个过程实现了屏幕配准的自动化。

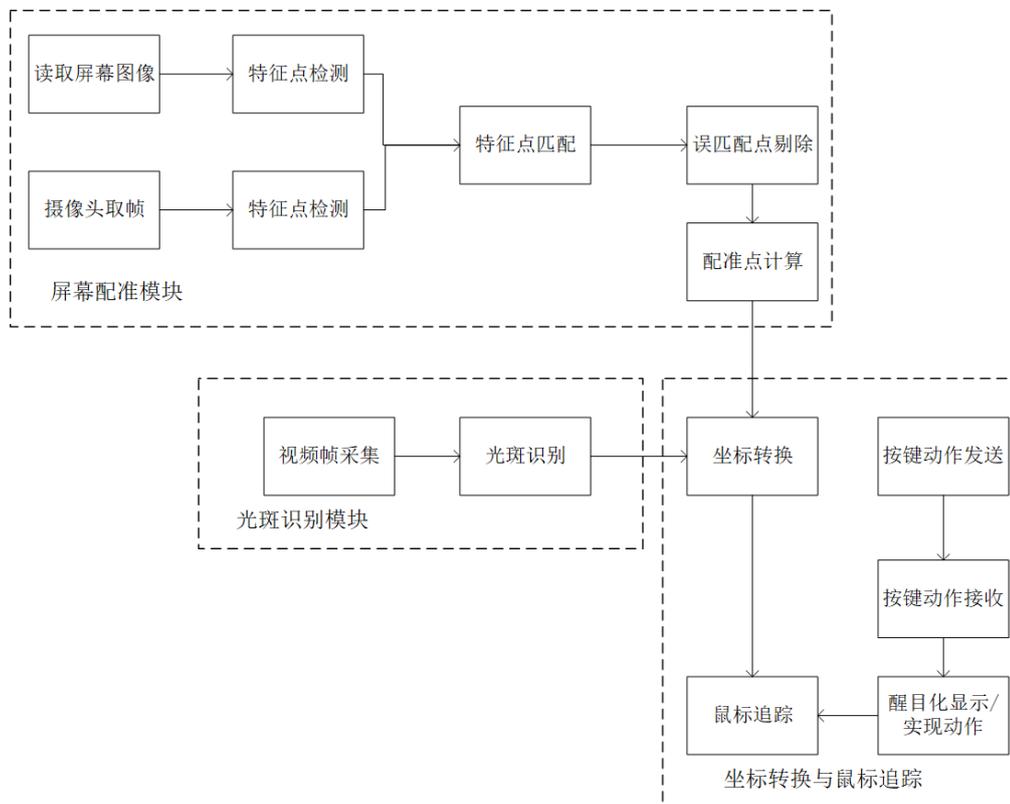


图 4-1 软件部分工作流程

Figure 4-1 specific workflow of software

2) 光斑识别模块主要用于实现光斑的识别和光斑的定位。由于采用激光笔作为交互设备，需要对激光笔发射的红外光斑进行识别与定位。在硬件选择部分，本文研究系统选用的红外激光笔能够同时发射两种光线，其中可见光光线主要用于实现光斑指示位置的可视化，红外光线主要用于实现计算机对光斑位置的识别。由于屏幕自身光线、灯光、可见光等无法完全避免的影响会对光斑的识别造成较大影响，本文研究系统为摄像头前面配置了滤光片，滤光片滤除了大部分可见光，强化了红外光斑在捕获图像中的显示效果，但是由于红外激光笔的红外线具有闪烁频率，并不是处于一直发送红外线的状态，这为光斑的捕捉和识别时间带来了挑战。

3) 坐标转换与光标追踪模块主要用于实根据激光光斑的识别结果和屏幕配准中的配准点计算结果，通过分析摄像头坐标系和屏幕坐标系之间的几何关系，实现摄像头坐标系下坐标值到屏幕坐标系下坐标值的转换。当屏幕距离摄像头连接的计算机较远时，通过 UDP 协议实现光斑具体坐标和激光笔按键动作消息的传输，屏幕主机部分根据接收到的光斑坐标实现光标的追踪。

本文在同一台电脑上实现了屏幕的自动配准与激光光斑的识别与控制工作，该电脑处理器为 Intel(R)Core(TM)i7-8550U CPU@1.80GHZ 2.00GHZ；内存(RAM)为 16GB；系统类型为 64 位 windows 10 操作系统。本文研究系统程序实现部分主要基于 Visual Studio 2019 平台并调用以 C++ 语言编写控制台程序调用 OpenCV4.1.1 库函数来实现，其中 Microsoft Visual Studio 2019（简称 VS2019）是一个可视化工具集合，在同一个开发环境中集成了代码编辑、资源编辑器等工具，利用项目、项目集等组织概念，简化了项目开发到发布的流程；OpenCV 是开源的计算机视觉库，主要由 C 函数和少量 C++ 类构成，效率高是其突出优点，具有丰富的图像处理 and 计算机视觉算法。

## 4.2 屏幕校准

### 4.2.1 概述

由于摄像头摆放位置的限制，摄像头采集到的屏幕和真实屏幕之间存在偏差。这种偏差如图 4-2、图 4-3 所示。



图 4-2 屏幕信息显示

Figure 4-2 screen information display



图 4-3 摄像头捕捉到的屏幕

Figure 4-3 screen captured by camera

屏幕校准部分需要解决的问题如图 4-2、图 4-3 所示。在对本文研究的智能沙盘系统最初实现的智能实体沙盘进行研究实现时，也存在摄像头采集到的实体沙盘与空间中的实体沙盘存在偏差的问题，最初的只能实体沙盘设计人员手动的方式对沙盘平面坐标系进行矫正，主要方式为在系统初始阶段用户手持激光笔对矩形实体沙盘的四个点进行点击指示，摄像头对这一系列行为进行捕捉，计算机通过对摄像头输入的视频流进行分析得到矩形实体沙盘的四个顶点的坐标位置，通过这四个点的光斑坐标与摄像头捕获区域坐标系之间的几何关系，对摄像头捕捉到的实时光斑坐标进行转换，转换结果通过串口通信的方式发送到 DMX512 灯光控制设备，实现了对光斑的醒目化提示。

这种方法存在的显著缺点就是不够智能化，需要用户手动进行屏幕的校准，用户无法准确对矩形实体沙盘进行点击时，屏幕校准的速度和准确度都是受到影响。考虑到这个缺点，本文研究的智能沙盘系统考虑了一种自动化的屏幕校准方式，屏幕校准的流程如图 4-4 所示。下面将对具体实现方式进行简单介绍。

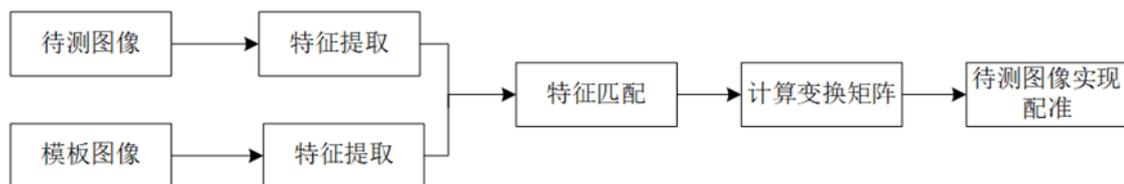


图 4-4 屏幕配准流程

Figure 4-4 screen registration process

## 4.2.2 特征点提取

根据第二章相关研究的介绍，主要的特征点提取方法有：SIFT 特征点提取算法、SURF 特征点提取算法、ORB 特征提取算法。其中 SIFT 特征提取算法以及 SURF 特征提取算法相对于 ORB 特征提取算法的耗时很长，基于用户交互行为需要实时性较高，所以本文选择 ORB 算法进行特征点的提取。

ORB 特征点提取算法主要由 FAST 角点检测算法和 BRIEF 描述子两部分构成。其中 FAST 角点检测算法的实现流程为：

- (1) 将图像转化为灰度图像后，选取一点  $P$ ；
- (2) 设定阈值  $t$ ；
- (3) 以  $p$  为圆心画圆，设定半径为 3，对圆上的全部 16 个像素点进行分类；

$$S_{px} = \begin{cases} a & I_{px} \leq I_p - t \\ b & I_p - t < I_{px} < I_p + t \\ c & I_p + t \leq I_{px} \end{cases} \quad (4-1)$$

其中  $a$  类表示亮度比  $p$  点暗的点， $b$  类表示亮度和  $p$  点相差不大的点， $c$  类表

示亮度比  $p$  点亮的点。

(4) 当  $p$  点附近包括或者超过所有像素点的四分之三的分类不是  $b$  类时,  $p$  点被判定为角点;

BRIEF 描述子的生成流程为:

(1) 将特征点周围的一个大小为  $S \times S$  的正方形区域选定为建立描述子的区域;

(2) 对选定的这个区域做高斯平滑处理, 因为 BRIEF 算子仅考虑图像邻域中单个像素的对比, 所以对噪声敏感, 预平滑处理能够提高 BRIEF 描述子的稳定性。

该描述子定义了准则:

$$\tau(p; x, y) = \begin{cases} 1 & p(x) < p(y) \\ 0 & p(x) \geq p(y) \end{cases} \quad (4-2)$$

其中  $p(x)$  和  $p(y)$  分别代表邻域  $p$  内图像位置  $x$ 、 $y$  处的像素值。测试集  $(x, y)$  位置对的生成方式有很多种, BRIEF 描述符的采样方法是  $x_i$  和  $y_i$  服从以特征点为中心, 高斯核为  $\frac{1}{25} S^2$  的高斯分布。位置点对的数目可以是 128、512 等。

本文选择选用  $590 \times 334$  像素大小的屏幕截图作为训练图像,  $590 \times 504$  像素大小的摄像头捕捉图像作为待配准图像对 ORB 算法的特征点检测算法进行了测试, ORB 算法中第一个参数 `nfeatures` 表示最多可以采集到的特征点数目, 这里设置为 500, 第二个参数为 `scaleFactor` 表示金字塔抽取比, 也就是金字塔图像之间的尺度参数, 这里设置为 1.2f。处理结果如图 4-5、图 4-6 所示。



图 4-5 训练图像特征点提取结果

Figure 4-5 feature point extraction results of training image

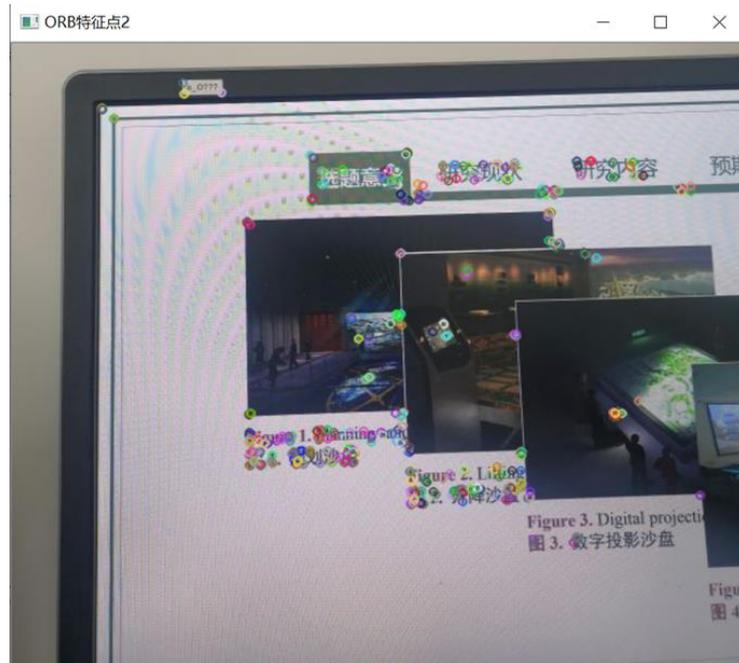


图 4-6 待配准图像特征点检测结果

Figure 4-6 detection results of feature points of image to be registered

### 4.2.3 特征点匹配

取得特征点及其描述之后需要进行特征的匹配，目前的匹配方法主要有两种，分别是暴力匹配（Brute Force, BF）和快速最近邻逼近搜索函数库（Fast Library for Approximate Nearest Neighbors, FLANN）。这两种方法的区别在于第一种会对所有可能的匹配进行尝试，从而找到最佳的匹配，而 FLANN 可以在短时间内找到相对好的算法进行特征的匹配。

FLANN 包含了一系列的查找算法和一种自动选取最快算法的机制。Muja 和 Lowe 为了减少计算机视觉以及机器学习需要大量训练问题，对随机 k-d 森林法等算法进行了大量的研究。快速最近邻库（FLANN）是一种开源库，其优点在于能够在较短的时间内进行一次相对好的匹配，但是无法保证是最佳匹配。而暴力匹配就是通过搜索所有可能匹配点来进行最佳匹配的寻找，该算法的计算效率较低，但总能找到最佳的匹配。

本文主要基于汉明距离对匹配进行最佳匹配的筛选，计算每个匹配点对之间的汉明距离，分别计算出距离最大的点对之间的距离和距离最小的点对之间的距离。当描述子的距离大于两倍的最小距离是即认为匹配有误，通常点对之间汉明距离的最小值可能会很小，所以本文选择 30 作为经验值。

基于 4.2.2 部分特征点检测结果，本文的特征点匹配结果如图 4-7 所示。

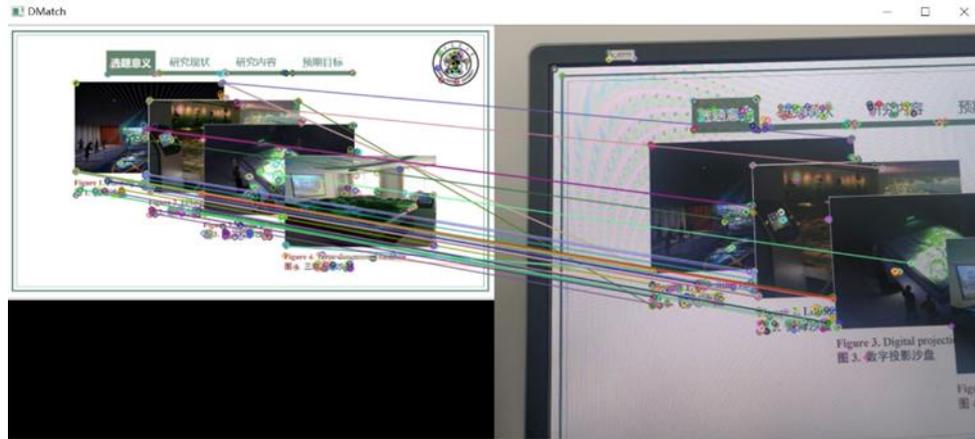


图 4-7 暴力匹配结果图

Figure 4-7 violence matching results

#### 4.2.4 错误匹配剔除

经过暴力匹配对图像特征点进行匹配后，可以根据匹配点之间的关系计算点对之间的变换矩阵  $H$ 。取出训练图像的角点坐标，通过变换矩阵  $H$  对训练图像的四个角点坐标进行透视转换，转换完毕的角点为在合成图像上的绘制的矩形的角点。

在变换矩阵  $H$  的计算阶段，本文首先通过对暴力匹配得到的匹配点对进行了  $H$  矩阵的计算。屏幕配准的结果如图 4-7 所示。该计算结果耗时 10.2962ms。

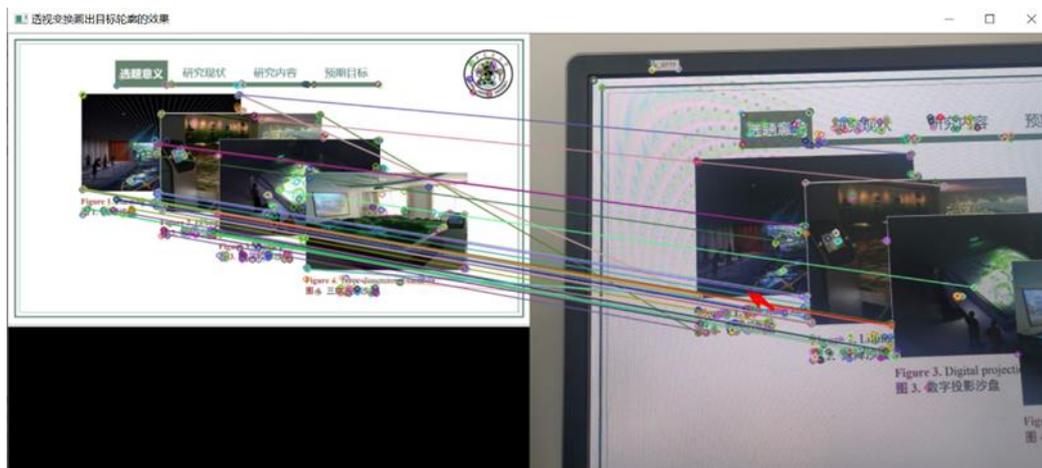


图 4-8 利用所有点进行屏幕配准结果

Figure 4-8 screen registration results using all points

通过图 4-8 所示结果可知，直接利用暴力匹配得到的特征点对进行屏幕配准，基本无法完成，根据对图 4-7 进行分析，发现暴力匹配之后存在错误匹配的特征点对，后续本文考虑了剔除错误匹配的特征点对。首先采用 LMedS 算法进行错误匹

配点对的剔除。LMedS 算法当错误匹配点占全部点对的比例达到 50%依然可以准确预测模型。已知匹配点对的集合  $\{X_i | X_i = \{(x_i, y_i), (x'_i, y'_i)\}, i = 1, 2, \dots, N\}$ , 其中  $(x_i, y_i)$  表示训练图像特征点坐标,  $(x'_i, y'_i)$  表示待配准图像特征带你坐标, 这种算法求取变换模型的流程:

(1) 样本子集的选取。从样本集中随机选取四个数获得样本子集  $S = (i_1, i_2, i_3, i_4)$ 。

(2) 通过样本子集  $S$  得到特征点子集  $\{X_i | X_i = \{(x_i, y_i), (x'_i, y'_i)\}, i \in S\}$ , 对该特征点子集根据最小二乘法计算的粗投影变换矩阵  $H$ 。

(3) 计算平方偏差中值:

$$r_s^2 = \text{median}(\|(x'_i, y'_i) - H(x_i, y_i)\|^2) \quad (4-3)$$

其中  $i=1, 2, \dots, N$ 。

(4) 计算出重复  $m$  次(3)计算中的最小平方值  $r_{jLMedS}^2 = \min(r_{Si}^2), i = 1, 2, \dots, m$ ,  $r_{jLMedS}^2$  的矩阵参数即为投影变换矩阵, 并通过该矩阵做错误匹配点的剔除。

通过 LMedS 算法对图 4-7 进行透视变换矩阵计算并得出的屏幕配准结果如图 4-9 所示。该方法得出配准结果耗时 30.1569ms。通过对图 4-8 的观察, 该算法较好的实现了屏幕校准功能。当错误匹配点数占全部匹配点数的比例较低时, LMedS 可以做到错误匹配点的完全剔除, 并得出最优的投影变换矩阵。因为 LMedS 记录的是中值偏差, 当错误匹配点占全部匹配点的比例超过 50%之后, LMedS 算法将失效。

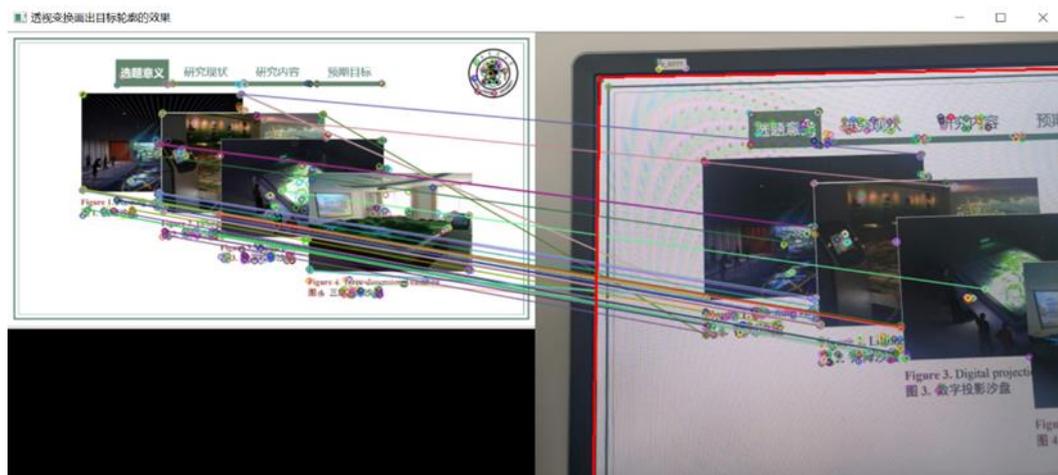


图 4-9 LMedS 算法屏幕校准结果

Figure 4-9 screen calibration results of LMedS algorithm

通过与本文尝试的第一种方法对比, 该方法对然能够得到准确的屏幕配准结果但是时延过大, 本文系设计统的初衷是带给用户灵活的交互, 所以时延问题需要

重点关注。所以本文对 RANSAC 算法进行了尝试，该算法消除了错误匹配比例带来的限制，在当错误匹配点超过全部匹配点的一半时依旧可以准确的实现匹配点的提纯。该算法的实现步骤为：

- (1) 设置最佳估计内点数目  $N_i$  为 0，随机的从所有匹配点中抽出 4 组匹配点，其中任意 3 组不共线，据此计算出变换矩阵 H。
- (2) 计算剩下没被抽取的匹配点经过变换矩阵的坐标值，并计算其与预匹配坐标之间的差值。若这个差值小于预设阈值，则将其加入正确匹配点集。
- (3) 将当前正确匹配点击数目与最优匹配集进行比较，若大于  $N_i$  则当前转换关系 H 即为最佳矩阵估计，并对  $N_i$  的值进行估计。
- (4) 经过不大于 k 次的迭代次数后，选择正确匹配点数最多的变换矩阵参数作为图像间的变换矩阵，最终利用该矩阵做是否是正确匹配点的判断。

通过 RANSAC 算法对图 4-7 进行透视变换矩阵计算并得出的屏幕校准结果如图 4-10 所示。

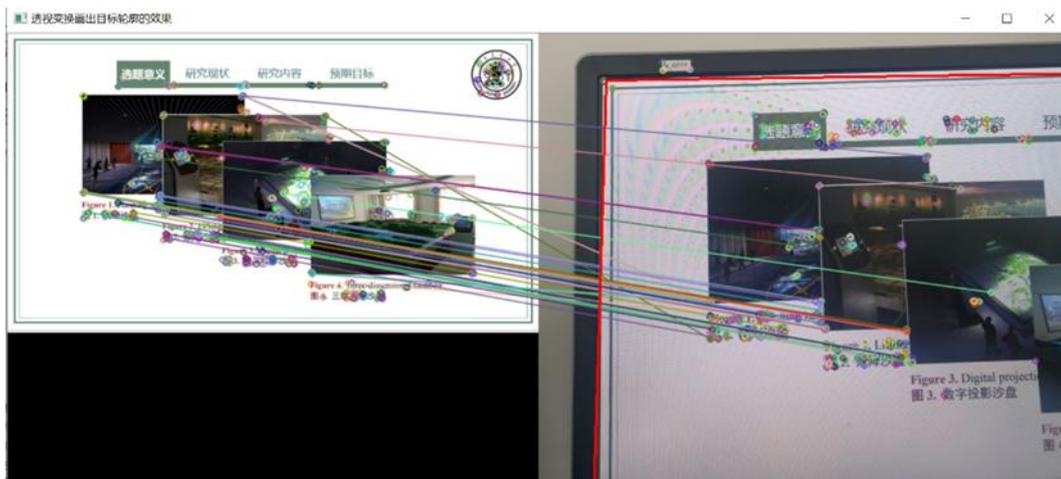


图 4-10 RANSAC 算法屏幕校准结果

Figure 4-10 RANSAC algorithm screen calibration results

通过对图 4-8、图 4-9、图 4-10 的屏幕校准结果分析，并对各个算法的耗时进行了计算。耗时结果如表 4-1 所示。

表 4-1 三种算法时延比较

Table 4-1 delay comparison of three algorithms

方式	用时 (ms)
不剔除	10.2962
LMedS	30.1569
RANSAC	4.6604

根据表 4-1 所示结果，本文选择 RANSC 算法进行变换矩阵的计算，该算法能

够在时延较小的情况下实现屏幕的准确矫正。

## 4.3 光斑识别

### 4.3.1 视频采集

本系统计划采用单个主机同时对多个摄像头的信息进行处理，在此视频采集的过程中计划采用多线程技术以实现多摄像头的循环取帧，这样使得各个摄像头获取的图像信息准确而且路径不发生冲突。其中多线程技术主要是指在 windows 操作系统下，每个线程被分配了不同的时间片，在某个时刻，CPU 仅执行一个时间片内的线程，多个时间片中的相应线程在 CPU 内轮流执行，由于每个时间片的时间很短，所以对用户来说仿佛各个线程在计算机中并行处理一样。基本流程如图 4-11 所示。

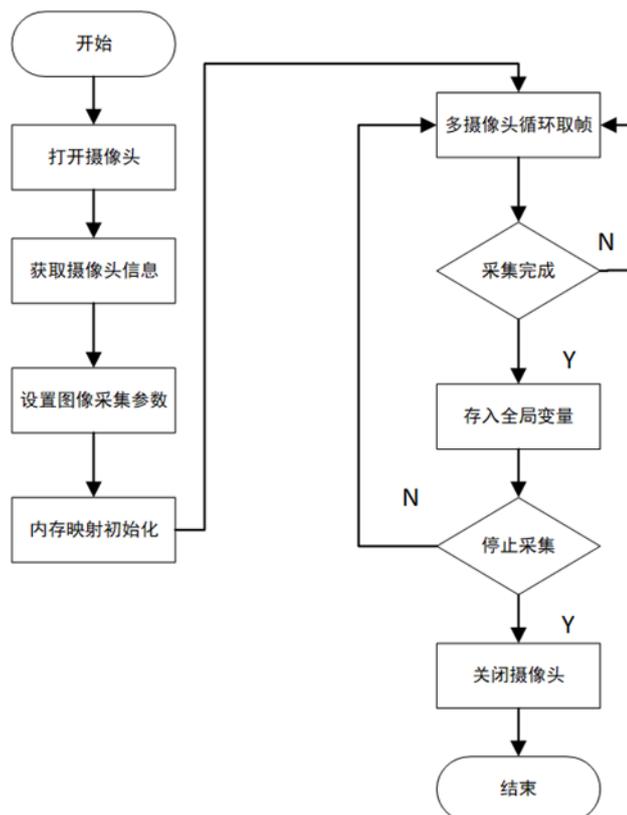


图 4-11 视频采集基本流程

Figure 4-11 basic flow of video acquisition

### 4.3.2 激光光斑定位

本模块的主要任务是从摄像头采集到的图像中检测出激光笔发射的光斑。由

于激光笔发射的光斑占整个摄像头拍摄区域非常小，其模式特征不明显。利用模式特征进行激光光斑的识别与跟踪得到效果可能会不是十分理想。激光光斑在屏幕上显示的亮度值非常大，在较暗的环境中，屏幕上显示为一个亮度很高的红色亮斑，但是当周围环境光线很亮的时候，激光光斑很容易就被周围光线盖住。为了解决这个问题，本文在选择硬件设备的时候在摄像头前面加了一个滤光片，滤光片的作用是使红外光经过，而其他波长的光线被过滤掉。这样处理使得不论外部光线如何变化，摄像头都可以捕捉到清晰的激光光斑。

对于实现对光斑等亮度较高的小目标跟踪，目前已经有了很多研究，例如基于顶帽运算与 PNLOG 算子的激光光斑检测算法以及采用尺度自适应图像分割与各方向灰度变化相结合的方式实现光斑的跟踪检测。本文使用基于亮点灰度值的检测算法进行光斑的检测方案，能够保证效果良好和实时性。此方案流程图如图 4-12 所示。

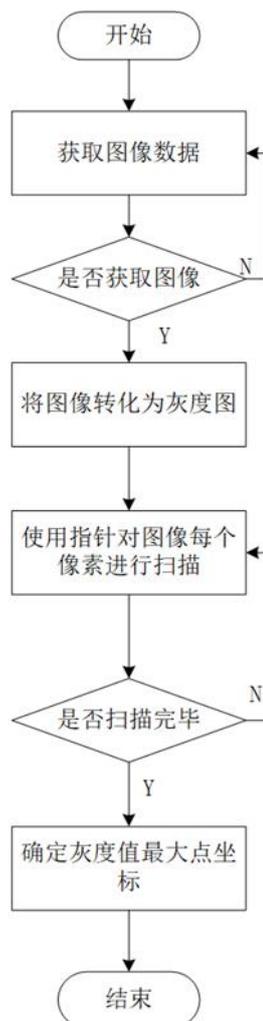


图 4-12 激光光斑定位流程图

Figure4-12 flow chart of laser spot location

激光光斑定位结果如图 4-13 所示。

```

选择Microsoft Visual Studio 调试控制台
[ INFO:0] global C:\build\master_winpack-build-win64-vc15\opencv\modules\videoio\src\videoio_registry.cpp (187) cv::anonymous-namespace::VideoBackendRegistry::VideoBackendRegistry VIDEOIO: Enabled backends(7, sorted by priority): FFMPEG(1000); GSTREAMER(990); INTEL_MFX(980); MSMF(970); DSHOW(960); CV_IMAGES(950); CV_MJPEG(940)
[ INFO:0] global C:\build\master_winpack-build-win64-vc15\opencv\modules\videoio\src\backend_plugin.cpp (340) cv::impl::getPluginCandidates Found 2 plugin(s) for GSTREAMER
[ INFO:0] global C:\build\master_winpack-build-win64-vc15\opencv\modules\videoio\src\backend_plugin.cpp (172) cv::impl::DynamicLib::libraryLoad load C:\WINDOWS\SYSTEM32\opencv_videoio_gstreamer411_64.dll => FAILED
[ INFO:0] global C:\build\master_winpack-build-win64-vc15\opencv\modules\videoio\src\backend_plugin.cpp (172) cv::impl::DynamicLib::libraryLoad load opencv_videoio_gstreamer411_64.dll => FAILED
(30, 144)
(32, 134)
(34, 124)
(36, 120)
(38, 120)
(64, 106)
(66, 102)
(68, 102)
(70, 104)
(74, 106)
(76, 100)
(78, 98)
(80, 104)
(90, 108)
(92, 106)
(94, 106)
(96, 100)
(98, 102)
(100, 102)
(41, 55)

```

图 4-13 激光光斑定位结果

Figure 4-13 laser spot location results

## 4.4 坐标转换与鼠标追踪

### 4.4.1 坐标转换

由于受到空间条件的限制，摄像头的位置不可避免使得屏幕上显示的图像发生一定地倾斜。此使采集到的激光笔光点坐标也会跟随这个倾斜发生一定角度的旋转。这样会影响到鼠标功能的准确度，进而降低用户的使用体验。这种倾斜问题导致的问题有坐标原点的差异和坐标轴不平行的差异。所以得到的坐标只是摄像区域的坐标值，而不是计算机屏幕的逻辑坐标。其中  $X-O-Y$  坐标轴区域是摄像区域坐标， $X'-O-Y'$  坐标轴区域是计算机屏幕逻辑坐标。这两种坐标轴之间的几何关系如图 4-14 所示。

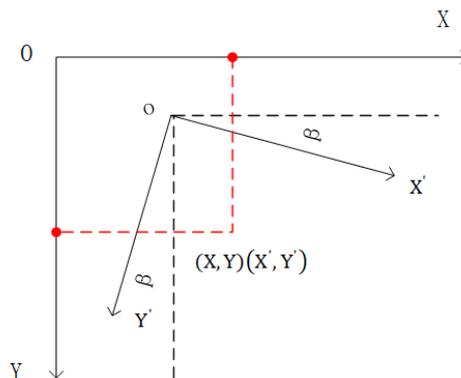


图 4-14 坐标转换几何关系

Figure 4-14 geometric relationship of coordinate conversion

其中  $(X, Y)$  表示在  $X-O-Y$  坐标轴中的坐标， $(X', Y')$  表示在  $X'-o-Y'$  坐标中的坐标， $\beta$  表示两个坐标轴之间存在的角度，坐标  $(a, b)$  为  $o$  点在  $X-O-Y$  坐标中的坐

标。由屏幕配准的模块可以得出  $o$  点相对于  $O$  点的坐标，由图上几何关系可以进行两个坐标轴之间的坐标转换，由几何关系推出的转换关系为：

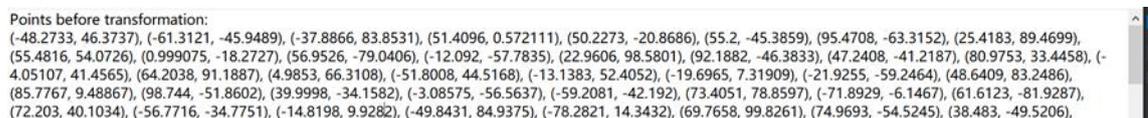
$$X' = (X - a) \cos \beta + (Y - b) \sin \beta \quad (4-4)$$

$$Y' = (Y - b) \cos \beta - (X - a) \sin \beta \quad (4-5)$$

因此，可将得到的倾斜矫正前的光点坐标  $(X, Y)$  代入上述公式就可以得到矫正后的光点坐标  $(X', Y')$ 。

该部分将 RANSAC 算法计算出来的屏幕角点坐标作为输入，对该坐标进行反三角函数计算即可获得坐标系之间的偏转角度。RANSAC 算法计算出来的坐标为  $(74.99, 52.08)$ ，偏转角度为  $48.58^\circ$ 。坐标转换部分的时延是 88ms。

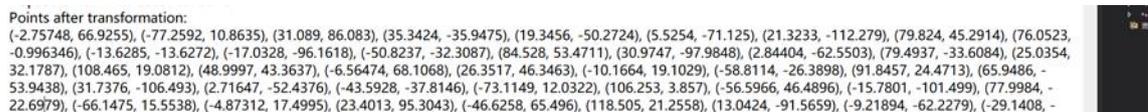
通过以该点为参考点进行坐标转换。坐标转换结果如图 4-14、图 4-15 所示。



```
Points before transformation:
(-48.2733, 46.3737), (-61.3121, -45.9489), (-37.8866, 83.8531), (51.4096, 0.572111), (50.2273, -20.8686), (55.2, -45.3859), (95.4708, -63.3152), (25.4183, 89.4699), (55.4816, 54.0726), (0.999075, -18.2727), (56.9526, -79.0406), (-12.092, -57.7835), (22.9606, 98.5801), (92.1882, -46.3833), (47.2408, -41.2187), (80.9753, 33.4458), (-4.05107, 41.4565), (64.2038, 91.1887), (4.9853, 66.3108), (-51.8008, 44.5168), (-13.1383, 52.4052), (-19.6965, 7.31909), (-21.9255, -59.2464), (48.6409, 83.2486), (85.7767, 9.48867), (98.744, -51.8602), (39.9998, -34.1582), (-3.08575, -56.5637), (-59.2081, -42.192), (73.4051, 78.8597), (-71.8929, -6.1467), (61.6123, -81.9287), (72.203, 40.1034), (-56.7716, -34.7751), (-14.8198, 9.9282), (-49.8431, 84.9375), (-78.2821, 14.3432), (69.7658, 99.8261), (74.9693, -54.5245), (38.483, -49.5206),
```

图 4-14 转换前坐标

Figure 4-14 coordinates before transformation



```
Points after transformation:
(-2.75748, 66.9255), (-77.2592, 10.8635), (31.089, 86.083), (35.3424, -35.9475), (19.3456, -50.2724), (5.5254, -71.125), (21.3233, -112.279), (79.824, 45.2914), (76.0523, -0.996346), (-13.6285, -13.6272), (-17.0328, -96.1618), (-50.8237, -32.3087), (84.528, 53.4711), (30.9747, -97.9848), (2.84404, -62.5503), (79.4937, -33.6084), (25.0354, 32.1787), (108.465, 19.0812), (48.9997, 43.3637), (-6.56474, 68.1068), (26.3517, 46.3463), (-10.1664, 19.1029), (-58.8114, -26.3898), (91.8457, 24.4713), (65.9486, -53.9438), (31.7376, -106.493), (2.71647, -52.4376), (-43.5928, -37.8146), (-73.1149, 12.0322), (106.253, 3.857), (-56.5966, 46.4896), (-15.7801, -101.499), (77.9984, -22.6979), (-66.1475, 15.5538), (-4.87312, 17.4995), (23.4013, 95.3043), (-46.6258, 65.496), (118.505, 21.2558), (13.0424, -91.5659), (-9.21894, -62.2279), (-29.1408, -
```

图 4-15 转换后坐标

Figure 4-15 coordinates after transformation

#### 4.4.2 鼠标追踪

本节主要用于实现基于红外激光笔的电子沙盘鼠标操控核心功能。经过 4.2 节的介绍获得了屏幕在摄像头拍摄视野中的位置，4.3 部分获得激光光斑在摄像头拍摄视野中的位置，根据 4.4.1 部分的几何变换关系得到了激光光斑在屏幕中的准确位置。本文研究系统主要实现电子沙盘部分的鼠标操控功能，其实现流程如图 4-16 所示。本部分首先获取鼠标得权限，并根据光斑的准确位置移动鼠标位置，在后续过程中监听鼠标的移动和按键的动作消息。

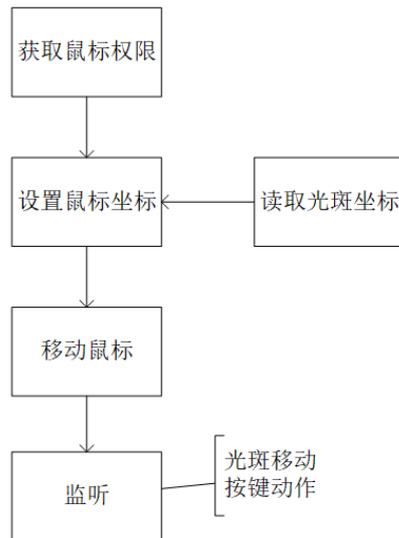


图 4-16 鼠标控制功能实现流程

Figure 4-16 implementation process of mouse control function

本部分主要通过预先设置鼠标图案实现电子沙盘部分的光斑醒目化指示，鼠标效果如图 4-17 所示。



图 4-17 鼠标效果图

Figure 4-17 mouse effect

## 4.5 本章小结

本章对系统的软件实现部分进行了具体描述，第一部分对软件部分需要解决的问题进行了总结，并对软件部分各个模块进行了具体的设计；第二部分对屏幕校准模块进行了具体实现，主要包括特征点提取、特征点匹配以及错误匹配点剔除桑格步骤；第三部分主要介绍了光斑位置的坐标转换和鼠标的追踪功能的实现流程。

## 5 系统评估

本章对上文提出的基于激光光斑的智能沙盘控制系统进行了实验评估。首先介绍了评估的环境和评估的标准，之后对评估的实验结果进行了分析，评估了系统实现屏幕自动配准的准确率和光斑识别的准确率，并对整体时延进行了分析。

### 5.1 评估方法

#### 5.1.1 软硬件环境

本文在同一台电脑上进行了屏幕的自动配准与激光光斑的识别与控制工作。软硬件环境如下：

处理器：Intel(R)Core(TM)i7-8550U CPU@1.80GHZ 2.00GHZ

内存(RAM):16GB

系统类型：64 位 windows 10 操作系统

网络环境：校园网

其余硬件部分激光光斑发射器、无线接收器、摄像头、滤光片见 3.4.2 部分。

#### 5.1.2 评估标准

##### (1) 系统的定位精度

系统的定位精度具体是指用户期望操作的具体位置 and 实际光标之间的差距，导致系统定位精度存在偏差的主要原因来自两个方面，一个是信息输入的准确性，另一个是计算机程序计算的准确性。针对第一个方面，比如对于屏幕来说，可见光的干扰就是一个不可避免的问题，我们需要谨慎的选择合适的硬件设备进行信息的收集和输入。对于第二个方面，则需要对收集到的信息进行合理分析，并根据其特征进行程序的设计。有许多外在因素的不可避免性使得误差不可能完全被消除，我们需要做的就是尽量合理化的进行该系统各方面的设计并尽最大可能将误差控制在可以接受地范围内。

##### (2) 系统的响应速度

系统的响应速度主要是指系统光标速度跟随交互设备运行的速度。响应时间是演示系统一个非常重要的因素。较小的响应时间可以带给讲述者清晰的思路和流畅的表达，而迟缓的使用则将会出现分散使用者注意力等不利因素，在一定程度上大大降低了使用者的使用体验。这一问题主要来源于两个方面，一个是硬件方面，

其收到硬件信息采集能力的限制。另一方面是受到软件处理流程的限制，这一方的限制主要和算法的复杂度有关，一次性处理的数据量也会对光标的跟随速度产生一定的影响。此外，系统的环境对于光标跟随时间也存在一定限制。

### (3) 系统的兼容性

基于本系统需要运行在各种日常的应用场景中，所以本系统需要在各种 Windows 环境中运行正常，并具有良好的兼容性。

### (4) 系统的稳定安全

演示系统的软件部分能够实现稳定安全的运行，如果系统经常崩溃需要各种维修，那么将失去研究与实现的意义。

## 5.2 评估结果分析

通过摄像头对屏幕信息进行采集，受到摄像头摆放的空间位置限制，摄像头采集到的屏幕并不能和实际中的屏幕完全重合，所以本模块需要实现屏幕的校准。本部分主要采用 ORB 算法对屏幕的特征进行提取，第二步尝试采用暴力匹配算法和 FLANN 算法对采集到的特征点进行匹配，第三步尝试使用 RANSAC 算法和 LMedS 算法对错误匹配的特征点进行剔除，最后一步对单应性矩阵进行计算实现屏幕的校准。

本部分利用 ORB 算法进行训练图像与屏幕图像的特征点提取，考虑到真实系统的应用场景，本文选择了不同的图片进行了屏幕配准测试。由于本文实验环境的限制，本文选择笔记本电脑屏幕进行算法的测试，其大小为 14 英寸。首先对容器进行初始化，使用 `ORB::create()` 函数进行关键点得创建并形成描述符，该函数中支持得参数很多，一般只对前两个参数进行设置，后面参数保持默认值即可获得较好的结果，其中第一个参数 `nfeatures` 表示最多可以采集到的特征点数目，这里设置为 500，第二个参数为 `scaleFactor` 表示金字塔抽取比，也就是金字塔图像之间的尺度参数，这里设置为 1.2f。其中特征点匹配结果如图 5-1 所示。

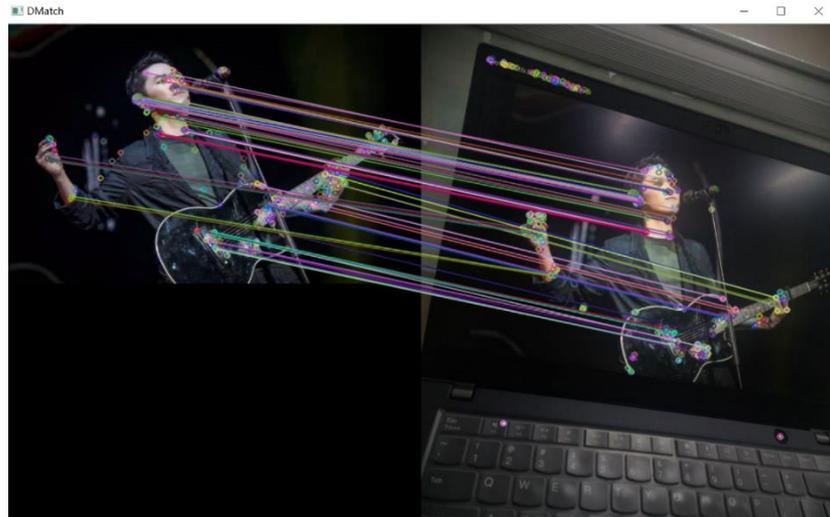


图 5-1 ORB 特征点匹配结果图

Figure 5-1 matching results of orb feature points

根据第四章的分析结果，采用 RANSAC 算法对透视变换矩阵进行计算，将点对视为内点的最大允许重投影错误阈值设为 3。可置信度设置为 0.995。经过透射变换矩阵进行变换后，计算出的屏幕角点坐标如图 5-2 所示。

```
透视矩阵变换后 在合成的匹配图像中绘制的标记目标的矩形的四个角点坐标为
115.94106, 70.00248;
565.44409, 211.67102;
501.13562, 473.54492;
55.445408, 407.823]
```

图 5-2 屏幕角点坐标

Figure 5-2 screen corner coordinates

根据该角点坐标实现的屏幕校准结果如图 5-3 所示。

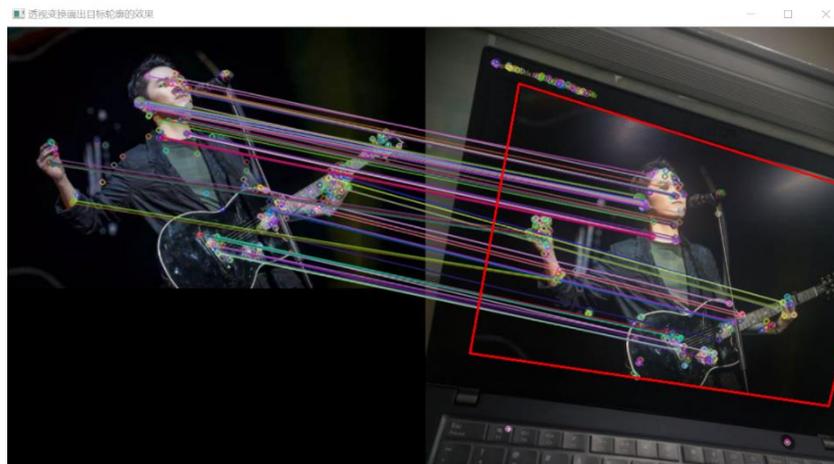


图 5-3 屏幕校准结果

Figure 5-3 screen calibration results

利用激光光斑对大屏幕进行交互主要指的是利用激光笔代替鼠标完成远程对计算机的控制，主要包括鼠标的左击、右击、双击和移动等。在交互的过程中，鼠标的操作可能会使大屏幕背景发生变化，滤光片的应用效果较好的解决了背景变化带来的光斑识别问题。但是当光斑打在屏幕校准生成的红框上时，将无法实现光斑的定位，所以会对光斑识别的准确率造成影响，本文定义光斑识别的错误概率为屏幕校准边框像素点的个数除整个屏幕的像素点个数，本文系统计算出的光斑识别准确率为 89%，该结果证明光斑定位的准确率较高。

根据上述时延计算，计算出本系统实现全部功能所需的时延如表 5-1 所示。

表 5-1 系统时延分析

Table 5-1 system delay analysis

	用时 (ms)
自动配置	1316.32
光斑识别	0.0674
坐标转换	88
总时延	1404.3874

根据上表分析，本系统的总时延在 1.4 秒左右，而且大部分的系统时延来自屏幕的自动配准过程，光斑识别与坐标转换部分的时延较小，基本上可以满足系统的实时性交互要求。

### 5.3 本章小结

本章对第四章实现系统进行了测试与评估。对屏幕的自动校准进行了验证，也对光斑识别的准确率进行了计算。通过对整体系统的时延分析，证明当在系统进行初始化配置的时候时延略大，但并不影响用户的实时交互。整体评估结果表明，通过激光光斑对智能沙盘系统进行交互控制时可行的。

## 6 结论与展望

本章对本文工作进行了总结，并且展望了未来的工作内容和仍需进一步研究的问题。

### 6.1 本文总结

本文主要对基于实体沙盘的智能沙盘系统进行了设计，并主要针对电子沙盘的交互问题进行了具体研究。根据目前物理沙盘和电子沙盘存在的主要问题，本文设计系统解决了物理沙盘和实体沙盘的操控问题，实现了基于红外激光笔的电子沙盘鼠标操控核心功能，解决了电子沙盘的灵活交互问题。本文讨论了所设计系统的可行性，实验证实了核心功能的可用性。本文具体工作为：

(1) 设计了一套基于实体沙盘和大屏幕展示相结合的智能沙盘系统，该系统包括屏幕自动配置模块、光斑识别模块、坐标转换与光标追踪模块，实现了屏幕的自动化配置并利用激光光斑实现了相对便捷的人机交互。

(2) 实现了一种基于图像识别的屏幕自动配置功能，以匹配电子显示屏角点坐标和摄像头坐标。该自动配置功能不严格约束摄像头与屏幕的空间摆放位置，而是通过分析摄像头图片与屏幕显示图片的特征点对应关系自动校准摄像头对应的屏幕区域。

(3) 实现了基于红外激光笔的虚拟鼠标操控功能，以实现灵活方便的大屏幕远程操控。该操控功能技术流程包括，红外光斑提取、光斑的坐标转换（从摄像头坐标到屏幕坐标）、虚拟鼠标操控。

### 6.2 工作展望

智能沙盘的应用领域非常广泛。随着人机交互技术的不断发展，人们对交互的便捷性与灵活性的要求越来越高，智能沙盘系统还存在很大的优化空间。未来可以从以下角度进行系统的优化设计：

- (1) 完成系统设计需求的硬件设备应适当减少。
- (2) 实现实体沙盘和电子沙盘的联动互控。

## 参考文献

- [1] 罗琪斯, 罗雅丹. 一种实体沙盘与电子沙盘双向互控系统研究[J]. 测绘与空间地理信息, 2020, 43(08):26-30.
- [2] 沈颖. 基于 VRMap 平台的电子沙盘研制[D]. 长安大学, 2005.
- [3] Z. Zhou, Z. Bian and Z. Zhuo, "MR sand table: Mixing real-time video streaming in physical models," 2017 IEEE Virtual Reality (VR), Los Angeles, CA, 2017, pp. 239-240, doi: 10.1109/VR.2017.7892265.
- [4] 程忠波. 浅析数字沙盘的设计与应用[J]. 电子制作, 2014(24):68.
- [5] 严一鸣, 郭星. 基于计算机视觉的交互式电子沙盘系统研究[J]. 计算机技术与发展, 2017, 27(06):195-198.
- [6] 姚世明. 增强现实电子沙盘协同交互技术研究与应用[D]. 南京理工大学, 2019.
- [7] 马文学, 刘华伟. 可视化智能交互电子沙盘在公安一体化指挥调度平台中的应用[A]. 中国通信学会. 2019 年全国公共安全通信学术研讨会优秀论文集[C]. 中国通信学会:中国通信学会, 2019:4.
- [8] 刁学磊. 智能沙盘人机交互系统研发[D]. 哈尔滨工业大学, 2020.
- [9] 李志. 基于图像识别的大屏幕人机交互系统研究[D]. 安徽大学, 2010.
- [10] Thomas Zimmerman, Jaron Lanier, Chuck Blanchard, Steve Bryson, Young Harvill. A hand gesture interface device[P]. GI + CHI 1987, 1987.
- [11] 王万良, 杨经纬, 蒋一波. 基于运动传感器的手势识别[J]. 传感技术学报, 2011, 24(12):1723-1727.
- [12] Charayaphan C, Marble A E. Image processing system for interpreting motion in American Sign Language. [J]. Journal of biomedical engineering, 1992, 14(5).
- [13] Ying W , Ying S , Zhang Z , et al. VISUAL PANEL: From an ordinary paper to a wireless and mobile input device[J]. Microsoft Research, 2000.
- [14] Desale R . A Study on Wearable Gestural Interface - A SixthSense Technology[J]. IOSR Journal of Computer Engineering, 2013, 10(5):10-16.
- [15] Yeung Ling-Fung, Yang Zhenqun, Cheng Kenneth Chik-Chi, Du Dan, Tong Raymond Kai-Yu. Effects of camera viewing angles on tracking kinematic gait patterns using Azure Kinect, Kinect v2 and Orbbec Astra Pro v2[J]. Gait & Posture, 2021, 87.

- [16] 王辉. 基于视觉的实时手势跟踪与识别及其在人机交互中的应用[D]. 浙江大学, 2008.
- [17] 时旭. 人机交互中视觉目标识别与跟踪算法的研究[D]. 哈尔滨工程大学, 2010.
- [18] 朱杰. 触摸屏技术的交互式电子沙盘开发与应用[J]. 测绘科学技术学报, 2011, 28(04):304-306+312.
- [19] 郗瑶颖. 基于立体视觉的指尖定位与人机交互技术研究[D]. 长安大学, 2013.
- [20] 张玉军, 孟晓军, 王刚. 基于手势交互的三维电子沙盘系统设计与实现[J]. 指挥控制与仿真, 2016, 38(02):110-114.
- [21] 夏晗. 基于视觉定位的人机交互示教系统[D]. 哈尔滨工业大学, 2017.
- [22] Liu Hai, Nie Hanwen, Zhang Zhaoli, Li You-Fu. Anisotropic angle distribution learning for head pose estimation and attention understanding in human-computer interaction[J]. Neurocomputing, 2021, 433.
- [23] TAO Jianhua. Emotion recognition for human-computer interaction[J]. Virtual Reality & Intelligent Hardware, 2021, 3(1).
- [24] Fan Yongkui, Wang Yuanyuan, Zhu Zhongxiang, Lv Meng, Mao Enrong, Zhai Zhiqiang. Design of Human-Computer Interaction System of Electro-hydraulic Hitch for Hillside Tractors[J]. Journal of Physics: Conference Series, 2021, 1732(1).
- [25] Wenwen Jiang, Lei Chen. Design of Human-Computer Interaction Teaching System Based on Virtual Reality Technology[J]. Frontiers in Educational Research, 2020, 3(11).
- [26] Yi Feng Zhang, Gang Chen, Min Juan Wang, Muhammad Jahanzaib Khan. A Bluetooth-Based Laser Pen Instruction-Aid System for Face-to-Face and Mobile Teaching[J]. Applied Mechanics and Materials, 2012, 1945.
- [27] Xueya, Song, Li, Desheng. Development of Wireless Mobile Video Surveillance on Windows Mobile Using DirectShow Technology[J]. Journal of Computational and Theoretical Nanoscience, 2017, 14(7).
- [28] Jian Bing Xiahou, Zhen Xiong Wang. The Apply of DirectShow in Scenario Interactive Teaching System[J]. Advanced Materials Research, 2014, 3181.
- [29] Xin Chen, Xiang Rong Wu. A Courseware Recording & Editing System Based on DirectShow[J]. Applied Mechanics and Materials, 2013, 2346.
- [30] 张毅. 基于 ORB 特征的图像检索技术研究[D]. 中国矿业大学, 2020.
- [31] Jianfang Dou, Qin Qin, Zimei Tu. Robust image matching based on the

- information of SIFT[J]. *Optik*, 2018, 171.
- [32] 易飞, 许珊珊. 基于 SIFT 的图像匹配方法改进[J]. *计算机系统应用*, 2018, 27(10): 261-267.
- [33] 王嫣然. 视频的美学和有趣性计算角度[D]. 复旦大学, 2014.
- [34] Parashivamurthy Ravi, Naveena Chikkaguddaiah, Sharath Kumar Yeliyur Hanumathiah. SIFT and HOG features for the retrieval of ancient Kannada epigraphs[J]. *IET Image Processing*, 2021, 14(17).
- [35] Bassoma Diallo, Jie Hu, Tianrui Li, Ghufraan Ahmad Khan, Ahmed Saad Hussein. Multi-view document clustering based on geometrical similarity measurement[J]. *International Journal of Machine Learning and Cybernetics*, 2021 (prepublish).
- [36] Wu Hao, Cao Yu, Wei Haiping, Tian Zhuang. Face Recognition Based on Haar Like and Euclidean Distance[J]. *Journal of Physics: Conference Series*, 2021, 1813(1).
- [37] Saikat Singha Roy, Soma Barman. A non-invasive cancer gene detection technique using FLANN based adaptive filter[J]. *Microsystem Technologies*, 2021, 27(2).
- [38] 邓金雪. 可编程 DMX 设备的控制设计和实现[D]. 北京: 北京交通大学, 2019.

## 作者简历及攻读硕士/博士学位期间取得的研究成果

### 一、作者简历

赵映南，女，1998年7月出生，本科就读于燕山大学信息科学与工程学院，2019年获工学学士学位，硕士就读于北京交通大学电子信息工程学院，2021年获专业硕士学位。研究生就读期间主要从事信息系统方向的研究工作。

·  
·

## 独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作和取得的研究成果，除了文中特别加以标注和致谢之处外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京交通大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：赵映南 签字日期：2021年6月1日

## 学位论文数据集

表 1.1: 数据集页

关键词*	密级*	中图分类号	UDC	论文资助
智能沙盘系统; 屏幕自动化配 置;光斑识别;坐 标转换	公开			
学位授予单位名称*	学位授予单位代码*		学位类别*	学位级别*
北京交通大学	10004		工程	硕士
论文题名*	并列题名			论文语种*
智能沙盘中的自动配置与光斑控 制研究				中文
作者姓名*	赵映南		学号*	19125078
培养单位名称*	培养单位代码*		培养单位地址	邮编
北京交通大学	10004		北京市海淀区西直 门外上园村 3 号	100044
工程领域*	研究方向*		学制*	学位授予年*
电子与通信工程	信息系统		2	2021
论文提交日期*	2021 年 6 月			
导师姓名*	李纯喜		职称*	副教授
评阅人	答辩委员会主席*		答辩委员会成员	
	孙强			
电子版论文提交格式 文本 ( ) 图像 ( ) 视频 ( ) 音频 ( ) 多媒体 ( ) 其他 ( ) 推荐格式: application/msword; application/pdf				
电子版论文出版 (发布) 者		电子版论文出版 (发布) 地		权限声明
论文总页数*	52			
共 33 项, 其中带*为必填数据, 为 21 项。				