****

工程硕士学位论文

**基于FPGA的手势识别跟踪系统算法架构研究与实现**

|  |  |
| --- | --- |
| 作者姓名 | 李叙琼 |
| 工程领域 | 电子与通信工程 |
| 校内指导教师 | 殷瑞祥 |
| 校外指导教师 | 李辉宪 |
| 所在学院 | 电子与信息学院 |
| 论文提交日期 |  |

**Algorithm architecture research and implementation of gesture recognition and tracking system based on FPGA**

A Dissertation Submitted for the Degree of Master

**Candidate：Xuqiong Li**

**Supervisor：Prof. RuiXiang Yin**

South China University of Technology

Guangzhou, China

**分类号： 学校代号：10561**

**学 号：**

华南理工大学硕士学位论文

**基于FPGA的手势识别跟踪系统算法架构研究与实现**

作者姓名： 申请学位级别：工程硕士

工程领域名称：

校内指导教师姓名、职称： 校外指导教师姓名、职称：

论文形式：ꇶ 产品研发 ꇶ 工程设计 ꇶ 应用研究 ꇶ 工程/项目管理 ꇶ 调研报告

研究方向：

论文提交日期： 年 月 日 论文答辩日期： 年 月 日

学位授予单位：华南理工大学 学位授予日期： 年 月 日

答辩委员会成员：

主席：

委员：

**华南理工大学**

**学位论文原创性声明**

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

作者签名： 日期： 年 月 日

**学位论文版权使用授权书**

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属华南理工大学。学校有权保存并向国家有关部门或机构送交论文的复印件和电子版，允许学位论文被查阅（除在保密期内的保密论文外）；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。本人电子文档的内容和纸质论文的内容相一致。

本学位论文属于：

□保密，在 年解密后适用本授权书。

□不保密,同意在校园网上发布，供校内师生和与学校有共享协议的单位浏览；同意将本人学位论文提交中国学术期刊(光盘版)电子杂志社全文出版和编入CNKI《中国知识资源总库》，传播学位论文的全部或部分内容。

(请在以上相应方框内打“√”)

作者签名： 日期：

指导教师签名： 日期

作者联系电话： 电子邮箱：

联系地址(含邮编)：

摘 要

随着互联网技术与硬件终端设备的发展，多媒体技术和人机交互在尽量来得到了充分的应用。而传统意义的人机交互输入设备类似于键盘、鼠标已经无法满足多媒体高度发展的多元化需求。近年来，随着移动终端的普及，手机、平板电脑、可穿戴式设备等硬件平台辅以大量传感器的应用，构成了狭义上的“物联网”，丰富了人与机器的交互模式，也提升了用户体验。

从按压按钮实现对机器单一的控制，到用矩阵键盘实现的组合式控制，再到后来用触摸屏实现更加良好的用户体验。人机交互方式逐渐从功能的控制转移到以人为中心的人机一体互动模式。近年来，不依赖于传统硬件设备的人机交互方式已经开始商用，类似于微软的Project Natal，索尼的Play Station都给了用户前所未有的体感体验。利用人类的语音、手势、动作、姿势作为输入，以高速、并行、非精确的控制方式进行交互式新一代自然交互结束的核心。近年来基于机器视觉和神经网络的人机交互方式有了长足的发展。基于深度摄像头的非接触式人机交互方式在国内也初步应用于智能平板电视，但从技术层面上说仍然处于起步阶段，在复杂背景下、在光线不足的情况下仍存在着技术瓶颈。

本论文通过研究复杂背景下的人机交互算法，针对算法中的大数据量、高吞吐量等技术瓶颈问题，提出一套优化了实时性与复杂度、功耗的解决方案。该方案充分利用了FPGA并行处理的特性，采用了Verilog HDL语言对系统进行设计，着重提出系统的硬件框架，为人机交互芯片设计提供过渡产品。

本设计算法采用C++和openCV对算法和可用性和合理性进行了验证，最后采用Verilog HDL语言对系统进行重构，以Terasic公司提供的DE4-230与2D摄像头为硬件平台对系统进行了仿真、验证和评估。经过最后的评估，系统实现了对分辨率为640\*480、帧率为30帧每秒的视频源进行了实时的检测和跟踪。实现了一套高性能的人手检测跟踪系统。

**关键词**：手势识别跟踪；FPGA；算法系统；硬件架构

Abstract

目录

[摘 要 5](#_Toc369612784)

[Abstract 6](#_Toc369612785)

[第一章 绪论 10](#_Toc369612786)

[1.1课题研究背景及意义 10](#_Toc369612787)

[1.1.1非接触人机交互 10](#_Toc369612788)

[1.1.2非接触人机交互的应用 11](#_Toc369612789)

[1.2基于FPGA设计的自然人机交互国内外研究现状 12](#_Toc369612790)

[1.3本文的研究内容和主要工作 13](#_Toc369612791)

[1.4 本章小结 14](#_Toc369612792)

[第二章 手势识别跟踪算法系统介绍 15](#_Toc369612793)

[2.1 手势识别跟踪算法系统概述 15](#_Toc369612794)

[2.2 基于Harr特征的Adaboost手势识别算法 15](#_Toc369612795)

[2.2.1 Harr特征 16](#_Toc369612796)

[2.2.2 检测过程 16](#_Toc369612797)

[2.3 任意手形跟踪算法 18](#_Toc369612798)

[2.3.1 肤色分割算法 19](#_Toc369612799)

[2.3.2 帧差运动检测算法 21](#_Toc369612800)

[2.3.3 Mean-shift质心跟踪算法 22](#_Toc369612801)

[2.4 卡尔曼滤波算法 23](#_Toc369612802)

[2.5 人脸定位算法 25](#_Toc369612803)

[2.6 手势识别跟踪决策算法 25](#_Toc369612804)

[2.7 本章小结 26](#_Toc369612805)

[第三章 FPGA设计综述 27](#_Toc369612806)

[3.1 FPGA技术 27](#_Toc369612807)

[3.2 FPGA算法设计方法及其特点 28](#_Toc369612808)

[3.3 FPGA验证平台概述 29](#_Toc369612809)

[3.4 本章小结 30](#_Toc369612810)

[第四章 手势识别跟踪算法系统的硬件设计 31](#_Toc369612811)

[4.1 系统架构设计 31](#_Toc369612812)

[4.1.1 系统架构设计及模块划分 31](#_Toc369612813)

[4.1.2 系统存储架构设计 33](#_Toc369612814)

[4.1.3 接口及时序标准 34](#_Toc369612815)

[4.1.4 算法的硬件化改进 35](#_Toc369612816)

[4.2 Adaboost手势检测算法的硬件设计 37](#_Toc369612817)

[4.2.1缓存与压缩模块设计 39](#_Toc369612818)

[4.2.2积分图生成模块设计 41](#_Toc369612819)

[4.2.3选择器组模块设计 42](#_Toc369612820)

[4.2.4分类器组模块设计 43](#_Toc369612821)

[4.2.5主控模块设计 44](#_Toc369612822)

[4.3 任意手形跟踪算法的硬件设计 45](#_Toc369612823)

[4.3.1 肤色分割算法的硬件设计 45](#_Toc369612824)

[4.3.2 帧差运动检测算法的硬件设计 47](#_Toc369612825)

[4.3.3 Mean-shift质心跟踪算法的硬件设计 49](#_Toc369612826)

[4.4 卡尔曼滤波器的硬件设计 50](#_Toc369612827)

[4.4.1算法的硬件转化 50](#_Toc369612828)

[4.4.2 硬件设计实现 53](#_Toc369612829)

[4.5 本章小结 55](#_Toc369612830)

[第五章 硬件系统的仿真、验证与性能分析评估 56](#_Toc369612831)

[5.1 仿真、验证、评估的方案与方法学 56](#_Toc369612832)

[5.2 功能仿真与板级验证 57](#_Toc369612833)

[5.2.1部分功能模块仿真结果 57](#_Toc369612834)

[5.2.2静态图片板级验证结果 59](#_Toc369612835)

[5.2.3视频流板级验证 62](#_Toc369612836)

[5.2.4仿真验证结果分析 63](#_Toc369612837)

[5.2.5性能分析 63](#_Toc369612838)

[5.3 跨平台联合测试与分析评估 64](#_Toc369612839)

[5.3.1直观、定性地分析硬件实现效果 64](#_Toc369612840)

[5.3.2 滤波器性能的定量分析 65](#_Toc369612841)

[5.4 系统性能分析及交互评估 70](#_Toc369612842)

[5.4.1 系统的运行性能分析 70](#_Toc369612843)

[5.4.2 系统的硬件性能分析 72](#_Toc369612844)

[5.4.3 系统的交互分析评估 73](#_Toc369612845)

[5.5 本章小结 75](#_Toc369612846)

[总结与展望 76](#_Toc369612847)

[工作总结及意义 76](#_Toc369612848)

[后续工作展望 76](#_Toc369612849)

[参考文献 78](#_Toc369612850)

# 第一章 绪论

## 1.1课题研究背景及意义

本课题所涉及的领域包含了人机交互设计、图像处理、FPGA前端设计以及系统框架设计。旨在用硬件描述语言Verilog HDL的并行性一定程度解决大数据量、大数据吞吐率图形处理所面临的关于实时性和复杂度的矛盾问题，同时为图像处理芯片化提供过渡性产品。从算法研究、软件实现到FPGA硬件设计提供了一整套完整的解决方案。

### 1.1.1非接触人机交互

人机交互、人机互动（英文：Human–Computer Interaction或Human–Machine Interaction，简称HCI或HMI），是一门研究系统与用户之间的交互关系的学问。系统可以是各种各样的机器，也可以是计算机化的系统和软件。人机交互界面通常是指用户可见的部分。用户通过人机交互界面与系统交流，并进行操作。小如收音机的播放按键，大至飞机上的仪表板、或是发电厂的控制室。人机交互界面的设计要包含用户对系统的理解（即[心智模型](http://baike.baidu.com/view/2333986.htm)），那是为了系统的[可用性](http://baike.baidu.com/view/1436.htm)或者用户友好性[[1]](#endnote-1)。

非接触式人机交互，顾名思义，是指不需要与系统有直接接触的交互方式。例如通过摄像头捕捉用户的行为，与系统进行交互，或者用户通过语音与系统进行交互。这种交互方式的特点在于用户与系统之间可以进行无障碍交互，交互过程变得更加人性化、智能化和自然通畅[[2]](#endnote-2)。

从心理学角度，美国著名的管理学家道格拉斯•麦格雷戈指出，人的本性是懒惰的，一般人都有好逸恶劳、尽可能逃避工作的特性。人类的文明就在于文字和工具，因此工具的不断发展、不断完善、不断符合人类的使用习惯也是历史发展的潮流。

科技的进步、经济的繁荣加快了生活的节奏，也增大了当代人的生活压力。工作上的快节奏需要家庭的安逸来抚慰劳累的心灵，此时，广义上的家就需要有更多的工具来让人们从繁忙的工作中解脱出来。

然而，传统的家用电器的控制方式往往借助某种控制器，例如电视机的遥控器、空调的控制面板等。而控制器一般无法随身携带，因此给我们的生活带来了很多的不便利，此时我们需要一种全新的交互方式来解放人的双手和实现良好的用户体验。因此，类似于语音和手势控制等非接触式交互方式会成为未人机交互的主流。

### 1.1.2非接触人机交互的应用

人类社会的发展，要求了交互方式的更新换代，而什么样子的交互方式是最自然、最符合人类应用需求成为了一个热门的研究话题。早在2008年，比尔盖茨就提出了“自然用户界面”的概念，并预言了人机交互方式的发展路线图，他认为鼠标和键盘灯接触式控制方式将逐渐被更加自然的非接触式控制方式所取代，类似于语音控制、手势控制等。

目前，学术界和工业界对于非接触式人机交互的命题进行了大量的尝试和应用。并得到了一些成果。微软Xbox One继承了Xbox360的在体感游戏的理念，加大了用户对游戏非接触式的概念。通过捕捉人体躯干动作，识别人类的肢体语言，并在游戏中转化为相应的操作控制方式。让游戏的体验不再仅仅局限于鼠标、键盘、游戏手柄的输入了，而是通过摄像头，这种非接触式的输入工具实现了所见即所得的用户体验理念。将虚拟现实的游戏与现实中的运动健身相结合，完美诠释了非接触式控制的让我们摆脱工具，能够更加随心所欲的“懒惰”特性。麻省理工工学院的Pranav将手势识别、投影和云端计算机进行有机融合，设计出一套全新的“第六感设备”。摆出一个拍照的姿势就可以实现拍照的功能；只要一张白纸，通过投影仪就可以实现一个显示器的功能；还可以通过改变纸张的方向来实现游戏中速度和方向的控制。微软公司根据产品的需要设计了CamBot，实现了对Power Point、Media Center、Photo Viewer等多媒体工具的手势控制，虽然只是简单的换页、切换等控制，但是已经让演讲者摆脱了遥控和演讲笔的束缚。谷歌在手势控制方面也做过了类似的尝试，它旗下的Gmail Motion通过摄像头识别手势的动作，实现了原来操作邮箱所需要的控制。另外，安卓4.0也人脸解锁、Leap Motion的手势交互等等产品的也相继应用到现实生活中。

## 1.2基于FPGA设计的自然人机交互国内外研究现状

如1.1.2节所述，非接触式人机交互已经在逐渐在日常生活中应用。但是上述的产品和研究成果大部分是基于冯诺依曼体系计算机搭建的，辅以摄像头等硬件设备构成的软件系统。系统实现的基本要求是实时性，通过摄像头实时处理当前的数据。1824年，英国伦敦大学教授皮特·马克·罗葛特在他的研究报告《Persistence of vision》中首次提出了暂留效应（Persistence）的概念，也成为余晖效应。报告中指出，眼睛的一个重要特性是视觉惰性，即光象一旦在视网膜上形成，视觉将会对这个光象的感觉维持一个有限的时间，这种生理现象叫做视觉暂留性。对于中等亮度的光刺激，视觉暂留时间约为0.05至0.2秒。因此，实时性要求系统在0.05s内需要实现对单帧视频的处理，也就是要求系统处理视频的帧率需要大于等于1/0.05s=20帧。本文所设计的系统在进行硬件设计之前对算法进行了软件的验证。软件算法的验证平台采用了联想T440p笔记本电脑，该电脑配置了Intel Core i5-4200M CPU（主频2.5GHz），4G DDR3 1600M内存。基于openCV图像处理函数库对系统进行了算法级别的验证。实验表明，对于一张640\*480分辨率的灰度图，基于Haar特征对人手进行检测，人手大小可遍历搜索，每秒能够检测出5-6帧图像，无法满足实时性的要求。

由于软件算法依赖于计算机的配置，实时性也无法一概而论。而且在功耗方面也面临一个矛盾的问题。冯诺伊曼体系的计算机的处理器的功耗计算公式如下：

其中P表示功率，表示翻转因子，表示节点电容，表示工作电压，表示工作频率。为了实现系统的实时性要求，其中一种办法是加大处理器的频率，然而频率将线性加大处理器的功耗，因此也是一个性能和功耗的矛盾问题。

由于性能与功耗和硬件设备的依赖关系是两对矛盾，国内外的专家也在不停探索着解决这个问题的方法，随着FPGA工艺的不断提升，芯片设计的成本也不断降低，一种不依赖冯诺伊曼计算机体系的图像解决方案有可能解决这个问题，并利用其本身硬件的并行性应用到人机交互的实时性设计上。

随着智慧城市在近年来越来越成为了社会关注的热点，这就要求机器能够“自主”判断人类的行为和现实中发生的事情，从而做出正确的决定。在工程控制方面，传感器的发展和大规模使用一定程度上可以使机器更加智能化，但是并非“智慧”。所谓的智慧需要理解，行为理解中除了传感器还需要用到大量的图像处理技术，基于FPGA的数字图像处理技术变得尤为重要。其中原因有三：第一，FPGA作为集成电路的验证平台，为图像处理算法的芯片化提供了过渡解决方案；其二，FPGA本身的硬件特性能够处理大量的数据，相比于PC有着更优的性能；其三，低价位、低频率、低功耗便能够实现成本很高的PC或者DSP的功能。因此在这种大背景下，在车载视频终端、智能监控终端、安全门系统以及各种运动跟踪产品将越来越多采用FPGA作为处理单元。不仅仅有功耗低的优势，能够用于低成本的可穿戴式产品，也能用于对性能有要求的高性能产品。

## 1.3本文的研究内容和主要工作

该论文首先介绍手势识别跟踪系统等非接触式系统的现状、发展以及对于当前用户的体验的意义；然后从算法上分析该系统用到的算法的优劣并用软件和硬件的方式对其进行验证；最后通过子模块设计、仿真、时序约束等流程将系统在FPGA上进行验证。系统最重要的工作在于系统的硬件框架设计和对系统的时序仿真。

本论文架构设计如下：

第一章阐述了课题关于非接触式人机交互的现状、研究意义、工作安排；

第二章对系统使用到的软硬件平台进行介绍；

第三章对系统算法进行详细介绍；

第四章阐述系统的硬件架构设计以及核心模块设计；

第五章对硬件系统进行仿真验证、时序约束和班级验证；

第六章对该课题进行了总结，并就系统的不足之处展开讨论并希望得到进一步改进。

本论文主要的工作安排如下：

1. 关于非接触式人机交互系统的调研，包括背景、现状、可能的发展进程；
2. 学习掌握关于本系统需要基础知识，包括C++语言、OpenCV库的使用、FPGA开发平台的使用、Verilog HDL高级语言、时序约束、仿真；
3. 系统算法的学习研究，包括Adaboost算法、背景建模算法、滤波器算法、任意目标跟踪算法等算法的学习；
4. 根据软硬件的算法架构不同，提出适应FPGA设计的硬件架构，包括通信架构、存储架构、实时性与复杂度矛盾等瓶颈问题的解决方案；
5. 设计一套非接触式人机交互系统，包括视频采集、视频处理、处理结果回显的系统；
6. 使用C++与openCV对系统进行软件验证；
7. 使用Verilog HDL对算法进行逻辑验证，利用设计好的硬件架构，让系统兼备实时性、高数据吞吐率，让系统在处理30帧每秒的图像达到实时性；
8. 使用Time Request时序约束工具分析系统静态时序性能，提高系统鲁棒性；
9. 使用Modelsim仿真工具让硬件系统动态分析系统功能时序，验证系统鲁棒性；
10. 结合SignalTapII以及 Matlab等工具对系统进行板级调试。
11. 总结并撰写论文。

## 1.4 本章小结

本章结合了非接触式人机交互的背景和现状，分析了不同的方案的实现已经各种实现方案的优缺点及瓶颈，提出了一种算法上基于OpenCV、实现上基于FPGA的解决方案。在实现过程中，既利用到软件算法对视频数据对图像处理的研究成果，又利用FPGA并行化的硬件特性，解决了数据吞吐率和实时性的矛盾问题。到达良好用户体验的同时又为人机交互芯片的发展设计出一套过渡产品。但是由于FPGA算法的不成熟，距离一套成熟的可商用的人机交互系统来说仍然有一定的差距，需要在后续的工作中继续完善和优化。

本课题致力于推动数字化家庭终端的发展，为非接触式人机交互设计一套折中的方案，为人机交互芯片化进程提供过渡产品。

# 第二章 软硬件平台介绍

## 2.1 硬件平台及工具介绍

本系统使用的高级设计语言为硬件描述语言——Verilog HDL，系统充分利用其并行性为系统提供高速处理数据的能力。硬件算法验证阶段使用了Terasic公司的DE4-230作为验证平台，该平台搭载了StratixIV核心处理器。

### 2.1.1 FPGA技术简介

FPGA(Field－Programmable Gate Array)，既为现场可编程门阵列。是由前期的PAL、GAL、PLD、CPLD等其他可编程逻辑器件发展而来，属于半定制的集成电路芯片。从功能上来说比CPLD等老一代产品有着更加多的触发器和逻辑单元，更加适合做时序电路设计，满足我们的系统要求；从产品的角度上来说比全定制的集成电路芯片更加灵活，由于其可编程的特性，既兼容了硬件电路并行的速度优势又有可修改可编程的灵活性。

半导体的集成度与成本一直遵循着摩尔定律的演进历程，当然FPGA也不例外。俊龙科技的产品经理胡晟如是说：“每一次工艺的更新都会在产品功耗、运行效率、容量上体现出FPGA的优势。”FPGA的工艺在不断进步，也在不停适应着整个物联网兴起的大趋势。Lattice在近年来相继推出了ICE40、ICE40 ULTRA，Altera推出的MAX10，Xilinx的Zynq系列正式为了适应日渐兴起的物联网行业的发展。

文献【3】利用FPGA的SOPC（system on a programmable chip）功能模块和神经网络算法设计了机器人。实现了在同一块芯片上实现了复杂的功能。通过测试发现:机器人的移动躲避障碍物时速度的稳定性较好,移动误差较低,实现了果蔬采摘无人控制下的智能移动[[3]](#endnote-3)。文献【4】的研究表明,基于FPGA的OSGO-CFAR检测器在多目标环境中能大大提高小目标的检测概率[[4]](#endnote-4)。文献【5】介绍了一种基于FPGA的万兆光纤以太网高速传输的设计方法。在研究IEEE Std802.3和802.3ae规范的基础上,采用FPGA与Sharp Pcap的技术,实现了万兆光纤以太网的高速数据传输链路,其中利用FPGA实现了万兆以太网协议的媒体访问层(MAC)与物理层(PHY)的相应功能,并利用Sharp PCap在主机端实现网络数据包收发。经过实际系统测试和性能分析,传输速率达到6400 Mb/s,能够满足新体制雷达的高速数据传输需求[[5]](#endnote-5)。文献【6】也阐述了FPGA在图像处理方面的应用[[6]](#endnote-6)。在当前大多数的流行领域，FPGA都有着突出的应用。在关于实时性和算法复杂度矛盾的情况下， FPGA有着得天独厚的优势。

### 2.1.2 系统使用的硬件平台简介

算法验证阶段，采用了Terasic公司的DE4-230验证包对算法进行验证和板级调试。在完成了板级调试之后通过参考DE4-230的设计，重新设计了验证版并成功将算法移植到新设计的验证版卡上。

关于DE4-230与自行设计的FPGA验证板卡所使用到的Stratix IV，官网给出如下数据：

|  |
| --- |
| Stratix IV GX EP4SGX230 |
| 228,000 logic elements (LEs) |
| 17,133K total memory Kbits |
| 1,288 18x18-bit multipliers blocks |
| 2 PCI Express hard IP blocks |
| 744 user I/Os |
| 8 phase locked loops (PLLs) |

由上述表格可以得知，该FPGA处理器含有大量的逻辑单元（LEs）和相对比较大的内部存储空间，同时提供了8个锁相环（PLLs）。由于丰富的资源，让该FPGA芯片能够实现复杂的算法，真正意义上实现片上系统（system on chip），既在一个芯片上实现一个复杂的人机交互系统。锁相环IP核的使用，让设计者能够更好控制系统的时序，让系统更容易达到时序收敛。

## 2.2 软件平台及工具介绍

在进行硬件系统设计之前，需要验证算法的性能是是否符合系统要求。因此需要选择一种快速原型的方法。对比了Matlab与C++的优劣势，本人认为Matlab脚本语言有利于系统的快速原型，但是相比于OpenCV开源库的集成化程度，利用C++调用OpenCV来进行快速原型更具备优势。尽可能在短时间内验证算法的功能是否符合系统要求。

### 2.2.1 OpenCV技术简介

OpenCV全程是Open Sourse Computer Vision Library，在1999年由Inter建立，是一个跨平台的计算机视觉库，同时提供Python、C#、C等多门语言的API接口。随着计算机视觉巨大而持续的增长，OpenCV通过了C语言编写对其进行优化，并通过购买了Intel的IPP高性能多媒体函数库（Integrated Performance Primitives）使得其速度更一步优化。

本系统设计引用到了OpenCV中的Adaboost检测函数、RGB转HSV函数、直方图均衡化（Histogram Equalization）、Meanshift函数等，这几个主要函数在系统中的作用如下表所示，算法原理将在下一章节进行详述：

|  |  |
| --- | --- |
| 算法 | 主要作用 |
| Adaboost | 将样本经过训练生成xml文本，并对输入数据进行训练 |
| RGB转HSV | 颜色空间转化，在H空间肤色具有良好的聚类性，便于肤色分割 |
| 直方图均衡 | 肤色分割的数据预处理模块 |
| Meanshift | 通过迭代的方式查找跟踪框的质心，以达到手势跟踪效果 |

基于OpenCV的库函数仅仅用于验证系统的可行性，关于算法在FPGA的实现需要重新进行设计和优化。

### 2.2.2 软件实现效果简介

利用OpenCV库函数的快速原型作用，通过调用电脑的默认普通摄像头，实现了基于手势识别、任意手型跟踪的效果。系统的详细介绍在下一张将具体阐述，以下是大致效果图：

|  |  |
| --- | --- |
| 图2- 1 肤色分割的效果 | 图2- 2 识别及跟踪的效果 |

通过原型的系统可以看到，基于adaboost识别算法、meanshift跟踪算法能够达到非接触式人手跟踪的效果，通过该方法能够为非接触式控制提供可靠的数据输入，因此通过验证表明， 该方法可行。但是由于系统的数据量相对较多，需要对数据的处理时间较长，经过时间验证，基于OpenCV的非接触式人机交互系统每一帧的运行时间如下图所示：

|  |
| --- |
| 图2- 3基于软件的每帧运行时间 |

由实际实验测量结果看来，每一帧数据的运行时间大约为100ms，既一秒钟的帧率为10帧，达不到系统实时性所要求的的每秒30帧的速率要求。因此需要设计一种特殊的架构来解决这个矛盾问题。

## 2.3 本章小结

本章重点介绍了系统设计所需要用到的核心开发平台，包括了硬件的FPGA平台和软件的OpenCV开源库。其中，软件的系统实现仅仅是为了验证算法是否能够用于非接触式人机交互系统的前端数据处理，并不作为最终作品呈现，因此简略介绍。而对于本论文的最终成果是基于FPGA的片上系统，用单一的FPGA芯片就实现一套非接触式人机交互系统，不依赖于计算机，为该领域的算法芯片化提供一种解决方案和一个过渡性产品。

# 第三章 系统结构及算法介绍

本章将对本毕业论文所设计的非接触式人机交互系统从算法层面进行简述，着重介绍整个系统算法的结构和流程、各个子模块的算法原理。

## 3.1 系统结构与流程

如上一章2.2.2小节所述 ，经过优化后的软件系统达不到理论的实时性要求，因此需要为系统设计一个解决该问题的硬件架构。如第一章《绪论》中提到，FPGA采用的是硬件描述语言，能够将系统设计为并行的系统。因此我们可以将不同的算法模块并行化，提高算法的运行效率。下图中左图是软件算法的架构图，右图是改进后的算法流程图：

|  |  |
| --- | --- |
| 图3- 1 基于软件的串行框图 | 图3- 2 改进后基于硬件的并行框图 |

由上述的系统流程框图可以看出，将可以同时执行的模块进行了并行化，并在模块内部也进行了并行化，该部分在后续的模块介绍中进行详细阐述。根据右图的流程图，将系统进行了模块划分并各个模块分配存储器。详细情况如下图所示：

|  |
| --- |
| 图3- 3 系统模块框图 |

上图是基于图3-2设计的硬件框图，将系统重新进行模块划分并为各个模块合理的存储空间。上图红色字体“start”表示系统启动时候同时执行的模块。该“start”由控制中心输出。每个模块中的数字表示该模块对于存储单元的理论要求。系统的各个模块的设计的接口如上图所示。

## 3.2 系统时钟频率以及存储结构设计

系统的时钟频率决定了系统的性能，而更加高的速度将要求有更加先进的工艺并且增大系统功耗 ，因此采用合理的系统主频是必要的，将在3.2.1小节对系统主频进行详细计算说明。而系统的存储结构与主频相铺相成，在保证速度性能达到要求的情况下，尽可能减少芯片面积将可以极大程度减少后续流片的成本和减低功耗。由于存储器包括了片外存储器、片内存储器等多种类型，因此在3.2.2小节中将详细介绍系统如何分配存储器。

### 3.2.1 系统频率计算

如第二章2.2.2小节及本章3.1小节提到，在软件系统中未能达到每秒30帧的实时性要求，那么改进后的FPGA系统需要最小的时钟频率将成为系统最基本的时序要求，由于基于硬件结构的FPGA系统相比于CPU的指令型系统有着时序固定的特点，系统按照原先设计好的时序运行，时序结果是可预知的。因此，当每秒要求30帧时，系统处理一帧视频的速度为1s/30≈33.3ms，系统的输入采用的是标准的VGA视频标准输入，分辨率为640\*480。如果采用27MHz时钟频率的VGA标准视频格式，每一个像素点所需要消耗的时间为1/27M，标准的640\*480的VGA格式的视频加上消影部分，一帧图片的总像素点为800\*525=420000，因此一帧图像所需要用到的时间1/27M\*420000=15.6ms<33.3ms。按照这样子的计算公式，27M的系统时钟就能够使系统达到实时性要求。综上所述，本系统将采用27M的系统时钟来进行系统设计，各个模块都要求在15.6ms的时间内完成对一帧数据的处理。由于27MHz远小于CPU的运行主频，因此在功耗等方面都有着巨大的优势。

### 3.2.2 系统存储结构设计

系统采用的FPGA芯片为Altera公司的StratixIV，片内拥有丰富的片内资源，包括了片内RAM和逻辑单元LCs（Logic Cell）。系统同时也用到了片外存储器用来缓存原始的视频图像，外部存储器用到了一块DDR2（The 2nd Double Data Rate SDRAM）和一块64MB的SDRAM，前者用在手势识别的积分图计算时候的缓存，后者用于视频预处理中对于视频的解码。

文献【7】[[7]](#endnote-7)中，同样实现了对目标的跟踪，采用的存储器以片内存储器为主，虽然达到了性能要求 ，但是不利于后续的芯片化和产品化，因为其将耗费大量的硅片资源。就长远考虑，系统对于存储器的设计必须及考虑性能又要考虑设计的可用性，即是否能够符合工业上对芯片面积的要求，因此将耗费大量的片内资源的模块的存储部分移植到片外是必要的。

本系统在对存储进行设计的时候秉承面积优先的原则，即保证性能的前提下尽可能减少片内资源的使用。

图3- 4中给出了各个模块需要用到的存储器数量，并且将需要用到外部存储器的模块进行了说明，有改图可以明确知道不同模块将需要用到的资源数量，也用于存储结构进行设计的指导。

综上所述，按照各个模块的需求，做出这样子的分配，并在后续的实现中进行修改，以下分配属于理论分配。视频的预处理模块需要缓冲2帧数据，因此采用外存；背景建模模块同样需要用到缓冲2帧数据因此也采用外存；人手检测中的积分图缓存需要用到大概300KB的存储空间，因此采用外存；Meanshift模块、滤波模块、重定位模块等将设计为实时数据处理模块，不需要做数据缓存，因此采用片内存储缓存数据。由于外存与FPGA本身的数据交互需要用FIFO（First In First Out）进行数据同步，因此，人手模块、视频预处理模块、背景建模模块大部分的数据将采用外存的方式进行存储，但是仍然需要用到以RAM方式存在的FIFO。针对上述分配方式，用下图呈现：

|  |
| --- |
| 图3- 5 不同模块的主要存储器分配 |

该小节通过理论的计算，得到了各个模块可能需要用到的具体存储资源。属于粗略的估算，不代表哦最后系统的资源使用情况。但是一定程度上在FPGA选型、存储器选型、系统主频率选择等多方面为系统的设计奠定了基础。而每个模块的算法原理和存储资源具体计算方式将在下一个小节进行详细介绍，并在下一章中对系统进行硬件实现的时候给出实际综合的结果。用于与理论结果进行对比。

## 3.3 系统算法子模块介绍

按照本章3.1节系统框架设计中对各个模块的接口和时序要求，将系统划分为几个功能模块，并在系统的统一系统时钟27MHz的同步下协调工作。本节将对系统所使用到的不同模块进行具体算法分析和存储结构设计。

### 3.3.1 基于Haar特征的Adaboost人手识别

Adaboost（Adaptive Boosting）顾名思义是一种自适应的Boosting算法，属于图像处理领域中的经典算法。在机器学习领域与SVM合成两大嘴经典的算法，因此其理论基础成熟。其最主要的特点是训练慢，但是识别速度比较快。但是由于该算法大部分都是在软件层面上实现的，本系统也利用OpenCV和C++实现了基本功能作为算法验证。但是最终还是希望利用FPGA对其进行实现，因此本小节依旧将算法从基础开始分析并设计出FPGA的存储结构。

本系统采用了Haar特征作为人手识别的依据，在文献【8】关于人脸识别的期刊中提到Adaboost算法使用类Haar特征的优势是Haar特征相比于其他特征有更快的计算速度并且能够适应尺寸不同的目标识别[[8]](#endnote-8)。图3-6是小部分类型的类Haar特征图例，每一个特征都可以计算得到一个值，该值是用每一块直方图去乘以权重得到的值进行相加便是该特征所代表的的值。

|  |  |
| --- | --- |
| 图3- 6 一部分Haar特征 | 图3- 7 以其中一个Haar特征为例 |

所谓的权值是通过大量样本训练出来的的数值，是本章3.1所提到的xml文件里面的结果，本系统的训练样本采用OpenCV提供的样本，假设A,B,C,D四个直方图块的权值分别为、、，那么该Haar特征对应的值就是

其中，A、B、C、D的值根据定义为-1或者为1。由此得到每个特征对应的，不同特征的值相加将得到一张图像的特征值总和，将于某一个阈值进行比较用于判断该图像是否包含需要检测的目标。

在Adaboost算法中，Haar特征是基本单元，而核心的模块是分类器。分类器顾名思义就是用来区分图像是否包含了所需要检测的目标。该算法中，分类器分为弱分类器和强分类器。弱分类器和强分类器是通过训练得到的，训练得到的各个分类器又是后续进行目标识别的识别工具。假设已经拥有了这m个样本，其中有正样本也有负样本，正样本的值为1，负样本的值为-1，用来表示。在训练之前，每一个样本的权值是一样的，训练的过程就是通过迭代不停提高正样子的权重而降低负样本的权重。因此，训练之前将每一个样本的权值都设置为：

假设一个弱分类器对训练样本进行分类，将会有一些样本被分配到错误的类别中，错误的概率定义为，并将其定义为：

表示被假设的分类器分类的结果，上述公式表示在分类的情况下，所有的错误样本的权值相加。

将分类的错误率进行归一化，将其定义为：

根据归一化的来更新不同的样本的权值：

其中，表示第个样本的在t+1时刻的权值，为一个归一化的因子。根据上述公式的计算，对于不同的假设的分类器的最终的输入如下：

接下来通过一个例子对上述公式进行分析：

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

a图中共包含了10个样本，因此每一个样本的权值=0.1，b图中的是第一个训练的分类器，左边表示正样本，右边表示负样本，右边三个带圆圈的正样本被错误分为负样本，通过公式【】和公式【】计算得到,，通过计算不同样本的权值得到c图。重复上面的计算步骤可以得到d图、e图和f图，这样子就得到了三个弱分类器，在根据公式【】将三个弱分类器进行线性组合既可以得到强分类器，如g图所示：

|  |
| --- |
|  |

以上即是强分类器和弱分类器的训练过程。

关于对输入图像进行识别即是训练的逆过程，先后进行几个阶段，分别为对图像进行缩放、遍历所有候选的窗口、将没有包含检测目标的窗口淘汰、检测出目标窗口。具体的流程图如下所示：

|  |
| --- |
| 图3- 8 Adaboost人手检测流程图 |

其中输入图像为640\*480的原始灰度图，特征池为本节介绍的训练得到的xml文件。通过级联的方式将多个弱分类器串联成为强分类器，将图像通过多个强分类器进行判断，当图想通过了所有的强分类器且没有被淘汰，则说明该帧图像中包含人手，并返回包含人手的具体坐标和大小。

在FPGA系统主要是对目标进行检测，训练部分是一次性工作，利用PC训练出结果之后直接调用xml文件即可。因此需要设计你FPGA上对于人手的识别模块，该模块的框图如下图所示：

|  |
| --- |
| 图3- 9 adaboost模块FPGA实现的模块划分及存储结构 |

Adaboost模块是系统的初始化模块，为后续的人手跟踪模块提供初始化位置，由于该方法复杂度较高，因此本人的本科毕业论文提出一种基于挥手对人手的初始化位置进行检测的算法 ，但是本人设计的算法无法得到手的大小，不便于对手势进行跟踪，因此本论文采用经典的Adaboost算法对人手进行检测和定位。

### 3.3.2 基于meanshift的任意手型跟踪

在3.3.1小节中，利用了Adaboost算法对人手进行了初始化定位，而非接触式人机交互系统需要捕获任意时刻人手的位置。捕获人手的方法有多种，通过查阅文献大致分为以下几种：

1. 利用高斯粒子滤波对人手进行跟踪；
2. 通过提取肤色或者运动信息对人手进行跟踪；
3. 直接利用Adaboost的检测结果作为跟踪路径实现对人手的跟踪；
4. 建立背景模型，提取运动中的人手作为跟踪目标。

方法a在文献【9】中得到了应用，并有着良好的跟踪稳定性，文章中表示粒子滤波采用了高斯随机分布[[9]](#endnote-9)，极大程度提高了跟踪的鲁棒性。

方法b是经典的算法，在文献【10】中使用了肤色作为跟踪的主要特征，文章中表示肤色作为固定的特征，能够快速提取，提高跟踪效率[[10]](#endnote-10)。

方法c 是一种理想化的方法，目前没有实际的应用，通过实际的软件实现发现，并不是每一帧含有人手的图像都能够正确检测到目标，这个是Adaboost算法的强分类器级联的苛刻要求所决定的，漏检在Adaboost中是允许的，因此方法c并不适合实时跟踪系统。

方法d在文献【11】中有着详细介绍，文章着重讨论了仅仅利用背景模型对前景进行提取 ，是否能够实现对目标进行跟踪，得到的结论是肯定的[[11]](#endnote-11)。

本人针对上述四种办法分别利用OpenCV对其进行了快速原型，并得到了各种跟踪算法优势及劣势，分析如下：

方法a基于粒子滤波的方式的特征在大部分文章中都使用了SVD降维算法进行提取，SVD特征值矩阵的求解涉及到SVD矩阵分解，QR分解等大量的数据运算，而这不是FPGA所擅长的，也是难点。至于结果而言，系统目标是实现任意手型跟踪，而基于粒子滤波的方法，对突变的手型无法快速适应，容易跟丢。

方法b基于肤色的跟踪算法能够弥补方法a对于手型变化跟丢的局限性，而且算法简单，容易实现，其不足之处在于容易受到类肤色背景的干扰，包括了纯色背景墙、裸露的手臂、人脸等。

方法c由于Adaboost本身的局限性造成了无法连续进行跟踪，而且不能适应任何手型的变化，因此不符合系统要求。

方法d和方法b的难度相当，但是容易受到运动的背景目标的影响，直接导致了跟错目标的情况。

基于上述分析本系统选择了基于肤色和背景建模相结合的方案，两种算法的优劣势是互补的，符合系统要求。

1. 自适应肤色分割

肤色分割是利用肤色在某一个颜色空间具有聚类性而对颜色特征进行非精确性提取的一个过程。自适应就是肤色分割的过程中能够根据不同人的肤色、不同的光照条件以及不同的背景来调整肤色特征提取的参数，使得非精确性的特征提取尽可能可靠。

肤色在Ycbcr颜色空间具有聚类性，即在该颜色空间中，肤色的值会集中在某一个区域体现。利用肤色在该颜色空间的特点，通过含有肤色的图片样本可以训练出大致的肤色分布范围。加上自适应的策略，即可以让该肤色分布范围在肤色识别的过程中动态调整，以适应不同的光线亮度条件以及不同的肤色特点。首先是肤色训练的过程。取得一百张仅含肤色的图片，利用脚本语言工具MATLAB以Cb为横坐标，以Cr为纵坐标，画出二位的平面图：

|  |
| --- |
| 图3- 10 肤色分割训练效果 |

由上图可以看出 ，肤色聚类在一个小范围内，识别过程中可以利用Cb与Cr的范围判断来确实输入的像素值是否属于肤色范围。

自适应肤色分割是利用3.3.1中介绍的人手识别算法，由于adaboost能够精确定位到人手的位置，因此统计计算手心的肤色的Cb与Cr值的一阶矩，用来作为图3-10肤色范围的中心，并根据不同的Cb、Cr值来对肤色范围进行移动，便可以一定程度上实现了自适应肤色分割，在不同的肤色、不同光照条件下都能够正常进行人手肤色的识别。如下图所示：

|  |
| --- |
| 图3- 11 自适应肤色分割图示 |

假设A为图3.10中聚类的效果，由于光照条件或者人与人肤色的不同有可能聚类在B、C、D等地方，那么需要一种办法来自适应这个聚类效果，使其能够适应多种不同的情况。本系统将采用下图的方案：

|  |  |
| --- | --- |
| 图3- 12 自适应肤色分割图例 | 图3- 13 自适应肤色分割效果 |

利用3.3.1节的人手检测出手的外形，在寻找手心中纯肤色的64个像素点，求其Cr,Cb的均值作为聚类后结果的质心，如图3-11所示，箭头即为自适应的中心偏移，以此来达到自适应肤色分割的效果。图3-13为自适应肤色分割的效果图。可以看出，对肤色分割的效果基本能够符合系统要求。

这种肤色特征提取的方式就是上述中方法b，假设用肤色进行跟踪，在进过人脸的时候会有跟丢的情况（跟踪框在脸上，误认为人脸是人手）。因此本系统采用b)小节中的背景建模的方法减少人脸部分的干扰。

1. 背景建模

背景建模顾名思义就是建立一个背景模型，并运动中的前景进行提取。以下图为例：

|  |
| --- |
| 图3- 14 背景模型前景提取实例 |

图3为原图，图1为车辆行进过程中的前景提取图例，图2为前景提取的结果。可以看出，利用背景模型与当前的图像的差值可以得到相对较为合理的运动特征。同理，在人手跟踪的过程中，人手需要对系统进行控制，相对于人脸或者背景等肤色信息是运动的，因此可以利用背景模型来减少类肤色背景对人手跟踪的干扰。基于上述分析，需要选取背景模型并提取前景。

背景模型有多种，在文献【12】[[12]](#endnote-12)中提到一种利用混合高斯模型作为背景模型来对运动的目标进行检测的方法，该方法能够准确作为一种前景提取的算法，但是由于GMM算法的非线性特点，并不适合在FPGA平台上实现。文献【13】[[13]](#endnote-13)提到一种利用三帧连续的图像建立背景模型的一种帧差背景建模的方法，同样用于前景的提取。该方法的优势在于当前景速度快速运动的时候能够准确捕捉，但是由于本系统是人手对系统的控制，有可能有很多的微距离操作，手势的速度并不会很快，该方法对于速度较慢的前景捕捉性能较差，不适合本系统的算法要求。

基于前两篇文献的算法的优势，本人将文献【12】对于慢速运动的优势结合了文献【13】的帧差算法，适合在FPGA上实现。

设计的算法流程如下图所示：

|  |
| --- |
| 图3- 15 帧差背景建模算法流程 |

该算法是本人自行设计的一个背景模型，旨在用最简单的算法来达到系统的要求。算法的思想是将第一帧图像作为初始化的背景帧，之后按照当前当前帧与上一帧的差值的绝对值diff以及当前帧与背景帧的差值的绝对值diff\_bg的大小去判断背景的更新率。利用当前帧的像素值与背景帧做差，即可以得到前景图像。

1. Meanshift算法

Meanshift算法（均值偏移）是一种迭代的算法，用于对a）与b）中得到的前景进行跟踪。作为一种对概率密度无参数的估计算法，在实现过程中是不能保证能够得到全局最优质的，只能够向概率密度的局部最优值收敛。其算法原理及流程图如下图所示：

|  |  |
| --- | --- |
| 图3- 16 meanshift算法原理图 | 图3- 17 meanshift算法流程图 |

图中右上角的方框表示跟踪的初始位置，箭头表示跟踪的迭代收敛过程。最左下角为目标跟踪的位置，也就是局部最优解。图中“+”点类比于前景特征。

Meanshift的算法实现步骤如下：

|  |
| --- |
| 1,设定搜索窗口参数，包括窗口初始位置、形状、大小； |
| 2,在搜索窗口内计算概率密度分布的局部中心m(x)； |
| 3,重新设置搜索窗口的中心，即令窗口中心x=m(x)； |
| 4,循环执行（2）和（3），除非满足‖m(x)-x‖<ε或循环次数达到设定的阈值。 |

该算法最重要的概念是矩，用来描述各中概率分布的宏观特性。该系统中使用到了中心距和原点矩。矩的定义如下：

零阶矩是当i，j都为0时m的取值：

一阶矩为i，j任意一个为1，另一个为0时m的取值：

根据一阶矩和零阶矩即可以算出质心：

利用该算法即可以计算出下一次迭代的质心c，并通过迭代得到局部最优值，当‖m(x)-x‖<ε或循环次数达到设定的阈值时，停止迭代并认为所得到的,就是局部最优解，并作为帧图像跟踪的最后结果。

根据3.1节对系统的规划和接口设计，为本模块设计了的接口与内部结果如下图所示：

|  |
| --- |
| 图3- 18 跟踪模块的接口定义以及模块划分 |

该小节论述了本系统的跟踪核心，通过肤色分割和背景建模为meanshift跟踪算法提供前景的输入，并由后者对人手进行跟踪。当一帧图像经过该模块，模块通过运算得到了该帧图像的质心并将其返回给系统，作为当前人手位置的判断依据。结合adaboost的结果以及肤色点（上图中的Skin\_num）数去判断到底采用何种结果作为最终的跟踪位置。

### 3.3.5卡尔曼滤波算法

在3.3.4中提到了对人手进行跟踪，但是由于非接触式人机交互系统对手的运动轨迹有一定的要求，要求跟踪的过程抖动不能过大，如果抖动过大将影响用户人机交互的体验。因此需要一个滤波器来让人手运动轨迹的捕捉更加平滑。本文引入卡尔曼滤波器，用于减少运动信息噪声，并通过过去和当前的手势左边特点预测下一帧手的位置，用来矫正实际跟踪位置。

通过阅读黄小平的《卡尔曼滤波原理及应用》中关于卡尔曼滤波器的基础知识，总结其算法原理及简单实现关键流程。定义一个关于描述人手运动信息的矩阵：,其中x，y表示手部的位置，表示当前手势的运动速度。定义观测矩阵为,其中第一个元素，第三个元素表示上一小节meanshift得到的结果。另外需要定义协方差矩阵P以及卡尔曼增益矩阵。

在极短的时间间隔内，手势的运动可以被认为是匀速直线运动，因此在短时间内手部的在运动轨迹应该遵循直线运动定律：

定义一个状态转换矩阵F如下：

以及一个过程噪声矩阵为：

在二维坐标中可以将观测矩阵M定义为：

因此引入噪声矩阵为：

根据书中的理论知识可知，卡尔曼滤波器主要依赖于以下五个公式进行迭代计算：

本小节仅介绍算法原理，下一个章将进行详细介绍。

## 3.4 本章小结

本章作为本论文的核心章节，首先从整体上为系统设计了一个硬件框架，为系统的硬件结构奠定了基础。然后分析了系统的性能以及实时性对系统资源、系统时钟的要求。最后针对系统中需要应用的模块进行了算法分析和理论推导。本章从理论的角度尝试寻找一种能够解决软件端遇到的瓶颈问题的方法，对比其他文献中所使用的方法并尝试通过实验的方法选择最适合本系统的算法。而对于算法在硬件上的实现办法，将在第四章详细设计。

# 第四章 手势识别跟踪算法系统的硬件设计

## 4.2 Adaboost手势检测算法的硬件设计

Adaboost手势检测算法是本手势识别跟踪算法系统的关键算法，也是最复杂的算法，本章将重点介绍其设计过程。Adaboost手势检测算法在系统的初始化和跟丢后快速找回人手起关键作用，为手势跟踪的迭代过程提供精确地初始位置信息，直接影响整个系统的橹棒性。因此，该算法的硬件设计的运行性能需达到原PC系统的99%以上，才能保证整个硬件系统的性能。

本节主要人手检测系统的整体结构和运作流程。对于其中关键模块的设计进行了具体的论证。并从硬件角度对结构进行了有针对性的描述和改进。

根据算法的特点，进行特定模块的建立：积分图产生模块和分类器模块。由于系统定位在高速、实时系统，所以本系统在资源消耗允许的范围内建立了分类器组，并行处理大大提高运算效率。

考虑到硬件选择器消耗逻辑资源巨大，所以对选择器部分做了优化，根据训练好的特征量来进行归类合并，建立相应的译码表，大大减低逻辑资源的消耗。

图像处理部分的功能模块是本系统中最重要的部分，它对能否把系统做成实时检测系统起着最主要的参考。在系统设计中，时序设计是按照每秒30帧的速率设计。由于系统设计成流水线运算架构，分类器最大的运算速度只需108MHz，这为以后系统的升级提供了很大的空间，如处理图像的尺寸可以更大和检测目标可以更小。

针对本系统是用硬件结构实现，在系统设计中，有针对性的做了结构上的优化：

1.选择器组模块。在算法原理中，所有的特征都在特征池中产生，然后根据每个特征的值，从积分图中进行抽点运算。但在硬件过程中，抽点过程就是一个选择器选择点的过程，由于选择器消耗大量的逻辑资源，在全幅640\*480\*8bit的图像中进行选择将消耗大量的资源。所以此处有针对性的做了优化：设计一个选择器，对于送到对应分类器的数据才提供选择通道，其他的数据不提供选择。这样可以大大节约了逻辑成本的消耗。

2.分类器组模块。考虑到要使FPGA得运算速度达到最高，需要用流水线的运算处理方式。所以设计了一组能进行快速流水运算的分类器。并根据上一节算法硬件化的改进，将分类器组的结构设计为并联分类器结构。

整个手势检测硬体架构如图4-11所示：



图4-11 手势检测硬件架构

人手检测系统工作的流程主要分为三大部分：图像采集与存储、图像处理和图像显示。下面具体介绍这个系统的工作流程。

1、图像采集与存储

在DE2中，通过视频采集模块TV-in Connector，这是一块TV全制式解码芯片（NTSC/PAL/SECAM），使用I2C总线对其进行配置后可输出视频图像，把视频图像格式从ITU-R BT.656 YCrCb 4:2:2 视频数据转成16位ITU-R 601格式图像数据后，存储到SDRAM里面，供后续处理使用。

2、图像处理

（1）在DE2上把采集到的原始图像数据进行灰度转换，把彩色图像变成灰度图像。转换过程为把图像的Y量（亮度分量）分离，以Y值作为灰度值。

（2）把灰度图像传送到DE4上的BRAM缓存。BRAM开辟两块空间用来做乒乓操作，充分发挥硬件处理的流水线操作。

（3）从BRAM里面读数据出来做直方图均衡，以改善图像质量。

（4）把直方图均衡化后的图像按压缩因子进行压缩，以产生不同尺寸的图像，来适应不同大小的人手。

（5）对压缩后的图像进行积分处理，产生积分图像。其中积分图产生的模块包括列积分和行积分模块（行积分模块嵌在积分图阵列中）。根据需要，还要产生一副平方积分图来对图像进行光照补偿。

（6）积分图阵列的点根据特征值预处理后直接连接到相应的选择器模块中。

（7）控制状态机根据特征值控制选择器的数据输出和存放在片内RAM的权值、分类器阈值、影响因子和每级阈值的输出。

（8）分类器组进行高速的特征值运算，算出每一级所对应的特征值，与阈值比较，判断当前窗是否为人手。

（9）对所有窗口做筛选，选择合适的窗口（一般认为是最大的窗口）作为人手窗口输出。

（10）把坐标发回DE2，并在图像上显示检测到的人手窗。

（11）重复步骤1-10，进行下一帧图像的检测。

3、图像显示

当图像数据处理完后,要将视频显示在显示器上。在图像数据输出时要进行D/A转换操作，一般的VGA解码芯片包含了D/A转换过程，其控制时序一般为标准的VGA控制时序，故该模块具有较好的通用性。本系统采用的是AD公司的ADV7123 VGA解码芯片。这是一款三通道、10位数模转换器。它的吞吐量可达330MSPS。

下面是对关键模块设计的具体阐述：

### 4.2.1缓存与压缩模块设计

该模块主要完成缓存图像和压缩图像的功能。图像压缩的方法比较常用的有两种：一种是简单的采样抽点的图像压缩方法；一种是基于局部均值的图像压缩方法。第二种方法相对于第一种方法，其压缩的图像更加平滑，但计算复杂度较高。但对于图像的平滑度对于抽取Harr特征这种矩形状特征而言，影响并不明显。故本设计采用第一种简单的方法。

压缩图像的操作实质上是根据压缩因子的大小，对图像进行抽点操作。如一副640\*480分辨率的图像，进行压缩因子为1.25的一次压缩，则压缩后的尺寸为512\*384。那就需要从原图像中抽出512\*384个像素点组成一副图像。本设计中的做法是：分别计算需要的行、列值，抽出相应的行、列对应的点。本系统这样确定需要抽出的行、列值：设需要抽出第*i*行、第*j*列，*i,j*的值由式4-1和式4-2确定

(4-1)

(4-2)

在硬件设计中，进行浮点运算非常复杂且消耗较大的资源，所以从上式看出，需要把它转为定点运算。除4的操作实际上可以看做直接把结构的后两位舍去，即移位操作。这样，运算被大大简化。

一副图像的数据是640\*480\*8bit，把数据缓存在480片容量为640\*8bit的BRAM中。480片BRAM存储空间对应着480行的图像数据。在进行图像缩放的操作中，把计算得到的j值作为地址写进每块BRAM中，BRAM输出的数据就是对应的列数据，完成对列数据的抽取。根据计算得到的i值，选择对应行的数据输出，完成对行的抽取功能。并且需要同时缓存两帧图像做乒乓操作，保证压缩图像数据不间断输出。

这种方法简单直观，但要消耗600KB片内存储器，大大超出了本课题的片内存储的上限。对此，需要优化设计，这里可用外部存储器作为图像帧缓存，由于外部存储器DDR RAM无法对单个不连续的数据进行间隔读取，因此，需要再用少量的片内BRAM作图像压缩缓存。其结构如图4-12所示：

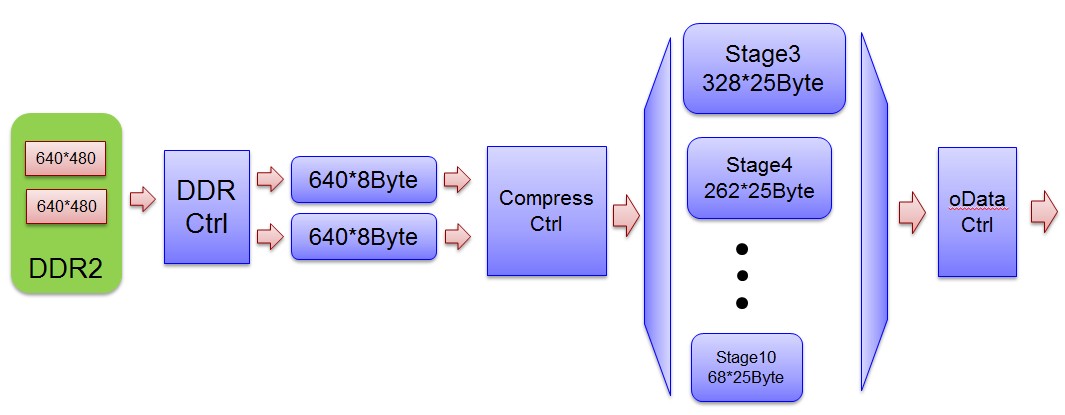


图4-12 图像缓存压缩硬件结构

在该硬件结构中：

（1）两帧图像缓存在DDR2 RAM中，并由DDR\_Ctrl模块控制作乒乓操作，保证图像数据不间断输出。

（2）同时DDR\_Ctrl模块计算图像压缩抽取的行地址，然后整行读取缓存在两块640\*8Byte的BRAM中，即缓存两行经行抽取的图像数据。

（3）经行抽取的图像数据，由Compress\_Ctrl模块进行列抽取，将抽取的数据输出给8块子窗口图像缓存BRAM。这里，从实际应用考虑，图像经8级压缩即满足手势检测需求。

（4）8块子窗口图像缓存BRAM，是为了将压缩后的图像按后一级子窗口积的大小25\*25像素输出，故行数均为25行，并由oData\_Ctrl模块控制，实现子窗口数据按1个像素步进，在8幅压缩图像遍历。

至此，图像压缩步骤完成，输入到下一级进行积分图的产生。整个模块消耗片内存储43.28KB，比原来600KB的片内存储大大减小，很好地达到了本课题的100KB以内片内存储器的要求。使后面模块设计预留了较大的裕度。

### 4.2.2积分图生成模块设计

由上一章阐述的对Adaboost图像积分方式的改进可知，本积分图生成模块是指生成25\*25像素子窗口的积分图。而窗口是按一个像素步进遍历整幅图像的，也就是说在一个像素时钟周期内生成一个子窗口积分图。这按照直观的逐像素累加的方法是无法实现的。因此，需要设计特殊的硬件结构来完成这样的运算。在本设计中，积分图产生模块分为列积分、行积分和存储阵列三个部分，每个部分包含加法器阵列及寄存器阵列，并采用流水线结构，协同完成这样积分图生成。其中，列积分的结构原理如图4-13所示：

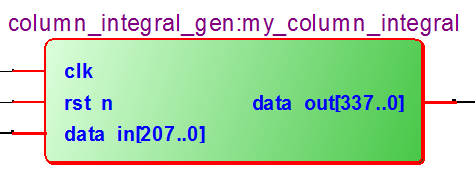
 

图4-13 列积分模块

设图4-17上每个寄存器中的值为*Ij,k*，j为横坐标，k为纵坐标，模块输入值为*il*，则*Ij,k*的表达式为式4-3。

(4-3)

行积分模块原理图如图4-14所示



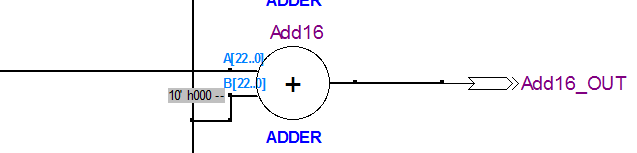
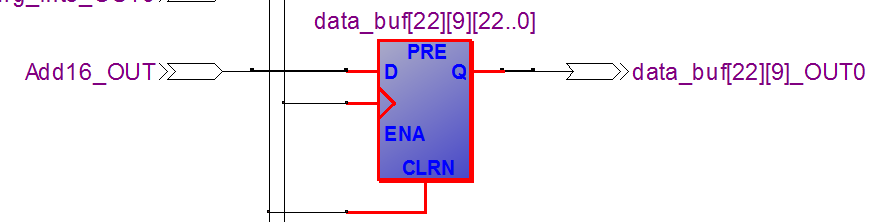


图4-14 行积分模块

行积分模块完成每行从列积分模块出来的数据的累加功能。设模块输入端数据为*il*，寄存器中当前数据为*Ik(n)*，则寄存器中下一刻的数据*Ik(n+1)*的值由式4-4得出。

(4-4)

连接列积分和行积分两个模块后，得出的数据存储在积分图阵列中。设积分图阵列中的值为*Ij,k*，原图像每个像素值为*ij,k*，则由式4-3和式4-4联合可得式4-5。

(4-5)

由上式可见积分图阵列代表原图像的积分图。接下来可以利用积分图进行特征运算。

### 4.2.3选择器组模块设计

选择器组的工作是要完成从积分图中抽出特征指示的点，然后输出到分类器组中进行运算分类。理论上特征点可能分布在整个图像上，所以选择器需要从整个积分图像上抽点。这样会造成选择器逻辑资源的大量消耗。例如积分图像的尺寸大小为20\*25，则每个点的抽取需要一个500选一的大型选择器。每个特征需要抽取12个点，即每个分类器需要对应着12个500选1的大型选择器。本设计中预计使用26个并行分类器组。所以需要312个500选1的大型选择器。这会大量消耗逻辑资源，使设计变得臃肿。考虑到检测过程中，特征值是已经训练确定的，所以每个点可以直接从积分图中抽出。按这种结构最大的消耗计算，整个检测过程一共22级，即每次需要检测的点最多有22种可能，所以每次只需要22选1，最终的系统最大消耗也就312个22选1的选择器。这

样会大大节约了逻辑资源的使用。一个22选1的选择器如图4-15所示：

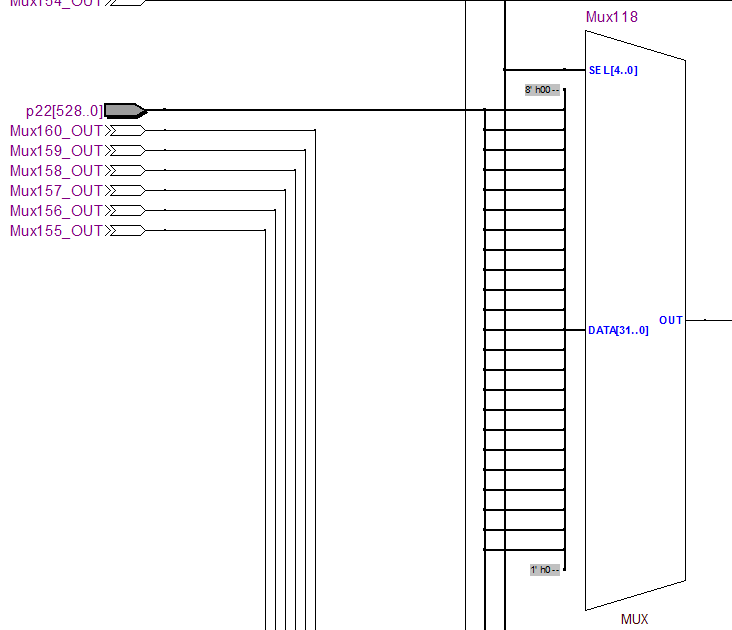


图4-15 选择器模块

### 4.2.4分类器组模块设计

分类器组要完成的工作是计算特征值并与训练阈值作对比，给出特征影响值，最后输出特征匹配结果。每个直方图的积分图像值乘以它的权重。所有直方图积分图像值乘以其权重的求和就是这个 Harr 特征分类器的结果。这个结果与特征阀值比较。如果这个结果小于特征阀值，那么这个 Harr 特征分类器的最近结果就选择left\_val，否则就是right\_val。这个最近结果称为影响因子。每个分类器中计算得出的影响因子会在模块中累加。累加值会与每级阀值比较。如果累加值大于每级阀值，那么就继续下一级以及接下来的其他级，以决定这个图像窗口可否通过所有级。每个分类器的原理如图4-16所示：



图4-16分类器模块

### 4.2.5主控模块设计

系统主控模块主要完成特征值选择，阈值选择，影响因子选择以及每级运算控制的工作，模块最终判定当前窗口是否是目标窗口，是的话一并输出当前窗口坐标值。因为系统设计成特征点值直接从积分图中抽取，所以主控模块需要根据特征值选择相应的点，其中需要建立一个译码表，记录相应的点连接在哪个选择上。如图4-17所示，选择点P11则选择端需要选择“0”。



图4-17 主控模块选择单元

阈值、影响因子和各级阈值是训练出来的常值，存放在ROM里面。如同上述抽点的操作，对于每个阈值、影响因子值和每级阈值都通过主控模块进行选择输出。

## 4.3 任意手形跟踪算法的硬件设计

### 4.3.1 肤色分割算法的硬件设计

根据第二章肤色分割算法的阐述，将肤色区域限定在一个四个线性方程组构成的容器中,在YCbCr空间中对输入的Cb,Cr值只需实现如下运算:

(4-6)

(4-7)

*Cb\* DOWN1SLOPE + DOWN1INTCPT > Cr*

(4-8)

(4-9)

其中，方程组的参数由计算机经大量样本训练得到：

*UP1SLOPE* = -1.222222222*UP1INTCPT* = 267.33333333

*UP2SLOPE* = 0.875 *UP2INTCPT* = 29.375 *DOWN1SLOPE* = -1.333333333 *DOWN1INTCPT* = 316.33333333 *DOWN2SLOPE* = -0.06451612903 *DOWN2INTCPT* = 170.61290322

对于FPGA的硬件结构，浮点数的运算无法在一个时钟周期内完成。这样，就无法实现图像视频的实时处理。对此，我们需要在一定的误差范围内将浮点数运算转换为定点数运算。在本设计中，根据其运算特点，将每个参数扩大1024倍后取其整数部分。经分析，其误差小于0.1%。则转换后的参数：

*UP1SLOPE* = -1252*UP1INTCPT* = 273749

*UP2SLOPE* =896 *UP2INTCPT* = 30080 *DOWN1SLOPE* = -1365 *DOWN1INTCPT* =323925 *DOWN2SLOPE* = -66 *DOWN2INTCPT* = 174708

通过对肤色分割算法的分析，本模块只需Cb，Cr输入，并将属于肤色的前景进行提取，输出二值图。其硬件设计相对简单，模块设计的部分代码如图4-18所示：

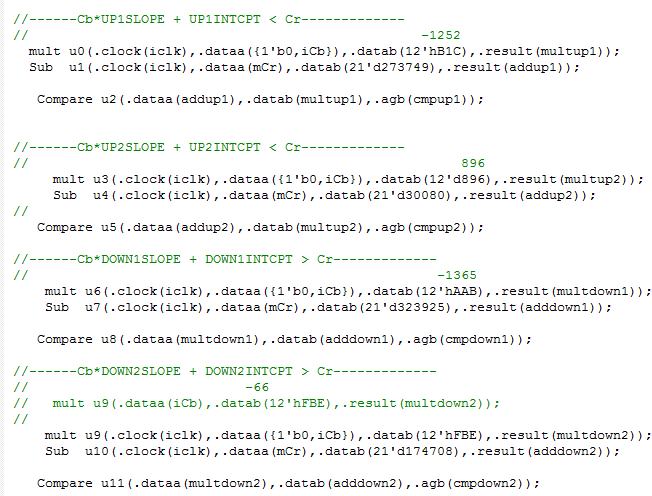


图4-18 肤色分割模块关键代码

通过视频图像的肤色分割提取的板级验证效果如图4-19所示：

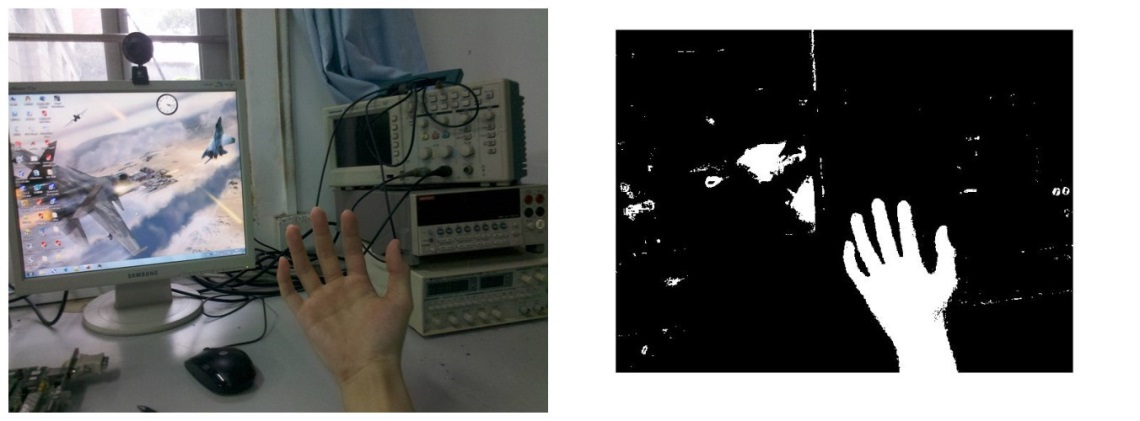


图4-19 板级验证效果图

从图4-19效果可以看出，该模块对手的分割效果明显，但是由于背景有类肤色干扰，并不能非常好地消除背景类肤色干扰，但是由于手势运动跟踪过程中，仅对包含手型的方框进行肤色统计且肤色与帧差是同时作用，因此该效果较为理想。

### 4.3.2 帧差运动检测算法的硬件设计

根据算法的功能原理，该模块的硬件设计结构图如图4-20所示。其中，速度预测模块是对速度的线性预测，主要针对出现运动过慢时，调整模块的帧间隔数及阀值参数。

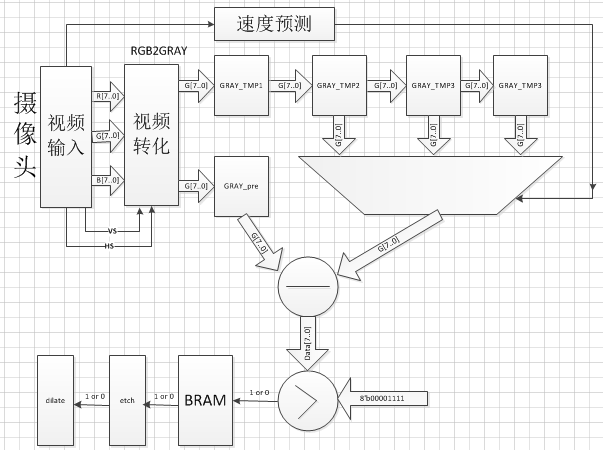


图4-20 帧差运动检测模块

经过硬件设计综合验证，速度预测会耗费大量的资源，并且，在计算机的软件验证的效果也不明显，无论使用相邻两帧的帧差还是隔着一帧做帧差效果没有影响，根据系统主要用于手势跟踪，不会出现运动过慢的情况，所以在实际应用中做了简化，如图4-21所示。其中，去除了速度预测模块，只做相邻两帧的帧差运算，并且简化了控制模块。这样，节省了不少的硬件资源。

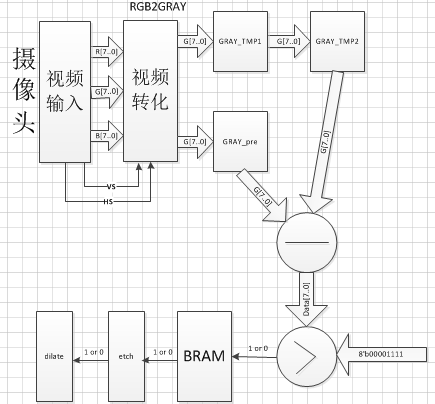


图4-21 简化后的帧差运动检测模块

设计过程中，利用SDRAM缓存前一帧的数据 ，将当前帧的数据与前一帧的数据做差并缓存，准备于后一帧数据进行对比。

帧差运动检测模块是一个相对比较简单的模块，难点在于数据的缓存，需要对SDRAM进行读写操作，主要通过视频采集、存储、回显架构中的四端口SDRAM控制器实现相邻两帧视频图像的并行读写操作。

图4-22是该模块的测试结果，通过对静止，手部运动，全身运动的对比来显示该模块的效果：



图4-22 模块验证效果

通过图4-22三幅图的比较，静止时候只有背景，效果较好。手部运动和全身运动的图片表明有别于背景的目标运动时，可以在运动目标的边角处提取运动信息。

### 4.3.3 Mean-shift质心跟踪算法的硬件设计

本章的4.3.1节和4.3.2节分别对前景提取算法模块的硬件设计进行了分析。本系统中，前景提取是为了对人手进行跟踪采集有效地数据。但要实现跟踪还需对这些有效数据进行计算得出质心点的坐标，本系统采用的算法是概率密度无参数估计方法均值偏移（Mean-shift）算法实现质心跟踪。设计过程中，通过功能模块设计及接口时序分析对该算法模块进行了设计。硬件设计框图如图4-23所示：

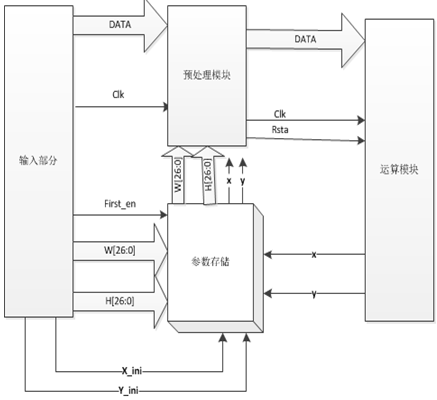


图4-23 Mean-shift质心跟踪模块

其中核心部分为运算模块，运算模块的细节如图4-24所示：

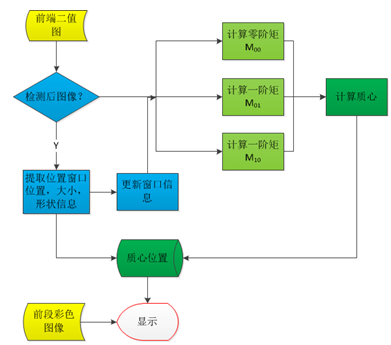


图4-24 计算流程

根据设计好的时序和模块，进行了板级调试验证，通过对连续的6帧包含有手的图片进行了验证跟踪，达到了预期结果，如图4-25所示：

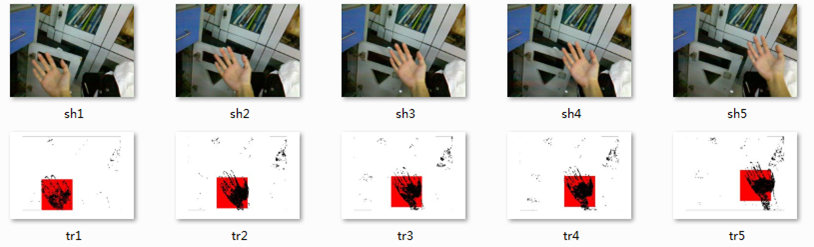


图4-25 板级验证效果

手从图片的左边缓慢移动至图片右边，红色方框为Mean-shift跟踪效果，结果显示该模块能够正常完成跟踪目标。

## 4.4 卡尔曼滤波器的硬件设计

### 4.4.1算法的硬件转化

由于FPGA属于纯数字电路，是硬件实现相关算法，因此需要对第一章所述卡尔曼滤波算法中的各项运算式向适合硬件运行的方向改写。针对FPGA逻辑电路的特性，需要将原来公式中的矩阵运算分解为代数运算，同时，尽可能地将乘、除法运算转换成位移运算。另外为了保证运算精度和时序稳定性，无法简化的多变量加减、乘除法运算需要使用专用的浮点数运算IP核。

首先，第一步的坐标预测运算中，矩阵运算可以分解为如：的代数运算。第二步中计算协方差矩阵的运算中，通过观察MATLAB运行376个样本数据的结果，发现协方差矩阵根据下面所示的对称性，可以简化为一个矩阵。

且矩阵中。

通过分解矩阵运算发现，上述两组元素确实相等。根据以上验证，卡尔曼滤波器中的5个公式，可以分解成以下代数运算式：

坐标和速度预测：

(4-10)

(4-11)

(4-12)

(4-13)

协方差预测：

(4-14)

(4-15)

(4-16)

(4-17)

如式4-15和式4-16所示，由于初始的和相等，迭代循环后所有的和相等。于是后面多处可以根据这个相等结论进行简化。

卡尔曼增益：

(4-18)

(4-19)

计算最优坐标和速度：

(4-20)

(4-21)

(4-22)

(4-23)

计算最优协方差：

(4-24)

(4-25)

(4-26)

(4-27)

由于和相等，由式4-24和式4-25可见，和相等。

为了硬件实现的便利性，用FPGA实现时，对上述代数运算作如下定义和修改。

首先，为了保证运算的精度和硬件设计相对简单，本设计采用定点数和浮点数混合运算的方式。在这种方式中，将所有的加、减法运算和能通过位移完成的乘除法运算设计成定点数运算，运算中各个变量（即矩阵中参与运算的元素）设为32位定点数，有符号整数部分和小数部分各16位。

然后，将间隙dt改为0.125，于是对该值的乘除法可以通过简单的三位右移操作完成。过程噪声矩阵中的和都是以整体参与运算的，整个运算过程中不用做乘方和除法运算，这两个数的32位定点数近似值分别为：0000000000000000\_0000000000101011和0000000000000000\_0000001000000000。在工程实现中，变量与固定数值的加法运算可以用简单的加法运算符“+”完成，综合后系统在时序上能够满足要求。而其余的变量与变量相加的运算需要用专用IP核完成，才能满足时序要求，否则综合后系统将出现无法预知的错误。

其次，为了保证计算卡尔曼增益时除法运算的精度，这里将 P00 (k)+1、 P00 (k)、P02 (k)以及和转化为单精度浮点数，使用浮点数乘、除法IP核，采用多级流水线方式，高效地完成式（4-1）到式（4-4）和式（5-1）到式（5-2）中的乘除法运算。由于系统输出只需要整数部分，另外为了节省芯片资源，其余的加减法运算将中间变量转化回32位定点数进行后面步骤的运算和最优值的输出。

### 4.4.2 硬件设计实现

本工程设计中，系统采用同步时钟设计，各系统运算模块采用50MHz时钟频率。前端手部位置信息采集模块每一帧运算得出一个二维坐标数据。视频采用640×480@60Hz的VGA制式，因此数据速率约为每16.6毫秒一个坐标数据。这里需要设计数据流控制模块，控制滤波模块分步执行各步运算。根据前一节中对各步运算的简化，整个滤波运算环路分为10步完成，各步骤运算用时不同，由于数据更新并不快，每一步都分配30个时钟周期，靠循环计数器计时，每30个时钟发出一个不同的控制信号，控制运算环路中的寄存器依次锁存前一个运算器的运算结果，为下一个运算器的计算做准备。图2-26为运算环路示意图：

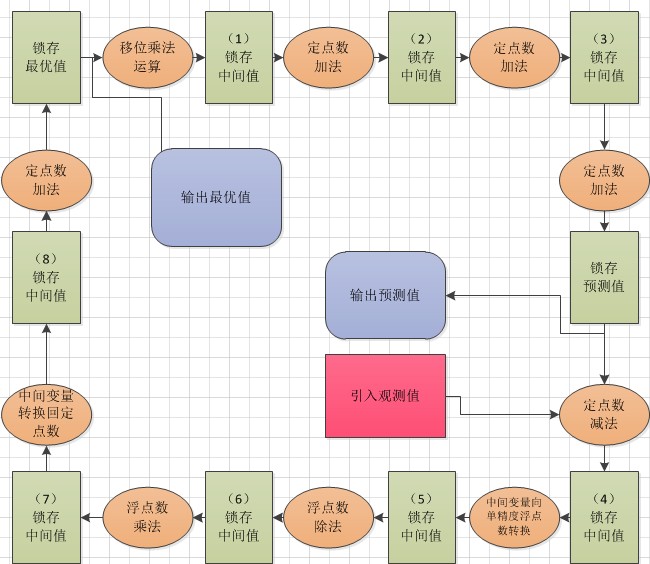


图4-26 环形运算器结构

环形运算的基本流程为：系统复位后，将最优值寄存器复位到k=0时刻的最优值，其他寄存器均复位到0值，位移乘法运算单元开始对k=0时刻的最优值进行运算，30个时钟周期后，控制模块发来第一个控制信号，寄存器（1）锁存位移乘法运算的中间值，准备定点数加法运算。再过30个时钟周期后，寄存器（1）之后的定点数加法运算完毕，下一个控制信号到来，寄存器（2）锁存运算中间值。以此类推，完成整个环路运算，控制模块复位，当下一个时刻k=1到来时，再次开始一次环形滤波运算。此时前一时刻的滤波结果已经保存在最优值寄存器和预测值寄存器中，为新时刻的滤波运算做好了准备。

根据上述思路，对环形运算器和运算控制器以及一些外部数据组合、处理模块进行Verilog HDL 建模，得到如图4-27的模块连接图。其中tmp寄存器用于对滤波数据更新信号进行边沿检测，采样到稳定的上升沿后发出高电平脉冲信号，消除信号抖动，对运算控制器复位，开始新的一个环路的滤波运算。ctrl\_port 寄存器在时钟上升沿锁存外部开关的电平信号，控制后端显示模块，选择显示滤波前的轨迹信号或者是滤波后的轨迹信号。

输入的iVGA\_VS信号为滤波数据更新信号，滤波数据在其上升沿位置更新。clk\_op为系统全局同步时钟信号，各模块在这个时钟控制下同步运行。rst\_n为系统异步复位信号，该信号的低电平脉冲使整个系统复位到初始状态。ctrl为控制输入信号，控制后端显示模块，选择显示滤波前的轨迹信号或者是滤波后的轨迹信号。obs\_x和obs\_y信号为滤波数据输入，该数据在iVGA\_VS信号的同步下更新。

输出部分各信号符合640\*480@60Hz的VGA制式。采用VGA时钟信号（oVGA\_CLOCK），VGA消隐信号（oVGA\_BLANK），VGA同步信号（oVGA\_HS，oVGA\_VS，oVGA\_SYN）和30位RGB数据。输出信号接入VGA数模转换芯片，生成驱动VGA显示器的模拟信号，显示滤波效果。

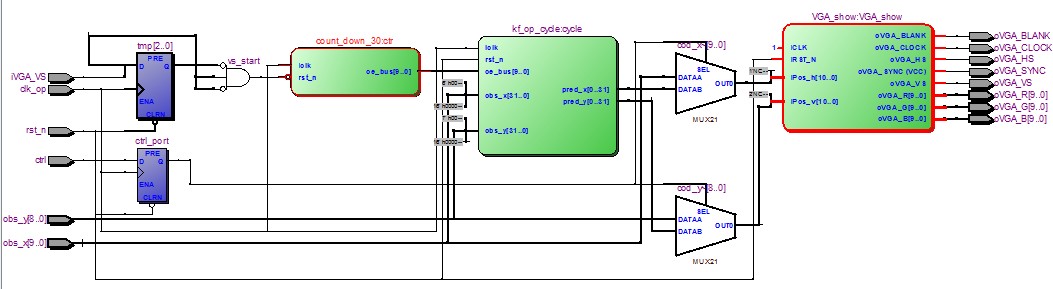


图4-27 模块间连接

## 4.5 本章小结

本章主要对整个算法系统的硬件设计进行了阐述，包括从硬件系统架构设计，算法硬件化改进，架构优化到子算法模块的硬件设计，是本文的核心部分。但受篇幅所限，本章只对关键部分设计作详细阐述。

# 第五章 硬件系统的仿真、验证与性能分析评估

## 5.1 仿真、验证、评估的方案与方法学

在一个数字系统开发中，其仿真、验证、评估是整个开发过程中极其重要的环节，甚至占整个开发周期的70%。在前面一章中具体阐述了整个硬件系统的设计过程，接下来是对所设计的模块及整个系统进行仿真、验证与性能分析评估。由于这是一个图像处理模式识别的人机交互系统，具有连续性、整体性和直观感受性，简单地从常用的仿真验证方法得到的数据是不能评估系统及其模块的功能和性能的。对此，需要针对系统的特性制定相应的及方法。

对于一套手势识别跟踪的人机交互硬件系统，除了相应运算数值结果，视觉上的视频处理效果，还包括人机交互的体验评估。那么，在整个系统及其模块的仿真、验证、评估方案中，除了包括常规的功能仿真，时序仿真及板级验证，还应与计算机联合进行跨平台测试分析，以及与应用端进行交互体验评估。整个系统及其模块的仿真、验证、评估方案的框架如图5-1和图5-2所示：

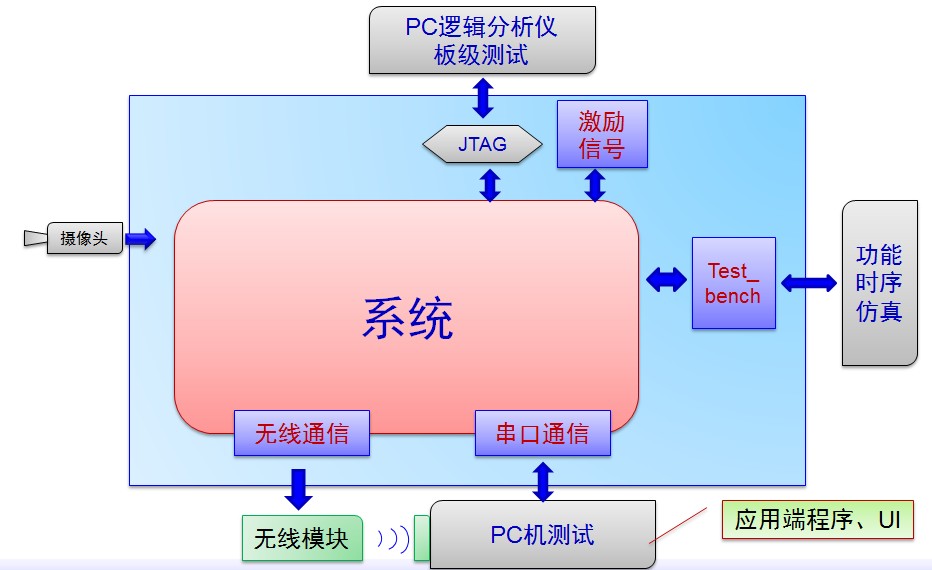


图5-1 系统仿真、验证、评估方案框架

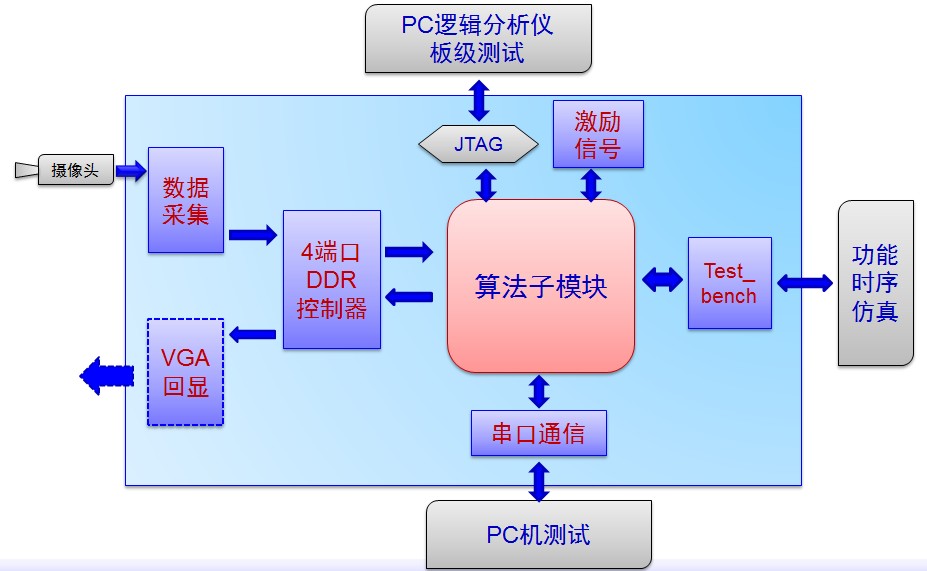


图5-2 模块仿真、验证、评估方案框架

接下来将从方法学的角度分别从两个层面分析两个模块的仿真、验证与性能分析评估的过程，以及对整个系统进行性能分析和交互评估。

## 5.2 功能仿真与板级验证

本节针对Adaboost手势检测模块及其子模块进行功能仿真、静态图片板级验证以及视频流板级验证分析。这是常规的数字系统仿真验证的手段，对于其他模块的功能仿真与板级验证的方法也是类似的。

仿真验证主要是应用Modelsim工具，编写仿真测试脚本testbench，进行模块功能验证，时序调节以及结果验证上。板级验证主要借助片上逻辑分析仪SignalTap II及计算机软件，其中静态图片验证主要完成对静态图片的检测，看接侧结果的精度以及误识别率。对视频流的验证是模块最终实现的功能，最终结果与软件系统相对比，分析两者之间的不同。

### 5.2.1部分功能模块仿真结果

积分图模块仿真结果如图5-3所示。测试图像为像素值全为1的定制图像，A图显示列积分模块输出结果，B图显示积分图阵列结果。



图5-3(A) ModelSim仿真

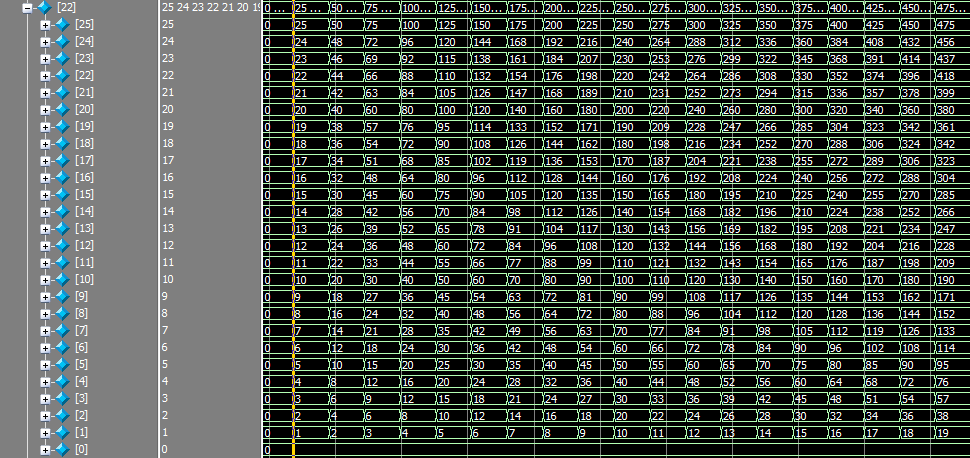


图5-3(B) ModelSim仿真

分类器模块仿真结果如图5-4所示。A图测试结果显示分类器运算的运算过程，由仿真图像可看出分类器流水线运算的特点和检验运算结果的正确性。B图测试结果是检测一副35x26的包含人手的图像为测试图像得出的结果，由仿真图像可看出人手被检测出来。C图测试结果是检测一副35x26的不包含人手的复杂背景图像作为测试图像得到的结果，由图可看出没有人手的窗口中不会检测出人手。

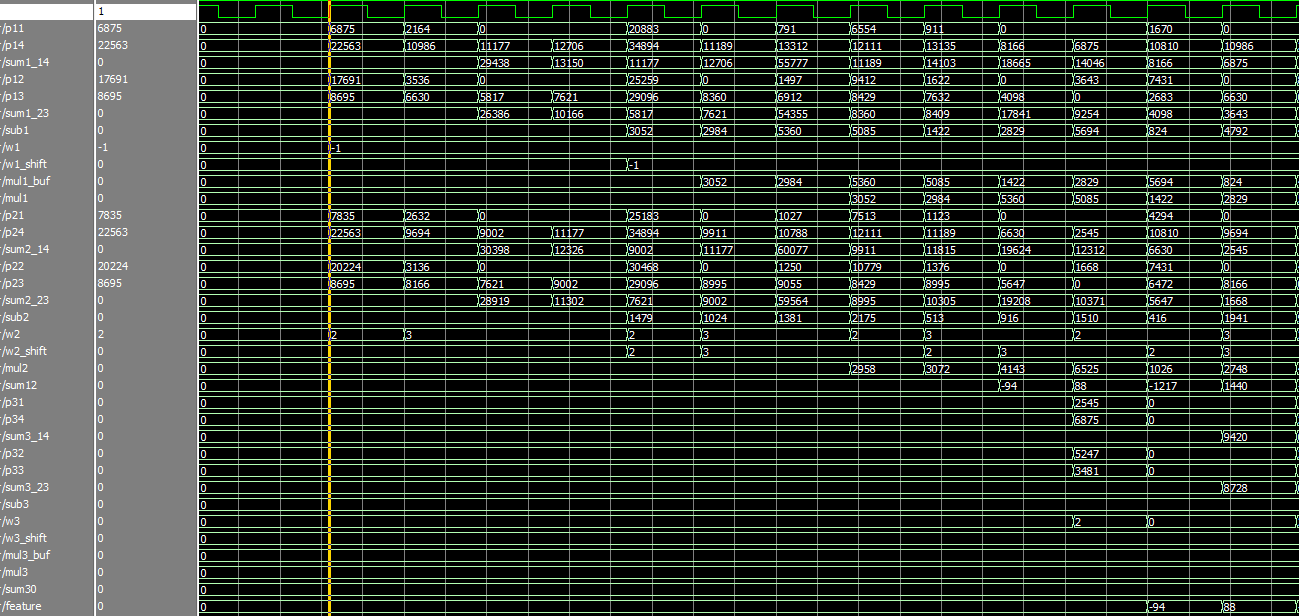


图5-4(A) 分类器仿真

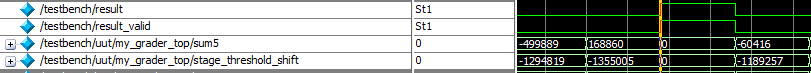


图5-4(B) 分类器仿真

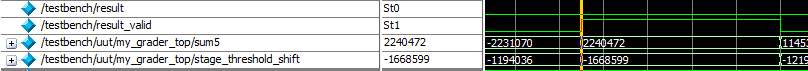


图5-4(C) 分类器仿真

主控模块仿真结果如图5-5所示。根据不同的特征，主控状态机在每一级会输出不同的经过译码后的选择值。因为特征值是训练后固定的，所以在主控模块中一次检测过程中输出的值是固定的，不会因为测试图片而改变。

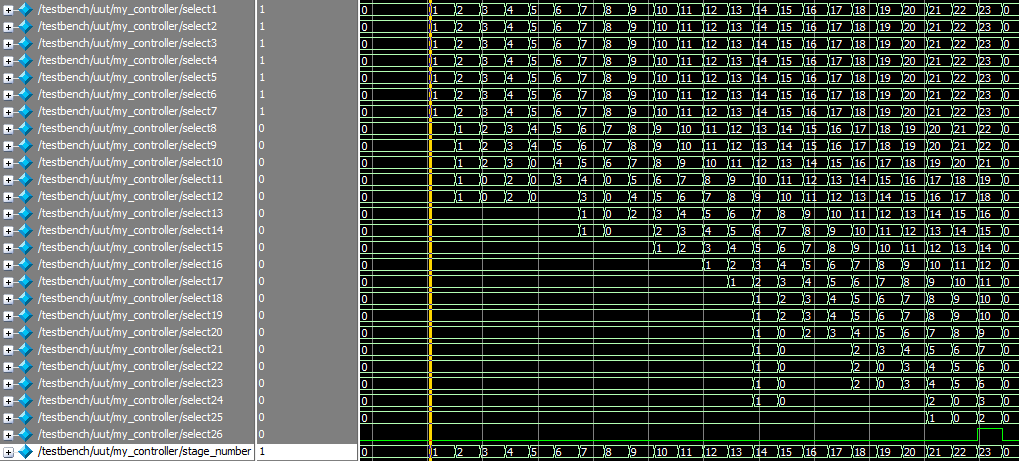


图5-5 主控模块仿真

### 5.2.2静态图片板级验证结果

采用静态图片板级验证的是进行视频流验证的基础。静态图片验证成功才能保证输入视频流时有成功实现的可能。进行静态图片验证只需要把事先采集好的图片放进板上的BRAM中，运行图像处理部分系统即可。如图5-6，图5-7，图5-8，图5-9所示为静态图片验证结果。蓝色框框位置是根据在线逻辑分析仪工具SignalTapII采集到的数据标注出来的。由图可见方框出现的位置就是人手的位置。

图5-6是手正直呈现，光照正常的情况下得到的测试结果。

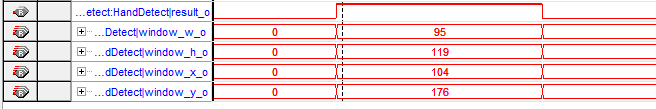




图5-6 静态图像板级验证

图5-7是的摆放角度倾斜，光照正常的情况下得到的测试结果。

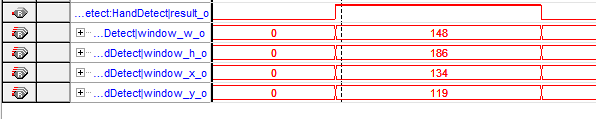
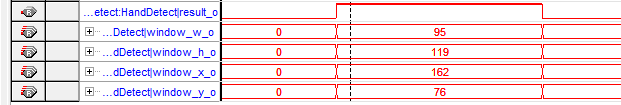


图5-7静态图像板级验证

图5-8是手部动作变形，光照正常的情况下得到的测试结果。



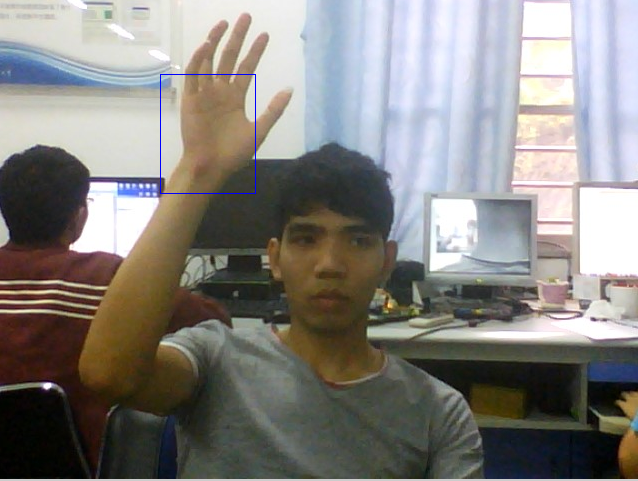


图5-8静态图像板级验证

图5-9是在光照条件不好的情况下得到的结果。

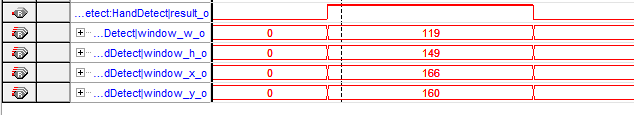




图5-9静态图像板级验证

另外，在本实验中，还做了光照条件更差，手部变形更厉害的静态图片的测试，测试结果显示检测不出人手。

### 5.2.3视频流板级验证

完成视频流的板级验证才是测试最终所要达到的目标。验证过程需要用到视频采集以及回显模块。通过摄像头采集视频流，然后用通信模块发送到图像处理模块的前端缓存，接着进行视频图像的处理——人手检测，检测结果从回显模块中与图像一并显示。如图5-10，图中红色框标注的位置为系统检测出的人手的位置。

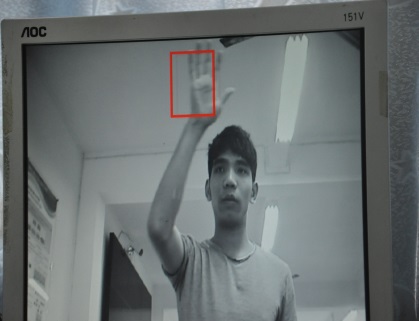
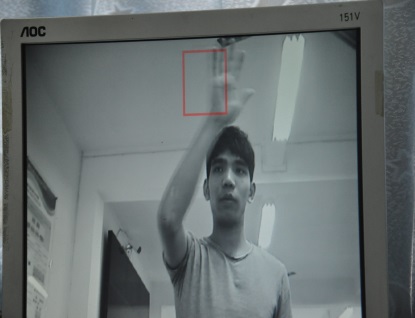
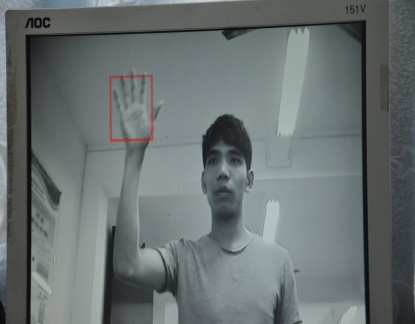
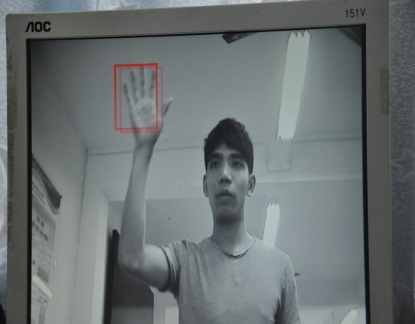
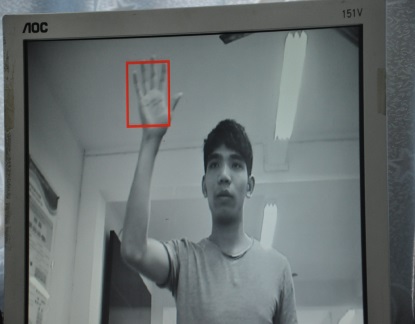
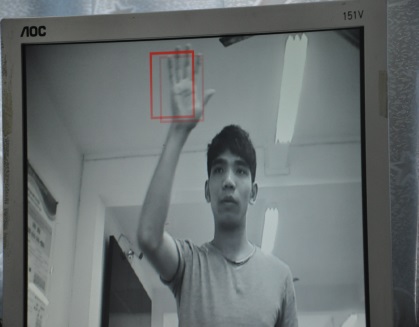


图5-10 视频流板级验证

图5-11显示在复杂背景下的测试结果。左图显示识别错误的情况，右图显示复杂背景下仍能识别正常的情况

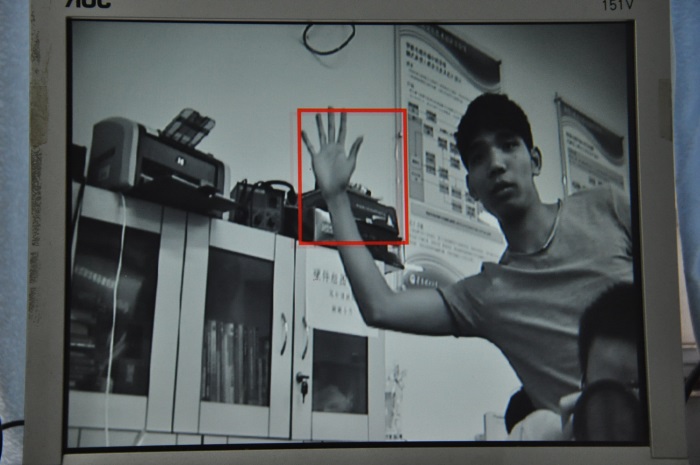


图5-11 复杂背景下测试

### 5.2.4仿真验证结果分析

1、部分模块仿真结果分析：

由图5-3、图5-4和图5-5可见，积分图模块、分类器模块以及主控模块功能仿真阶段测试结果跟预期结果相吻合，功能实现正常。

2、静态图片板级验证结果分析：

由图5-6，图5-7，图5-8可见当光照正常时，手正直呈现、在一定角度内偏斜、手部动作在一定程度内变形的情况下都是可以正确识别的。由图5-9可见，即使在光照条件不好的情况下，只要不低于某个阈值，算法还是能够通过光照补偿纠正过来，得到正确的识别效果。

3、视频流板级验证结果分析

图5-10是一段视频之中截取的一系列图片。在这一串图片中，由前大半部分可见在手正直或者在一定角度内旋转的情况下还是可以检测到手，但在后面几张看到手部旋转过大，系统检测不出手。

图5-11是显示复杂背景下系统的识别效果，由左图可见背景太复杂，几时光照条件良好，手部正直呈现，仍出现了检测错误的情况，检测出来的框显示在人脸附近。由右图可见在复杂背景中仍能检测出人手的存在。对比两幅图同样在复杂背景中出现的两种不同的结果，能得到结论如下：

1.背景的复杂度对检测效果有影响；

2.越复杂的背景，检测的效果越差；

3.类肤色的背景对检测效果影响很大。

### 5.2.5性能分析

本模块运行在分辨率为640\*480的视频中，达到30fps的时实检测速率，与视频的帧率一致，其运行的最高频率为108MHz。在该模块中，包含23级分类器，共700个Harr特征，采用行步进2像素，列步进1像素，进行10级压缩图像检测。

基于Adaboost算法的目标检测的硬件实现，在国内外均有典型的系统，表5-1是本系统的该模块与一些现有典型系统的参数对比

表5-1 与现在典型系统参数对比

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 平台 | 机构/学校 | 分辨率 | 特征数 | 频率 | 检测帧率 |
| 纯软件系统 |  | 640\*480 | 2135 | 2.8GHz | 5-6fps |
| SOPC | 华南理工大学 | 384\*228 | 2135 | 100MHz | 22fps |
| FPGA | 华南理工大学 | 640\*480 | 128 | 200MHz | 15fps |
| FPGA | 上海交通大学 | 176\*144 | 2135 | 100MHz | 50fps |
| FPGA | 卡耐基梅隆大学 | 640\*480 | 52 | 126MHz | 143fps |
| FPGA | 加利福利亚大学 | 640\*480 | 2153 | 不详 | 7.5fps |
| FPGA | 华南理工大学  （本系统） | 640\*480 | 700 | 108MHz | 30fps |

经对比分析，本模块在较分辨率下仍能达到较高的检测速率，其整体性能达到较优水平。另外，本模块与现在的典型的最大优势是其片内存储器仅使用了54.14KB，远低于上述系统。这就使得本系统在芯片化过程中可以控制其面积和成本，很好地推动系统实用化进程。

此外，本模块在设计过程中采用了参数化的设计思想，其当前30fps的检测速率是与摄像头的帧率同步，并不是该模块的极限速率。对于其最高频率仅有108MHz而言，本模块的检测速率尚有不小的提升空间，可根据摄像头的帧率的提高，调整设计参数，从而在一定范围内提高检测速率。

## 5.3 跨平台联合测试与分析评估

对于一些算法模块，如肤色分割、帧差运动检测、卡尔曼滤波等模块，在完成常规的功能仿真、板级验证后只能验证模块能正常运行，其运行效果和性能评估也是非常重要的。如何客观科学地评估性能和运行效果是本节就研究重点，除了主观地进行视觉评估外，重点采集有效数据进行定量数据分析对比，运用包括与软件运行数据纵向对比分析，以及进行方差分析等手段。这里需要进行跨平台软硬件协同测试与分析评估。本节将以卡尔曼滤波模块为例，对整个过程进行阐述。

### 5.3.1直观、定性地分析硬件实现效果

通过Quartus II软件综合与布局布线，生成了FPGA的烧录文件，测试平台采用Altera公司的DE2-70开发板，本设计用到该开发板上的Altera Cyclone II 2C70 FPGA芯片、用于下载配置的USB Blaster控制单元、一个用于选择显示信号的栓扣开关和驱动VGA显示器的VGA数模转换器。在评估测试过程中，还将用到232电平转换芯片和RS232串口。用于FPGA卡尔曼滤波器测试的轨迹数据与MATLAB使用的轨迹数据相同，这里使用只读存储单元IP核，在编译综合的过程中，将MATLAB中的轨迹数据写入FPGA用IP核实现的内嵌ROM存储单元，以上文所述的VGA帧率相同的时钟（频率约为60Hz）为数据时钟，依次读取ROM单元中的数据做滤波。图5-12分别为滤波前后小方块在二维空间运动的轨迹，滤波后小方块跳动明显减小，轨迹非常平滑。

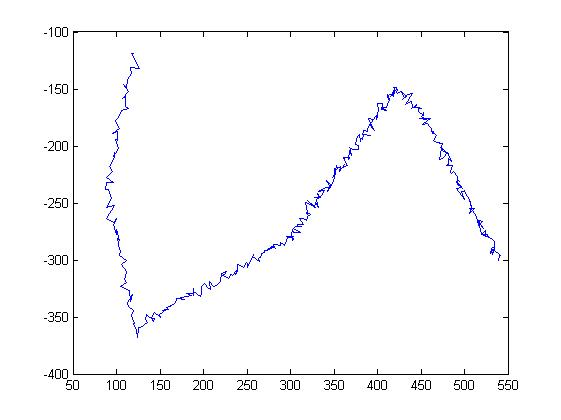
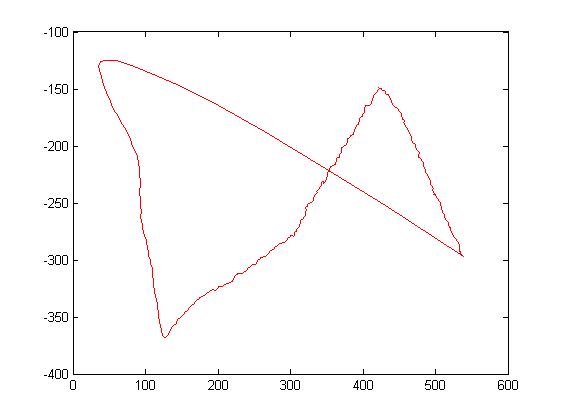


图5-12 FPGA卡尔曼滤波原始轨迹及滤波后轨迹

在非接触式人机交互系统中，手部的运动无法保持匀速直线运动，但是在一段较短的时间内的运动可以看作是线性的，如果将整个复杂的手部运动分解为多个匀速运动的线性组合，这样是近似符合卡尔曼滤波器的假设。所以，本设计中的卡尔曼滤波器可以用于非接触式人机交互系统中手部运动轨迹的平滑。

### 5.3.2 滤波器性能的定量分析

设计基于FPGA的卡尔曼滤波器，是通过将PC端运行的软件滤波器算法通过适当的修改和简化，在保证精度的前提下，使其计算复杂度降低或适合数字系统的硬件实现。因此，将PC端软件运行的滤波结果和硬件FPGA实现的数字系统滤波结果做定量对比将是评估FPGA数字系统实现的滤波器非常有效的手段。

通过FPGA将各点滤波时的协方差通过串口或者Quartus II软件中的逻辑分析仪工具传回PC，将这些数据用MATLAB中编写的程序进行处理和分析，便可以定量地分析、评估FPGA数字系统实现的卡尔曼滤波器与PC软件实现的滤波有何性能上的差别。

#### 5.3.2.1 FPGA传回滤波数据

FPGA设计中考虑到不是每个测试场合都有Quartus II的逻辑分析仪，本设计中加入串口模块，通过RS232串口线将滤波数据发给PC，PC端用串口调试助手等软件接收并将数据保存为excel表格。图5-13所示的是采集到的局部滤波数据。

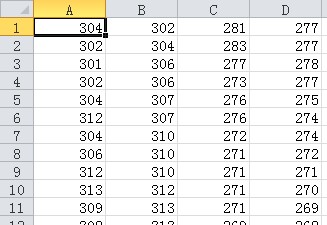


图5-13 PC采集到的部分滤波数据

通过编写MATLAB程序，只需要将上述表格放在MATLAB运行目录下，运行名为“DRE”的m文件，便能自动画出原始轨迹和FPGA完成的滤波轨迹。如图5-14所示。

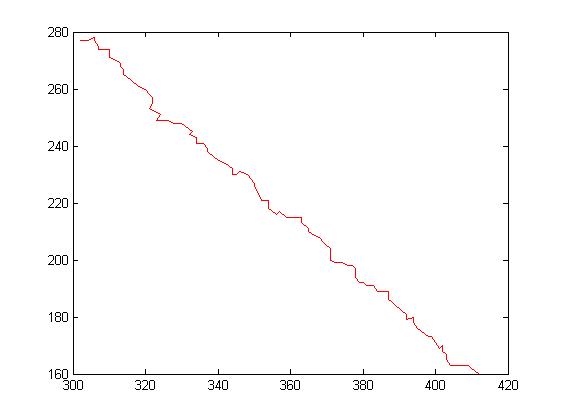
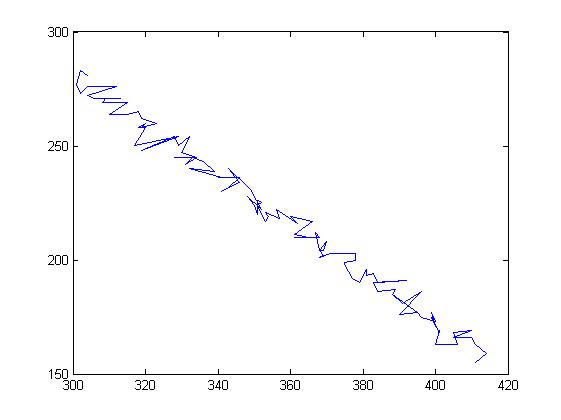


图5-14 PC端显示的FPGA滤波结果，各轴为平面坐标，无单位

#### 5.3.3.2数据的定性分析

滤波效果的好坏，一方面可以通过观察滤波前后的轨迹，直观地判断，但是这样的方法无法分辨出较细微的差别，对两种方法实现的滤波器无法做出比较。滤波效果的好坏，可以通过对比输出信号与需要信号之间的差值来判断。就本设计中的卡尔曼滤波器而言，就是运算中使用的协方差矩阵中的元素。通过FPGA卡尔曼滤波器中的串口模块或者Quartus II软件中的逻辑分析仪，我们可以得到滤波过程中对各点运算的协方差，通过MATLAB程序“RdAD”可以将FPGA卡尔曼滤波器的各步协方差数据和PC端MATLAB运行的卡尔曼滤波器各步协方差数据相对比，画出两者协方差变化的过程图像，以及两者协方差在对应位置的差值的图像。图5-15为FPGA卡尔曼滤波器运行时各步的协方差数据图像，可以看出，随着滤波过程的推进，协方差数据呈现出收敛状态。图5-16为PC端MATLAB运行卡尔曼滤波的结果，协方差数据同样呈现出收敛状态。图5-17是两者协方差之间的差值图像，由图像可见，随着协方差的收敛，两种方式实现的滤波协方差的差值趋于稳定。稳定的差值是由于FPGA滤波中采用32位定点数运算，局部采用单精度浮点数运算，数据格式在转换过程中存在固定的转换误差。

图5-15 FPGA卡尔曼滤波协方差数据

横轴为坐标点编号，纵轴为协方差值，无单位

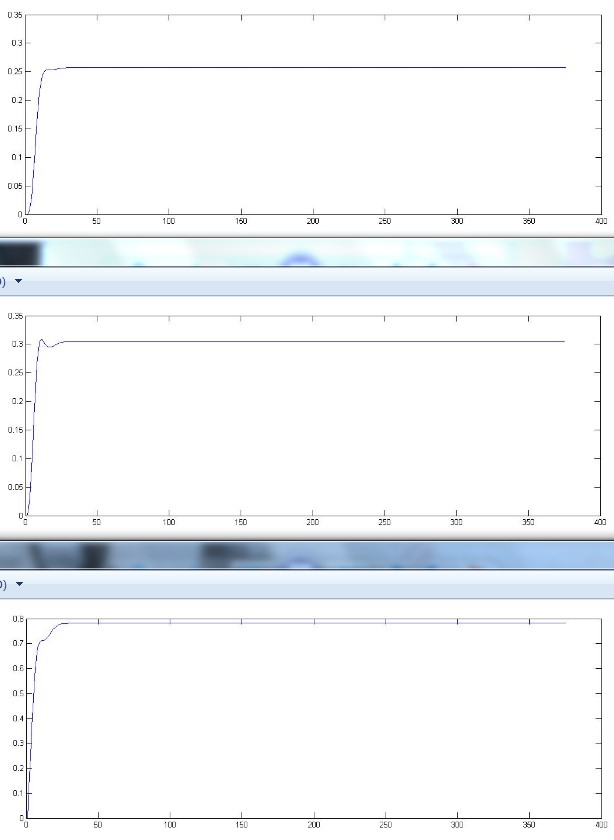


图5-16 MATLAB卡尔曼滤波协方差数据

横轴为坐标点编号，纵轴为协方差值，无单位

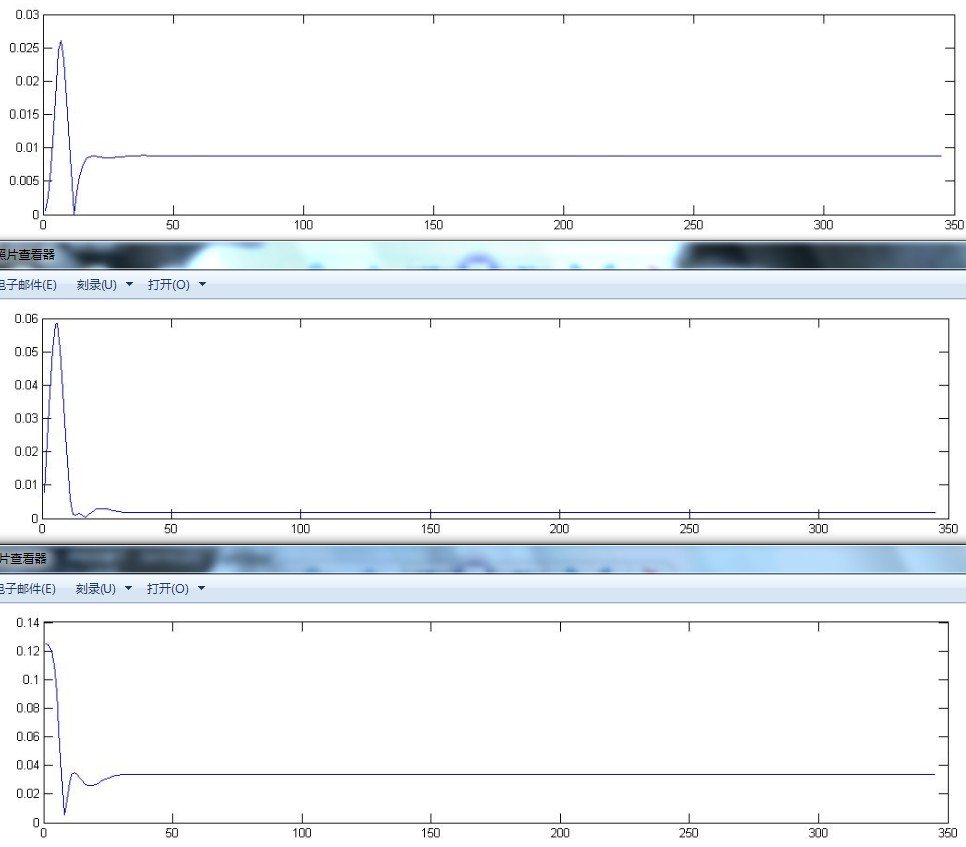


图5-17 FPGA卡尔曼滤波与MATLAB卡尔曼滤波协方差之差值

横轴为坐标点编号，纵轴为两种方式实现的协方差的差值，无单位

#### 5.3.3.3定量分析FPGA卡尔曼滤波器性能

将FPGA完成的卡尔曼滤波协方差数据上传到PC机查看其收敛特性，可以定量的分析FPGA卡尔曼滤波器的性能。观察数据可以发现，用PC机运行MATLAB程序实现的软件卡尔曼滤波和FPGA实现的卡尔曼滤波都可以在25个滤波时刻达到协方差数值的收敛，以后的协方差数值变化都小于0.001。另外观察到，稳定时，两组协方差数值分别有0.0088（3.4%），0.0018（0.59%）和0.0338（4.3%）[[14]](#footnote-1)的偏差。通过分析可以知道，这个误差是FPGA定点运算引入的。由于FPGA简化运算采用32位定点数（小数部分16位）和单精度浮点数混合运算，相对与MATLAB全双精度浮点运算必然有一定误差。而且运算的初始值是用0.000656的近似值替代其精确值的。由于简化运算，这个必然引入的误差在实际工作中带来的影响是非常小的，对于人机交互应用来说，这个较小（5%以内）的误差不足造成起手部跟踪系统的性能下降，因此完全可以接受的。而在对速度要求更高的场合，完全有余地扩大这个误差，以进一步简化运算，达到更高的运算速度。

## 5.4 系统性能分析及交互评估

在完成了各模块的仿真、测试、评估之后，将整个硬件系统整合，进行系统的测试和评估。包括系统的运行性能和硬件性能的分析，以及与PC应用层对接后的人机交互体验分析评估。

### 5.4.1 系统的运行性能分析

首先是对系统的运行性能进行分析评估，主要从系统运行的流畅度、稳定性，以及使用者的操作体验等方面进行评估，是一种定性的分析评估。图5-18是系统运行的整体效果图。



图5-18 系统运行

接下来从“手势识别🡪获取控制权🡪任意手形跟踪漫游🡪定位选中🡪跟丢后迅速重新捕捉🡪取消控制”这样一个完整的交互过程分析系统的运行性能，如图5-19至图5-28所示：

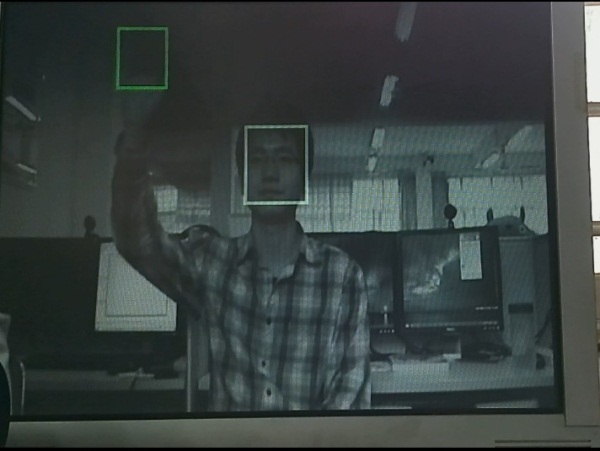
 

图5-19 初始化获取控制权 图5-20 手势漫游

图5-21 脸部区域漫游 图5-22 选中

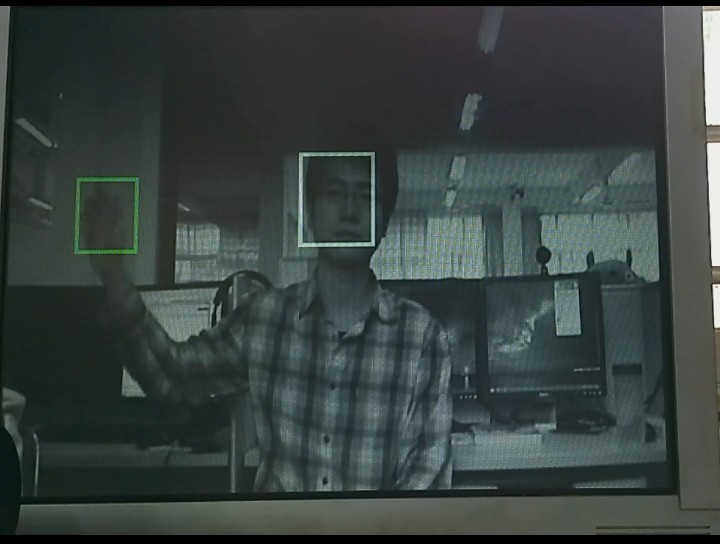
 

图5-23 任意手形漫游 图5-24 跟丢

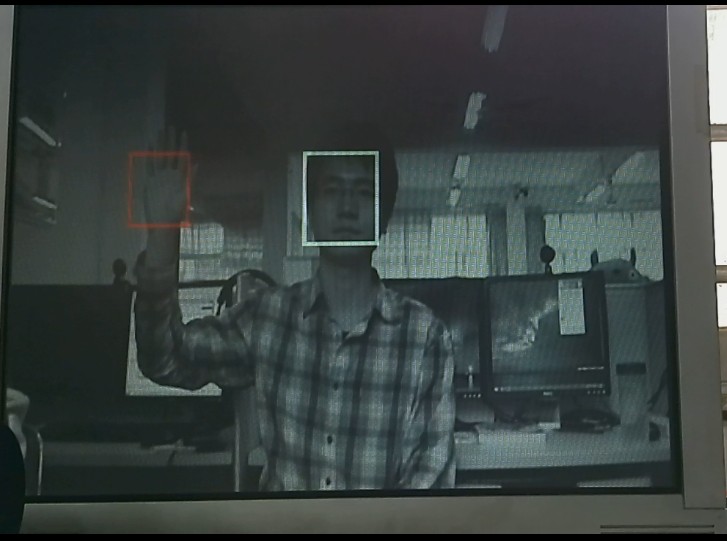
 

图5-25 重新捕获人手 图5-26 重新开始漫游

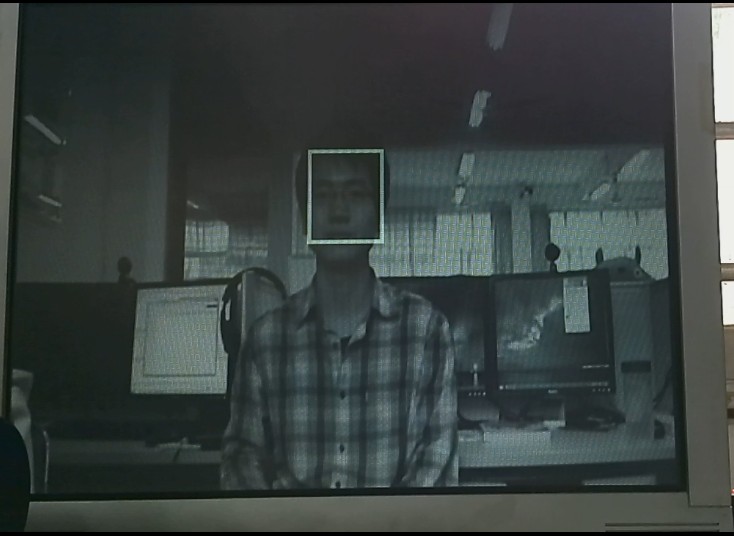
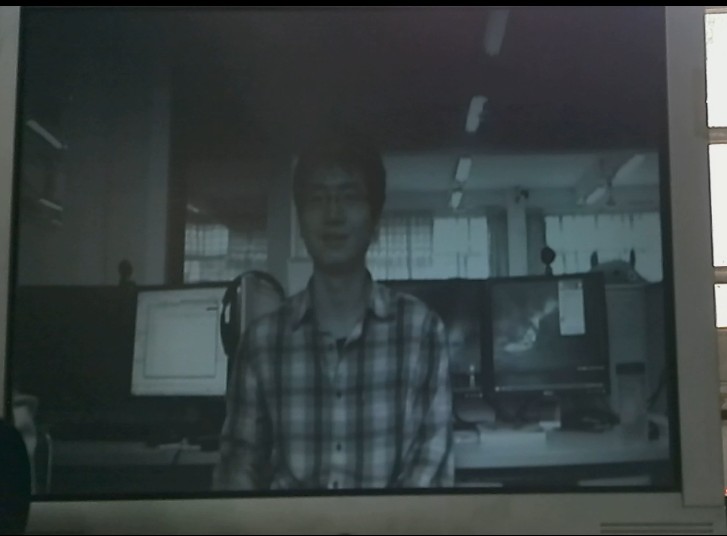
 

图5-27 人手离开 图5-28 撒消控制权

在系统中，用不同颜色的方框表示不同的运行状态以便观察。其中白色方框表示人脸区域；绿色方框表示系统在正常区域跟踪；蓝色方框表示在脸部区域跟踪，此时只有帧差运动检测算法进行迭代跟踪；黄色方框指手势在空中悬停表示选中，若悬停超过10秒则自动取消控制权；红色方框表示在跟踪过程中跟丢后启动Adaboost算法重新捕捉人手。整个跟踪操作过程简单自然，特别是任意手形的跟踪，符合人的动作习惯，使得人能与机器能够进行一个自然的交互。

### 5.4.2 系统的硬件性能分析

系统运行性能达到预期后，接下来从硬件资源消耗、FPGA等效门、系统频率等分析系统的硬件性能。表5-2列出了系统的资源消耗，表中列出了FPGA主要的四种硬件资源的使用情况，分别列出了验证平台每块开发板的资源使用情况以及整个系统的消耗的硬件资源，并与DE4-230开发板的FPGA片内资源比较。由于没类似的系统进行比较，这里只能做一个定性的分析。其ASIC的等效门经估算后约为600至800万门级，是一个接近千万门级的ASIC芯片的规模，相对于其算法系统的复杂度及当前的工艺制造水平，这样的规模已比较常见，其制造成本主要取决于产量。



表5-2 系统硬件资源使用

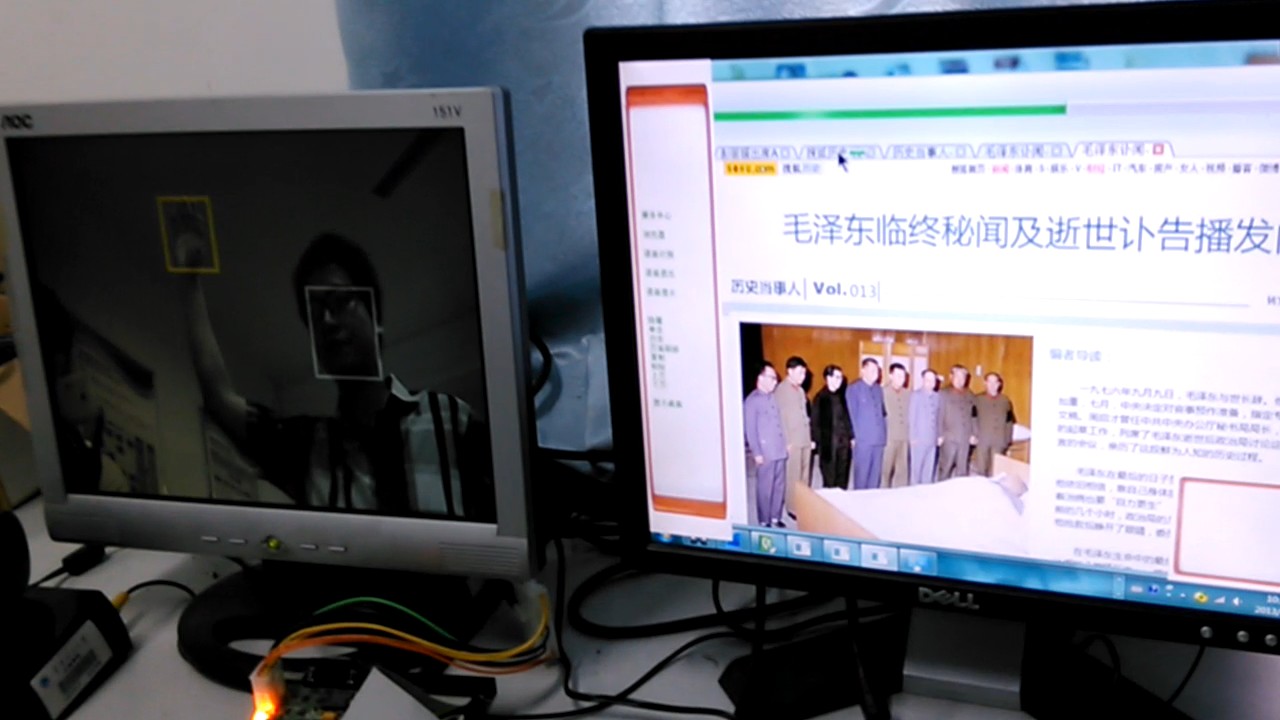
在系统的设计过程中，该系统按同步的思想进行设计，以视频的帧率为处理周期单位，以视频像素时钟27MHz为系统的时钟基频，整个系统最高时钟频率的模块是分类器的运算模块，仅为108MHz。由此可见，系统依据视频帧率（30f/s）实现实时跟踪，108MHz的最高频率尚有较大的提升裕度，可随视频帧率在一定范围内提升（如60f/s），以达到更佳的跟踪性能。

### 5.4.3 系统的交互分析评估

对于一个人机交互系统，用户的体验感受至关重要。本小节中，系统将与应用层的程序对接，即一个Web浏览器，进行非接触人机交互在浏览器中漫游操作的应用测试。

由于在本课题中，非接触人机交互的体验评估尚缺乏一套定性定量的分析方法，本小节仅作定性的感官体验的直观分析评估，以及与该算法系统在计算机上运行的效果对比分析。图5-29是系统与应用层对接进行浏览器中漫游操作的效果图。在本测试中，通过UDP协议将硬件系统输出的位置信息和相应控制指令传输到计算机，将手势在空中的漫游映射为计算机中的鼠标漫游。





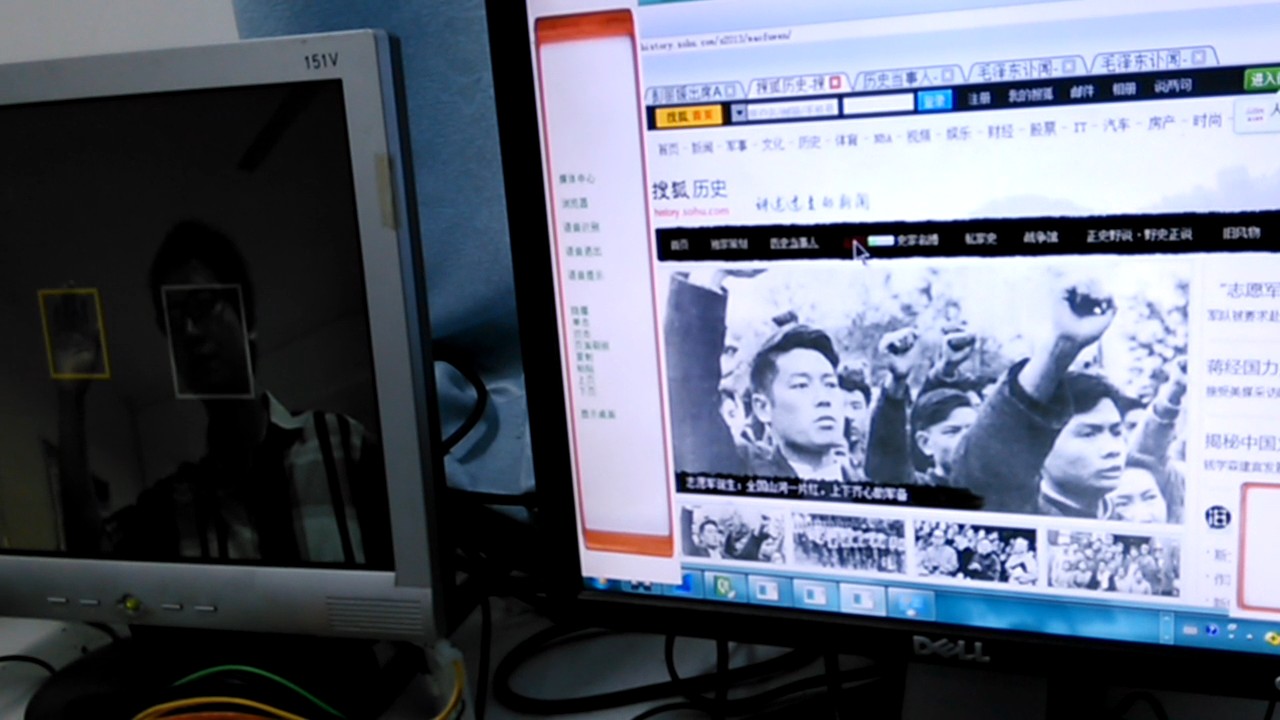


图2-29

本硬件系统中，由于是以视频的帧率为处理周期单位，每帧输出一个迭代跟踪结果，这与计算机运行中非周期的运算时间相对，其漫游过程更加平滑，这就使得本硬件系统在交互体验上优于计算机系统。并且，在悬停选中操作中，本系统采用消抖的一些设计技巧，使得交互操作显得平稳，交互体验更佳。

不过，在本交互体验测试中，非接触人机交互这种空间立体的交互方式仍映射为平面的交互，与鼠标漫游相比就显得繁冗吃力，未能体现非接触人机交互的优势。因此，在交互方法上尚有待理论的突破，让人拭目以待。

## 5.5 本章小结

本章从方法学的角度对基于图像处理模式识别的手势跟踪的硬件系统的功能仿真、板级验证、性能分析、交互评估进行详细地分析阐述。是针对特定的硬件系统而设计的一套测试评估方案，是对常规的数字系统仿真测试方法的一种扩展，实现了跨平台联合测试，是对智能机器视觉类的数字系统在测试评估上的一个有益的创新探索。

# 总结与展望

## 工作总结及意义

本论文研究学习了一套成型手势识别跟踪算法系统，对其进行硬件化设计，旨在推动非接触人机交互的实用化产业化。在整个设计过程中，贯彻了“自顶向下”的设计思想，首先对整个算法系统理解透彻后，对算法进行硬件化改进以及从算法模型向硬件模型抽象转化；接着进行整个硬件系统架构设计定型，对系统存储架构作了较精确的估算和布局，确定模块划分，定义接口标准和接口时序标准。在完成顶层的设计定义后，进行模块的设计，RTL代码编写以及功能仿真。标准化模块化的设计理念使得整个设计具有可测试性，易维护性，可升级性以及方便了模块间通信。特别是前期工作中开发的标准化的视频采集、存储、回显的架构，使得各算法模块可嵌入该平台进行板级测试。

在系统及其模块的仿真、验证、评估中，本论文从方法学的角度对基于图像处理模式识别的手势跟踪的硬件系统的功能仿真、板级验证、性能分析、交互评估进行详细地分析阐述。是针对特定的硬件系统而设计的一套测试评估方案；是对常规的数字系统仿真测试方法的一种扩展，实现了跨平台联合测试；是对智能机器视觉类的数字系统在测试评估上的一个有益的创新探索。

最后系统一套由DE4-230和DE2-70以及一台工业摄像机云台搭建的FPGA验证平台和计算机应用层终端联合的验证测试评估平台中运行，在视频分辨率640\*480图像中，以30f/s进行实时迭代跟踪，实现“手势识别🡪获取控制权🡪任意手形跟踪漫游🡪定位选中🡪跟丢后迅速重新捕捉🡪取消控制”这样一个完整的交互过程。

本课题为非接触人机交互系统的芯片化、实用化作出了探索和验证，在可查询的范围内尚无同类的硬件系统，是该领域的先行者，为推动非接触人机交互系统产业化做出有益的探索。

## 后续工作展望

虽然本系统能够实现一个完整的非接触人机交互，并且能与交互应用层对接，初步具备实用性，但仍无法应对复杂背景的现实场景。这是由于系统采用了肤色分割与帧差运动检测算法作为目标跟踪的前景提取算法，从实现效果看，当采用肤色跟踪时，容易受到类肤色干扰。并且由于参数固定，在光线变化的情况下，效果并不理想。在人脸附近采用了帧差运动检测跟踪，减少了人脸的肤色对跟踪的影响，也取得了相对较好的效果，但是帧差运动检测同样受到背景其他类似人或物移动的干扰。这两点是本系统明显的缺陷。在今后还需进一步研究，实现结合类似混合高斯模型算法（GMM）、增量子空间、光流运动目标跟踪等抗干扰能力强的算法的硬件系统设计，使非接触人机交互系统具备真正的实用性，为非接触人机交互系统产业化做出实质性推动。

# 参考文献

1. 百度百科 [↑](#endnote-ref-1)
2. 刘畅. 基于手势的非接触式人机交互中用户疲劳的评价与研究[D]. 华南理工大学, 2013. [↑](#endnote-ref-2)
3. 邓荣. 智能移动式果蔬采摘机器人设计——基于SOPC神经网络[J]. 农机化研究,2016,08:184-188. [↑](#endnote-ref-3)
4. 杨光,潘瑞云,蒋迺倜,尚文明. OSGO-CFAR算法检测性能分析及FPGA实现[J]. 雷达与对抗,2015,03:41-44. [↑](#endnote-ref-4)
5. 杨阳,刘剑,蒋廼倜,李赛辉. 一种基于FPGA的万兆光纤以太网高速传输方法[J]. 雷达与对抗,2015,03:23-27. [↑](#endnote-ref-5)
6. 周光宇,刘慧忠. 边缘检测算法的FPGA实现[J]. 计算机系统应用,2015,10:271-275. [↑](#endnote-ref-6)
7. Yamaoka, K.; Morimoto, T.; Adachi, H.; Koide, T.; Mattausch, H.J., "Image segmentation and pattern matching based FPGA/ASIC implementation architecture of real-time object tracking," in *Design Automation, 2006. Asia and South Pacific Conference on* , vol., no., pp.6 pp.-, 24-27 Jan. 2006 [↑](#endnote-ref-7)
8. Jie Zhu; Zhiqian Chen, "Real Time Face Detection System Using Adaboost and Haar-like Features," in *Information Science and Control Engineering (ICISCE), 2015 2nd International Conference on* , vol., no., pp.404-407, 24-26 April 2015 [↑](#endnote-ref-8)
9. Fan Zhao; Guizhong Liu; Xing Wang, "An efficient AdaBoost tracking algorithm based on the particle framework," in *Signal Processing, Communications and Computing (ICSPCC), 2011 IEEE International Conference on* , vol., no., pp.1-4, 14-16 Sept. 2011 [↑](#endnote-ref-9)
10. Bilal, S.; Akmeliawati, R.; Salami, M.J.E.; Shafie, A.A.; Bouhabba, E.M., "A hybrid method using haar-like and skin-color algorithm for hand posture detection, recognition and tracking," in *Mechatronics and Automation (ICMA), 2010 International Conference on* , vol., no., pp.934-939, 4-7 Aug. 2010 [↑](#endnote-ref-10)
11. Annan Li; Shuicheng Yan, "Object Tracking With Only Background Cues," in *Circuits and Systems for Video Technology, IEEE Transactions on* , vol.24, no.11, pp.1911-1919, Nov. 2014 [↑](#endnote-ref-11)
12. Dongxiang Zhou; Hong Zhang, "Modified GMM background modeling and optical flow for detection of moving objects," in *Systems, Man and Cybernetics, 2005 IEEE International Conference on* , vol.3, no., pp.2224-2229 Vol. 3, 10-12 Oct. 2005 [↑](#endnote-ref-12)
13. Lian XiaoFeng; Zhang Tao; Liu Zaiwen, "A Novel Method on Moving-Objects Detection Based on Background Subtraction and Three Frames Differencing," in *Measuring Technology and Mechatronics Automation (ICMTMA), 2010 International Conference on* , vol.1, no., pp.252-256, 13-14 March 2010 [↑](#endnote-ref-13)
14. 括号中的相对误差百分比是相对于MATLAB运算得到相应的协方差值的相对误差。 [↑](#footnote-ref-1)