国内图书分类号: G40-057

国际图书分类号: 378 密级：公开

**西 南 交 通 大 学**

**研 究 生 学 位 论 文**

基于Kinect的手语识别技术

在聋哑教学中的应用研究

年 级 二〇一一级

姓 名 朱明茗

申请学位级别 硕 士

专 业 教育技术学

指 导 老 师 景红 教授

二〇一四年四月十九日

Classified Index: G40-057

U.D.C: 378

Southwest Jiaotong University

Master Degree Thesis

THE RESEARCH ON APPLYING THE SIGN LANGUAGE RECOGNITION TECHNOLOGY FOR DEAF EDUCATION BASED ON KINECT

Grade: 2011

Candidate: Zhu Mingming

Academic Degree Applied for : Master Degree

Speciality: Educational Technology

Supervisor: Prof. Jing Hong

Apr.19, 2014

**西南交通大学**

**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，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权西南交通大学可以将本论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复印手段保存和汇编本学位论文。

本学位论文属于

1．保密□，在 年解密后适用本授权书；

2．不保密□，使用本授权书。

（请在以上方框内打“√” ）

学位论文作者签名： 指导老师签名：

日期： 日期：

西南交通大学硕士学位论文主要工作（贡献）声明

本人在学位论文中所做的主要工作或贡献如下：

（1）分析了手语对聋哑儿童的重要性，为了促进聋哑儿童与健全人之间的互动及改善人机交互方式，提出了手语识别系统。

（2）学习了相关的教学理论，并研究了手语识别系统对教学理论的支持。

（3）阐述了手语识别的流程和目前常用的算法，在此基础上提出了将Kinect应用于手语识别系统的优势与可行性。

（4）研究了使用Kinect和OpenNI/NITE进行手部跟踪的流程，及手部特征提取方法。本文针对手语的不同形式分三种情况进行了研究。对于手掌动态手语，使用了针对手掌运动方向角来提取特征向量的方法；针对手掌运动较小的情况，提出了根据手指运动，提取深度变化直方图的一阶差分的办法；对于静态手语，使用其轮廓的Freeman链码做为特征向量，同时对Freeman链码进行归一化和旋转不变性的处理。

（5）在识别阶段，本文针对手掌运动识别提出了改进的HMM模型，即分别提取手掌运动的平面和深度特征向量来训练HMM模型。针对手指动态手语识别，本文使用的是SVM/HMM模型。对于静态手语的识别，本文使用的是SVM模型，并对识别结果进行分析和统计。

（6）最后针对上述研究进行系统设计，实现了基于Kinect的手语识别系统。

本人郑重声明：所呈交的学位论文，是在导师指导下独立进行研究工作所得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中作了明确说明。本人完全了解违反上述声明所引起的一切法律责任将由本人承担。

学位论文作者签名：

日期：

摘 要

手语是聋哑人与其他人沟通的主要工具，应用手语识别技术使计算机能够理解并翻译读取的手语，帮助聋哑人与健全人沟通更加方便快捷。并且，随着体感技术的发展，使计算机更加容易理解人们动作，为手语识别技术的研究提供了新的方法与途径。由于传统的输入设备不能很好的满足人们对人机交互的要求。因此研究手语识别技术，不仅使聋哑人更加容易与健全人进行交流，同时有助于人机交互技术的发展。

本课题以聋哑学校教学需求为背景，在对传统的手语识别方法进行了深入研究的基础上，基于Kinect体感摄像机，对手语的特征提取方法和识别算法进行了改进，并实现了手语识别系统。本系统包含动态手语识别和静态手语识别部分，其中，动态手语识别包含手掌运动的手语识别和手指运动的手语识别。通过手语识别系统将手语进行翻译，并用人们理解的词语表达出来，提供聋哑学生与不懂手语的健全人沟通的途径，获得更多的学习资源，从而提高聋哑学生的教学质量。同时为改进教师与电脑的交互方式等研究奠定了很好的基础。

本文首先介绍了手语识别的背景和意义，并阐述了国内外的研究现状。接下来介绍了常用的特征提取和手语识别算法及与本文相关的技术和平台，同时阐释了将手语识别系统应用于聋哑儿童教育的理论支持。然后对手部特征提取和检测方法做了详细的描述。在动态手语的特征提取部分，针对Kinect独有的特点提出了手掌运动二次特征提取、手指运动变化深度直方图的方法。在静态手语特征提取部分，使用了手部的的Freeman链码，并对其进行了归一化和旋转不变性的处理。针对动态手语的手掌运动识别，提出了双层HMM模型，针对动态手语中手指运动的识别提出了改进的SVM/HMM模型，静态手语的识别使用了SVM算法。最后，将以上方法进行验证，同时给出了系统的设计和编码实现。

关键词：手语识别；动态特征提取；聋哑手语；Kinect；HMM；SVM/HMM；SVM

Abstract

Sign language is the most important communication tool between deaf people. Sign language recognition technology is to make computer understands sign language and interprets it in an understandable way to make healthy people and deaf people communicate more conveniently and quickly.With the development of motion-sensor technology, makes computer more easier to understand motions and provides a new way for researching sign language technology.Traditional input tools can’t meet the requirements of human-computer interaction. So researching sign language recognition technology can not only makes deaf-mute people easier to communicate with healthy people, but also contributes to the development of human-computer interaction technology.

This thesis based on the requirement of deaf-mutes school teaching, researches the traditional methods of sign language recognition.Using Kinect motion-sensing camera to improve the recognition technology and implements a sign language recognition system.This system includes dynamic sign language recognition part and static sign language recognition part.For dynamic sign language recognition part, it contains palm moving recognition and finger moving recognition. Using sign language recognition system makes deaf students get better access to various educational resources.And it lays the foundation of improving the interaction between teachers and computer.

First this thesis introduces sign language recognition’s background and meaning and describes the status at home and abroad. Next it introduces common feature extraction algorithms and recognition algorithms.And then it presents some technologies and platforms and illustrates theoretical supports of using sign language recognition system to deaf-mutes teaching.Then detail described the method of hands’ feature extraction and detection. In dynamic sign language feature extraction part, this paper according to Kinect characteristics proposes the methods of Hand Motion Twice Feature Extraction and Finger Motion Depth Histogram.In static sign language feature extraction part, this thesis uses Improved Freeman Chain Code. For hand motion recognition this paper presents Two Layer HMM Model .For finger motion recognition part this paper uses SVM/HMM Model,and for static sign language recognition part this paper uses SVM Model.At last, this paper validates the methods above and presents the design and implement of the system.

**Keywords:** sign language recognition; dynamic feature extraction; deaf and mute sign language; Kinect; HMM; HMM/SVM;SVM

目 录

[摘 要 I](#_Toc387910633)

[Abstract II](#_Toc387910634)

[目 录 III](#_Toc387910635)

[第1章 绪论 1](#_Toc387910636)

[1.1 课题研究的意义 1](#_Toc387910637)

[1.2 国内外研究现状 2](#_Toc387910638)

[1.3 研究内容和组织结构 3](#_Toc387910639)

[第2章 相关技术与开发平台 4](#_Toc387910640)

[2.1 常用手语特征提取 4](#_Toc387910641)

[2.2 手语识别常用算法 6](#_Toc387910642)

[2.3 Kinect体感设备 7](#_Toc387910643)

[2.4 OpenNI与OpenCV 9](#_Toc387910644)

[2.4.1 OpenNI简介 9](#_Toc387910645)

[2.4.2 OpenCV简介 10](#_Toc387910646)

[2.5 本章小结 11](#_Toc387910647)

[第3章 相关理论基础 12](#_Toc387910648)

[3.1 罗森塔尔效应 12](#_Toc387910649)

[3.1.1 理论概述 12](#_Toc387910650)

[3.1.2 对系统的理论支持 12](#_Toc387910651)

[3.2 社会学习理论 13](#_Toc387910652)

[3.2.1 理论概述 13](#_Toc387910653)

[3.2.2 对系统的理论支持 14](#_Toc387910654)

[3.3 学习金字塔理论 14](#_Toc387910655)

[3.3.1 理论概述 14](#_Toc387910656)

[3.3.2 对系统的理论支持 15](#_Toc387910657)

[3.4 人本主义学习理论 15](#_Toc387910658)

[3.4.1 理论概述 15](#_Toc387910659)

[3.4.2 对系统的理论支持 16](#_Toc387910660)

[3.5 本章小结 16](#_Toc387910661)

[第4章 手部跟踪及特征提取研究 17](#_Toc387910662)

[4.1 手语识别系统架构 17](#_Toc387910663)

[4.2 手部检测与追踪 18](#_Toc387910664)

[4.3 手部特征提取 18](#_Toc387910665)

[4.2.1 手掌运动特征提取 18](#_Toc387910666)

[4.2.2 手指运动特征提取 20](#_Toc387910667)

[4.2.3 静态手语特征提取 23](#_Toc387910668)

[4.4 本章小结 26](#_Toc387910669)

[第5章 基于HMM和SVM模型的手语识别 27](#_Toc387910670)

[5.1 HMM模型 27](#_Toc387910671)

[5.1.1 HMM基本原理 27](#_Toc387910672)

[5.1.2 HMM基本问题 28](#_Toc387910673)

[5.2 SVM模型 29](#_Toc387910674)

[5.2.1 SVM基本原理 29](#_Toc387910675)

[5.2.2 最优分类面 30](#_Toc387910676)

[5.2.3 SVM核函数 30](#_Toc387910677)

[5.3 基于HMM的手掌动态手语识别 31](#_Toc387910678)

[5.4 基于SVM/HMM的手指动态手语识别 34](#_Toc387910679)

[5.5 基于SVM的静态手语识别 36](#_Toc387910680)

[5.6 本章小结 36](#_Toc387910681)

[第6章 手语识别系统的设计与实现 37](#_Toc387910682)

[6.1 系统硬件及软件环境 37](#_Toc387910683)

[6.1.1 硬件环境 37](#_Toc387910684)

[6.1.2 软件环境 38](#_Toc387910685)

[6.2 手部检测与追踪 38](#_Toc387910686)

[6.3 手语特征提取的实现 40](#_Toc387910687)

[6.3.1 手掌运动特征提取 40](#_Toc387910688)

[6.3.2 手指运动特征提取 40](#_Toc387910689)

[6.3.3 静态手语特征提取 41](#_Toc387910690)

[6.4 手语识别的实现及运行效果 42](#_Toc387910691)

[6.4.1 手掌动态手语识别 43](#_Toc387910692)

[6.4.2 手指动态手语识别 44](#_Toc387910693)

[6.4.3 静态手语识别 45](#_Toc387910694)

[6.5 本章小结 47](#_Toc387910695)

[结论和展望 48](#_Toc387910696)

[致 谢 50](#_Toc387910697)

[参考文献 51](#_Toc387910698)

[攻读硕士学位期间发表的论文 54](#_Toc387910699)

1. 绪论
2. 课题研究的意义

手语是自然语言的一个组成部分，是聋哑人与健全人沟通和交流的主要工具，同时也是聋哑人学习口语和书面语言的工具。手语通过手势动作、手形及其变化来模拟音节或形象，同时配合表情和身体姿态来构成一定的词语和意思。尽管手语比口语的抽象程度要差一些，但和其他的交流方式，如读唇或笔谈相比会方便许多[1]-[3]。由于聋哑人阅读能力比健全人较弱，书写和阅读速度较慢，手语对于聋哑人的重要性更不能被书写和阅读取代。

随着计算机技术的发展，人机交互技术在人们的生活中逐渐扮演着越来越重要的角色。如今的人机交互方式大多数是通过键盘和鼠标等比较单一的输入工具，使用人体语言的输入方式与它们相比则要快捷方便的多。因此，使用自然语言与人体语言作为输入方式，加强计算机对人类语言和动作的理解，在人机互动领域有积极的意义和应用价值[4]。

手语识别的主要研究目的，就是使计算机能够理解人给出的手语，并将手语翻译成为人能够理解的文本或语言等方式输出，进而为聋哑人和健全人之间的交流提供方便，或者为人体语言的凌空操控提供支持。因此，将手语用到人机交互中，有着巨大的学术价值和广阔的市场应用价值 [5]。

全世界大约有6500~7000万人为聋哑人，我国大约有2075万，将近占中国人口总数的1.67%，7岁以下的儿童近80万。聋哑儿童大多由于先天性生理缺陷，致使丧失了一定的语言能力和听力能力，因为他们缺少了许多重要的获取信息的渠道，使他们与健全人的沟通产生了一定的障碍，许多信息他们了解的都不是很多甚至根本不知道。他们只能看到事情的表面，无法理解深层次的东西，造成他们对事物的认识比较肤浅。并且由于语言的障碍，聋哑儿童很难与健全儿童一起，看到其他人在聊天、唱歌、听音乐、讲故事等，他们无法融入到其中，总觉得自己不如别人，长期在这种心理作用下，很容易产生自卑感。因为绝大部分的聋哑儿童与健全人的沟通存在障碍，导致他们的朋友很少。他们讲的话别人“看”不懂，其他人讲的话他们也不理解，这种沟通障碍使他们会受到其他人的误解、耻笑、嘲讽和欺辱，导致他们过早产生了忧虑，烦恼、焦虑等不安的不良体验，更加不愿意与健全人接触，性格也逐渐变得孤僻起来，这种对他们的身心产生巨大的影响[6]。

因此，减少聋哑儿童和其他健全人之间的障碍，对聋哑儿童的学习和身心成长有着重要的作用。如何使用一种方便有效的办法，促进聋哑儿童与健全人的交流，让他们感到自己与其他健全人是平等的，也是聋哑学校研究的主要内容之一。健全人要明白了解聋哑儿童的世界，首先要理解他们的“语言”，但是大多数健全人不懂得手语，因此通过手语识别技术对聋哑儿童的手语进行解读，用健全人熟悉的文字或语音翻译出来，让更多的人理解聋哑儿童所表达的意思，了解他们的内心世界，减少对聋哑儿童的误会与歧视。与此同时，让聋哑儿童走进健全人的生活，加强聋哑儿童和健全人的沟通，获得更多的朋友，感受到同健全人在一起时的快乐，使他们觉得自己和其他人是平等的。 不仅如此，通过与健全人的接触，使聋哑儿童获得更多的知识，提高他们的受教育质量，使他们逐渐克服心里的自卑感，增强自信心，感受到集体和社会的温暖。

因此，手语识别的研究能够为聋哑儿童的心理健康和教育提供帮助，对聋哑学校教学质量的提高有很大的推动作用。本课题针对聋哑儿童与健全人之间存在的沟通障碍，使用Kinect体感摄像机，OPENNI+NITE，OPENCV开发平台，实现手语识别系统。为聋哑儿童提供一个与健全人交流的平台，帮助他们融入到健全人的环境中，进而加强他们的心理建设，提高学习效率。

1. 国内外研究现状

手语识别涉及到许多研究领域，如机器人学、概率统计学、模式识别学、计算语言学等，因此，在理论和技术上，手语识别的研究都是一种挑战[2]。

传统的手语识别包括基于视觉的手语识别和基于数据手套的手语识别。其中，基于视觉的手语识别通过数字摄像设备获取人的手部图像，对获取的图像进一步的处理，进而识别出相应的手语信息。基于视觉的方法之优点是价格较低和投入成本小，但是容易受到光照及背景等因素的影响，且目前这种方法识别的手语还不够精确。基于数据手套的手语识别，用户需要佩戴数据手套，计算机通过数据手套获取人手部的运动信息并实时的处理，并将结果实时显示出来从而实现人机交互。基于数据手套的方法之优点是输入数据少和识别度高，但是数据手套设备昂贵及佩带厚重，不利于自然人机交互。[7]

在国外，手语识别的研究开始于1982年，主要是对静态手语、手指字母、手势和连续语句识别的研究和开发，以Poizner和Shantz研究程序，用来识别美国手语。之后，一些科研院所和世界知名的高校，包括相关领域的大公司也开始逐渐对手语识别技术进行研究和探索，到目前为止，已经有很多大的突破。1983年，数据手套的发明专利最先由Grimes在AT&T取得，并且可以实现部分简单手势的识别[8]。1992年，Fukumoto等研制的Finger-Pointer系统，实验者不需要佩带任何辅助工具，该系统实现了一些简单的手形和手指方向的识别[9]。1993年，Hinton和Fels研制出通过Polhemus型跟踪设备和Mark Ⅱ型数据手套，并且使用神经网络进行手势的聚类[10][11]。1996年，Weng和Cui对区分性和表达性两种特征进行分析研究，将手形图像和肩胸区域提取出来，通过模板匹配法分类，得到93%的识别率[12][13]。2001年，Bauer等首次使用K均值算法对德国手语进行识别，由12个手势组成一个语句，识别率汏80.8%[14]。2004年，Hernandez通过模板匹配法，对200个美国手语词进行识别，识别率高达95%[15]。2007年，Oya等使用手语和其他身体动作信息，识别美国手语，识别率超过了81%[16]。

在国内，研究起步相对国外较晚，其研究成果如：哈工大的吴江琴和高文等组织的团队，主要针对中国手语中单个词汇和简单的语句来识别，先后提出了：使用HMM和人工神经网络相结合的算法训练手势，识别率为94%和90%；使用动态高斯混合模型对单个的手语词汇识别，识别率达98.2%；以及研发出了中国手语识别系统，使识别词汇数量和识别率显著的提高。2000年，任海兵等开发了单个动态手势与连续动态序列的手语识别系统，识别了12个手势样本，识别率约97%。2006年，张良国等使用混合序列捆绑的隐马尔科夫模型来识别单个的手语词汇，识别率达92.5%。2008年，中科院提出了动态建模方法，综合识别率达88.3%。[5]

1. 研究内容和组织结构

本文对手语识别技术进行全面分析，并详细的阐述了将手语识别技术应用于聋哑儿童教育的理论基础。根据手语的表达形式不同，本文将手语分为动态手语、手指语和静态手语三部分，并分别对这三部分手语的识别进行了研究，对选用的手语特征提取方法和手语识别算法进行了改进，实现了基于Kinect的手语识别系统。

本文的内容组织结构如下：

第一章绪论。对本文的背景及意义进行了详细的说明，介绍了开发手语识别系统在人机交互领域的意义，以及聋哑儿童使用手语识别系统的必要性。同时对研究内容的国内外现状进行了阐述。

第二章相关概念。首先给出了手语和手语识别的定义；接着介绍了当前手语识别常用的特征提取方法和分类算法，为后面的算法研究提供理论基础；最后对手语识别系统中使用的主要平台和技术，包括Kinect、OpenCV、openNI等进行了简要介绍；

第三章手语识别对教学的支持性研究。研究分析了相关教学理论，阐述了手语识别对这些教学理论的技术支持，为手语识别系统应用于聋哑儿童提供理论基础。

第四章手部跟踪及特征提取算法分析研究。详细介绍了三种手语识别时的特征值提取方法，主要阐述了方向角、颜色直方图和Freeman链码在特征提取上的应用。

第五章手语识别。详细阐述了HMM和SVM算法及其如何对手语数据的训练和分类。

第六章系统实现。对以上技术进行了验证，并详细介绍手语识别系统的设计和编码实现，并对分类结果进行分析。

1. 相关技术与开发平台

通过使用Kinect体感设备提取的是深度信息，它因不受光线和背景的影响而使识别效果更加准确。

本文基于Kinect提出的手语识别方法，由手部检测、手部追踪、特征提取和手语识别四部分组成，其具体识别流程如图2-1所示。其中，手部检测和手部追踪功能主要通过OpenNI实现。



图2-1 手语识别系统流程图

下面将着重介绍手语识别中常用的特征提取和识别算法，以及本文中涉及到的相关的概念和技术。

1. 常用手语特征提取

特征是所要识别的数字图像中的特殊信息，要对图像进行分析首先需要找到代表其内容的特征，可以说特征是图像处理结果是否成功的决定性因素之一。而对特征的选取要具有代表性和稳定性，即它能代表图像中的对象信息，并且在不同的场景中特征相近。对于特征提取就是通过计算机提取这些图像特征，用某个特征值或特征向量等形式表现出来，进一步代入识别算法中的过程。[17]

常用的特征包括图像的形状特征、空间关系特征、颜色特征、纹理特征等[17]。

1. 形状特征

如今使用的比较广泛的形状特征是轮廓特征和区域特征。其中，轮廓特征即图像中目标的外边界，而区域特征即目标所占据图像的整个区域。

形状特征提取包括形状不变矩法、几何参数法、边界特征法、傅里叶形状描述符法等[18]。

常用边界特征法包括边界方向直方图法和Hough变换检测法[18]。边界方向直方图法使用的图像轮廓信息是通过对图像微分获得，使用这些获取的轮廓信息生成边界方向直方图[18]。Hough变换是利用点线之间的对偶关系，将图像空间中的曲线通过转换变成为参数空间中的一个点，Hough变换检测法是提取图像中的目标区域，并将区域边界通过边缘像素连接起来的一种方法[19]。

傅里叶描述符法是将图像中的目标边界作傅里叶变换，将变换的值作为其特征的方法[18]。

几何形状法就是通过图像中对象的几何值表示其特征，如物体的面积、周长和角度等。但是，几何特提取的前提是要定位出所识别的对象，并且将对象轮廓从图像中提取出来。因此，几何特征提取效果的好坏与图像对象的分割有很大关系。

形状不变矩法就是将图像中的对象形状信息用矩表示出来，形状矩的优点是对图像的大小、旋转和平移等操作受影响小[20]。

1. 空间关系特征

有时一幅图像中含有多个需识别的目标，这些目标之间在图像中存在着某些位置和方向的关系。包括两个或多个对象相邻，对象和对象包含部分或全部重合，或者一个或多个对象包含在其他对象中。空间关系特征很容易表示出图像中对象本身和本对象与其他对象的位置信息，这些信息对于图像检索是十分必要的。一般空间关系可分为相对空间关系和绝对空间关系。相对空间关系指对象之间的相对情况，如对象之间的位置关系；绝对空间关系指对象之间的方位关系，距离远近等。目前空间关系主要应用于二维图像的处理，用于描述高维、复杂的空间关系的情况较少。并且，空间关系特征着重于表示图像对象的位置信息，当对象位置发生变化时效果并不理想。[21]

常用的空间特征提取方法有两种，一种是将图像根据设定均分成若干子部分，针对每个子部分的分配索引并提取特征向量。另一种办法是根据图像中目标的重要程度将图像划分成不同子部分，并对分别这些部分创建索引并提取特征[21]。

1. 颜色特征

颜色特征是图像处理中应用比较广泛和基本的特征，一般使用图像中的像素值作为图像的颜色特征，一般使用RGB和HSV颜色空间[22]。颜色特征不需要进行大量的计算，对图像中对象的位置、形状、旋转等变化依赖性小，但是也容易受到背景等相同或相似颜色特征的因素干扰，不能很好的表达出图像中对象的局部特征[20]。

常用的颜色特征提取方法有颜色直方图和颜色矩等。

颜色直方图就是统计图像中不同的颜色出现的频率，将统计出来的结果存储起来或用直方图的形式描绘出来，将颜色的频率信息作为图像的特征值[20]。颜色直方图对图像缩放，旋转，平移等物理变化不敏感，因此，使用颜色直方图进行计算时，丢失了图像中对象的位置信息[17]。同时，背景的变化对直方图的影响也较大，因此，他适用于图像背景变化不大且不需要考虑物体空间位置的情况。

颜色矩使用了线性代数中矩的概念，它是一种比较简单有效的特征检测方法。它使用颜色的一阶矩、二阶矩和三阶矩来表示图像的颜色分布特征。因为每个像素具有三个通道，因此，颜色矩只需使用九个分量就可以描述图像的特征，将高维的特征简化成低维特征使用[17]。

1. 纹理特征

纹理特征描述了图像中对象的表面特质，它描述了物体表面的结构和表现形式[20]。例如，表达图像中对象的表面光滑、粗糙、波纹等特征。纹理特征不受对图像的平移、旋转等造成的干扰，但是容易受到光照和图像分辨率等条件的影响。由于光照原因，图像中对象表现出来的特征可能并不代表其真实的纹理，并且图像中分辨率变化较大时也会对图像的特征准确性产生影响。

纹理特征提取的方法一般包括LBP(Local Binary Patterns)纹理特征方法和灰度共生矩。其中：

LBP方法就是计算某一像素和周围像素值的大小关系。提取某一像素点，按一定的顺序依次将该像素值与周围的八个像素值进行比较。将这周围八个像素做阀值话处理，当值比中间的像素大是，将该像素值置为1，小于该像素时置为0。最后，将这八个数字写成八位二进制数，最后将这八位二进制数转换为十进制，最后得到的这个十进制值就是LBP值。LBP方法在应用中有诸多改进，包括CLBP，LBPHF等。

而灰度共生矩即根据图像中的某个像素的灰度值与指定距离的其它像素的灰度值进行比较，将具有某灰度的两个像素状况进行统计[20]。

1. 手语识别常用算法

目前，常用的手语识别算法有模板匹配法、决策树分类器、人工神经网络法、SVM和HMM等算法，下面对这些算法进行简单的介绍：

1. 模板匹配法

模板匹配法在数字图像处理中使用的很广泛，是模式识别领域比较简单同时比较重要的算法。模板匹配法就是根据一幅图像中的对象，计算它与另一幅或多幅图像中对象的相似度，从而找到相匹配的目标图像 [23]。

模板匹配法的优点就是模板建立起来简单方便，适用于识别的目标较少的情况。但是当识别的目标较多时，识别率较低。

1. 决策树分类器

决策树分类器为一个树结构，树的非叶节点表示算法的判别标准，叶节点表示一个单独的类标签。一般构造决策树使用自顶向下的方法，训练样本从树的根节点开始，进行递归来构造决策树。当输入测试样本时，从决策树的根节点开始判别，根据判别结果决定其下一分支，直至到达叶子节点从而判别测试样本所属类别[24]。

1. 人工神经网络[24]

人工神经网络是一种模仿人类的大脑神经网络的结构和信息处理方式的模式识别算法。它将大脑的神经网络用数学模型表示出来，通过一个个简单的神经元之间相互连接、彼此传输信息、相互作用产生复杂的神经网络。这些神经元在不同的情况会处于不同的状态，并且是非线性变化。它们的这种结构使信息离散的存储起来。当神经网络局部出现问题时，其他部分并不受到影响，提高了算法的鲁棒性。同时，人工神经网络算法是按层分布的，每层独立的分部若干个神经元，它们之间可同步工作，使信息能够并行协同处理。

在神经网络中，算法的学习能力并不是由局部的神经元决定的，它取决于各个神经元间的相互作用。当神经网络中有大量的神经元时，其性能和功能也随之变得强大起来，能够处理多种复杂的问题。

不仅如此，神经网络的神经元之间连接强度会通过使用的权值决定，随着对算法的不断训练和优化，相应的权值同样会不断的调整，从而提高算法的灵活性。

1. SVM

支持向量机SVM是建立在VC维以及风险最小化基础之上的机器学习算法。它与其它算法比起来有诸多的优势，如支持小样本学习方法、支持非线性和高维模式。许多分类问题不能单纯的使用线性问题解决，但是使用SVM把这些样本通过核函数映射到高维空间中，可以得到很好的解决。 SVM的优点在于它只是使用了少数的支持向量，这样在很大程度上减少了计算的工作量，提高的算法的效率；而且它对个别的支持向量的增减不存在影响，具有很好的鲁棒性。SVM模型最初是针对二分类问题，在解决多分类的问题上， SVM算法给出了一对一、一对多和DAG-SVM等形式，有时还需要这些形式相应的组合[25]。

1. HMM

隐马尔科夫模型HMM，由状态转移矩阵和观察值矩阵组成，是由马尔可夫链演变的。当系统产生一系列的观察序列时，人们并不知道观察序列中某个值是在哪种状态下产生的，即它的状态转移过程是不可见的，因此不能直接得出模型的状态，但是人们可以观测到观察序列。假设模型中包含N种状态，那么状态转移矩阵为N\*N阶矩阵。通过观察值矩阵，推导出在时间t~t+1之间状态由i转为j的概率，将推导出的概率值记录在状态转移矩阵里。[26]

HMM一般涉及到三类问题：评估问题、解码问题和学习问题。其中，评估问题即给定模型参数和观测序列，获得最能生成这个观测序列的模型，通常使用forward算法；解码问题就是给定和观测序列，寻找马尔可夫模型中的转移矩阵，通常使用Viterbi算法；而学习问题即给定观测序列，且模型参数并不知道，要找到模型参数，使产生观测序列的概率最大，通常使用Baum-Welch算法实现[27]。

1. Kinect体感设备

微软在2010年发布Xbox360时，将Kinect其作为体感周边外设一起推出。Kinect提供了一种全新的操作方式，它通过语言、手势、动作，以自然用户界面实现操作。Kinect是一种利用光学感测原理识别人体动作的3D体感摄像机，它能够准确追踪人体骨骼、识别人体的动作，并实时处理相应操作。通过将Kinect体感设备与计算机结合起来，用户无需使用任何输入设备，即可通过自身的肢体动作向计算机发布指令，使之对计算机的控制具有更大的灵活性，进而增加了用户与计算机的交互性和用户体验。

Kinect有两个3D深度摄像头和一个RGB摄像头。左边的深度摄像头发射红外光谱，覆盖Kinect的整个视野范围并进行编码；右边的红外线接收器接受并读取编码后的光线并通过芯片处理，进而创建Kinect可视范围内的深度图像。中间RGB彩色摄像头拍摄可是范围内彩色图像，配合3D摄像头实现Kinect的图像识别检测操作。Kinect下部是一个麦克风阵列，由4个麦克风组成，用于采集和定位声音，同时帮助过滤背景噪声，使其作为语音识别的命令。Kinect的底座内置仰角控制马达，可通过编程控制Kinect的拍摄角度，使摄像头本体能够看到的范围更广[28]。

Kinect实物图如图2-2所示。



图2-2 Kinect实物图

随着Kinect的推出，它已被应用于游戏、电子商务、医疗、教育等多个领域。例如：模拟3D试衣间，客户可以快速试穿衣服；协助医生做外科手术，降低手术的时间和风险；模拟化学实验，降低了学生实验的危险性。

Kinect关键技术如下：

1. 深度提取

Kinect与传统摄像头最大的区别在于它不仅可以获取彩色信息，同时在可视范围内，能够采集物体的深度信息。他的基本原理是：激光照射粗糙物体会产生衍射斑点，这些衍射斑点具有高度的随机性，并且随着距离的变化，散斑也随着改变，在同一空间中不同的散斑图案都不同。只要使用散斑对空间编码，当有物体进入空间时，即可定位。Kinect首先利用2个3D摄像头对测量空间进行散斑编码，并读取编码的光线；接下来晶片对读取的编码计算，进而产生深度图像。深度图像如图2-3所示。



图2-3 Kinect深度图像

1. Kinect的骨骼追踪

Kinect在实现深度信息提取的基础上，能够实现对人体骨骼的追踪。它通过对深度信息的处理，计算人体各个部分的位置信息，通过(*x,y,z*)坐标表示。使用不同的开发平台，定位的关节点也有区别。例如，在OpenNI中定义了24个关节点，目前15个可以使用，如图2-4所示[51]。图2-5为使用OpenNI提取的骨骼图像。

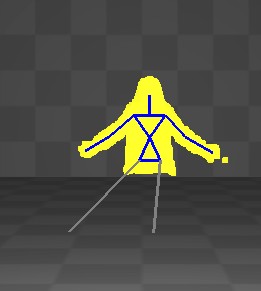
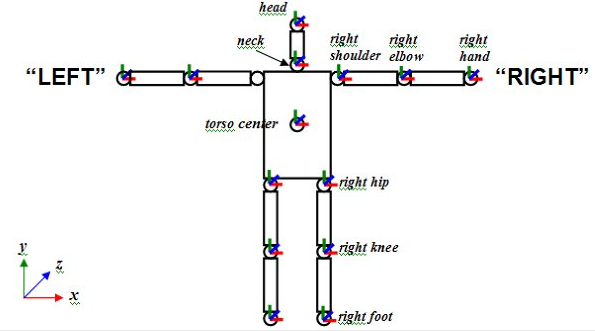


图2-4 关节点分布 图2-5 Kinect骨骼图像

1. OpenNI与OpenCV
2. OpenNI简介

目前针对Kinect的开发包一共有两种，一种是微软官方的Kinect for Windows SDK，另一个是OpenNI。两者之间主要区别如下：

1. Kinect for Windows只能追踪全身，不包含特定的追踪模式，如手部，上半身等；而OpenNI包含特定追踪，如手部追踪，手势识别等。
2. Kinect for Windows更消耗CPU，而OpenNI消耗的CPU更少。
3. Kinect for Windows只支持WIN7操作系统，不支持Unity3D游戏引擎；OpenNI不仅支持WIN7，同时支持其他Windows操作系统，包括Vista、XP；并且OpenNI支持Linux、Mac OSX，和Android操作系统。OpenNI自带的代码全面支持Unity3D和Ogre。
4. OpenNI在全身追踪上缺少头、手、脚和锁骨；与之相比Kinect for Windows SDK更精确和稳定。
5. OpenNI不支持音频和马达，Kinect for Windows SDK则支持对音频和马达的控制和操作。

由于本课题针对手语识别技术进行的研究，更偏向于对手部信息的捕捉。而且OpenNI支持更多的开发平台，为了方便后续对系统的扩展研究，本课题采用OpenNI作为Kinect的开发包。

OpenNI是Open Nature Interation的缩写，翻译成开放式自然操作。OpenNI对自然的操作包括手势、语音、身体动作等。

OpenNI本身定义了一系列的API，用来撰写和简化自然操作程序的开发，它是一个多语言跨平台的框架[28]。它使开发者使用和获取听觉与视觉等相关感应信息时更加方便。OpenNI的系统框架如图2-6所示。



图2-6 OpenNI系统框架图

图2-6中，最上层的为应用程序，这部分就是开发者自己需要撰写的部分；中间层则为OpenNI的部分，它包含了一些中间组件，用来做全身分析、手部分析、手势侦测、场景分析等功能。同时负责和底层的硬件打交道。目前，一般大多使用PrimeSense提供的NITE中间件来进行开发。最低层为硬件部分，目前OpenNI支持3D传感器、RGB摄像机、IR摄像头，麦克风这四类[51]。目前，主要是使用Kinect和Xtion。

在中间层，OpenNI定义了四种组件：

1. 手部分析：通过感应器追踪手的位置。
2. 手势侦测：辨识某些特定的手势，目前，OpenNI能够识别4种手势：RaiseHand，Wave，Click和MovingHand。
3. 全身分析：由感应器获取人体的相关信息，如骨骼信息，关节、相对位置角度等。
4. 场景分析：分析场景中的人和不同的物体。为场景中的物体标定坐标，并且将物体的前景和背景分离出来。

OpenNI中定义了相关的节点，用节点来表示内部的基本单元，它是OpenNI接口的基础。节点为一系列的组件，每个节点封装了特定的功能。通常情况下，产品节点只使用代表底层数据的节点，并对这些数据进行更进一步的分析。如Depth Generator 表示生产深度图的节点；Image Generator表示生成颜色图像的节点；User Generator表示生成3D场景中人的节点[28][51]。

1. OpenCV简介

OpenCV全称为Open Source Computer Vision Library，由Intel微处理器研究实验室于1999年建立，是一个开源、跨平台的计算机视觉库。它包含300多个用于图像和视频分析的算法开源库，最初的目标在于建立一个简单易用的计算机视觉框架，帮助广大研发工作群体用于计算机视觉领域的开发。它采用优化的C代码编写，并且充分利用了多核处理器的优点，加快了设计目标的执行速度，提高了系统实时处理能力。[29]

OpenCV应用在许多计算机视觉的相关领域。包括计算机视觉监控、信息安全、网络图像和视频、游戏界面、无人飞行器等。由于计算机视觉很多涉及到模式识别算法、机器学习等内容，因此OpenCV提供了机器学习库MLL(Machine Learning Library)来方便开发者进行模式分类等操作。该机器学习库主要涉及到模式识别和聚类两方面，其开发领域十分广泛和灵活，不仅包括计算机视觉，还可以应用于其他使用机器学习的场合。

通过OpenCV开发的程序不需要更改代码，可以在多个平台间使用，如Windows、Unix、Linux、MacOSX等。同时，OpenCV提供了不同语言的接口，如Python、MATLAB、Ruby等，能够与不同的软件兼容，使其涉及的图像处理和模式识别相关的算法更加全面。不仅如此，通过OpenCV开发的程序在调试运行时被编译为二进制可执行文件，使程序的运行时间更短，提高了程序的实时性。

OpenCV的主体一共分为5部分，包括HighGUI、CxCore、Cv、ML、CxAux。其中，HighGUI包含用户交互及图像和视频的输入输出函数；CxCore是OpenCV最重要的一个库，它提供了最基本的数据结构和函数，包括数组、矩阵等的基本运算，出错处理函数等。核心函数库Cv包含了图像处理方法和一些计算机视觉算法。ML是机器学习库，包含大多数的函数和工具用于统计分类和聚类。CxAux是用来存储即将淘汰的和最新更新的函数和算法。OpenCV其中四个的基本结构如图2-7所示[29]。



图2-7 OpenCV结构图

1. 本章小结

本章介绍了手语识别系统的基本设计流程，阐述了手语识别过程中的特征及提取方式，同时叙述了几种常用的模式识别算法。并且，对本系统中使用的Kinect体感设备和OpenNI和OpenCV开发平台进行了基本的介绍，为后面具体的具体相关技术说明奠定基础。

1. 基于深度图像的静态手势识别

本章描述了一套能够从深度图像中提取手势区域，并在此基础上通过识别手指点个数，完成静态手势识别的算法框架。基于Kinect的静态手势识别的算法流程如图3-1所示。

为了识别静态手势，我们首先要从深度图像中，提取出手势区域。本文考虑到算法通用性，没有使用Windows Kinect SDK自带的骨骼模型来定位手的位置，而是采用深度值阈值限定在Kinect摄像头平面前的指定区域提取手势区域。由于得到的手势区域包含的像素点数量太多，不可能逐一判断所有的像素点是否是手指点，因此，本文采用手势轮廓对手势区域进行近似描述，并且在手势轮廓上寻找手指点。根据手指点的数量，划分静态手势。



图3-1 静态手势识别算法流程

1. 基于阈值的手势分割

通过Kinect深度摄像头得到的深度图像，除了手势，还有身体的其他部分、周围的物品等无关的背景冗杂。为了准确地识别手势，我们需要利用Kinect深度传感器的特性，去除这些冗杂区域，从而获取比较精确的手势区域。

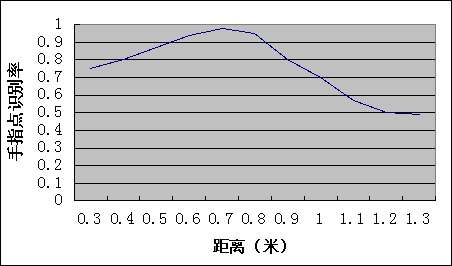
1. Kinect深度传感器

通过实验发现，当用户与Kinect距离过远时，单位面积内，Kinect深度传感器投射到手指上的定位点太少，导致深度图像中，手势区域变得非常模糊甚至消失。这种现象会严重影响后续的指尖识别算法的准确率。如图所示。

当用户与Kinect距离太近时，手势区域所占据整个深度图像比例过大，无法为后续的动态手势识别（尤其是书写轨迹识别）提供完整有效的数据。如图所示。

1. 阈值的确定

为了避免3.1.1中出现的情况，保证手指点识别的准确率，本文通过实验，确定手势区域的深度值阈值的取值范围。



通过实验发现：

1. 当距离在0.3~0.7米时，识别率逐步提高。在0.7米之后，识别率逐步降低。
2. 当距离在0.6~0.8的区间内时，平均识别率在94%以上。
3. 当超过0.8米时，随着距离的增加，识别率急剧下降。0.9米的识别率为80%。

考虑到用户在作动态手势时，手与Kinect的距离会因为手的移动，发生一定的变化，因此，为了既满足较高的手指点识别率，又给予足够的距离变动空间，本文确定0.6~0.8米为深度值阈值范围。

1. K-Means分割手势区域

借助3.1.2中实验得到的阈值范围，我们将深度图像中，在阈值范围内的像素点提取出来，作为手势区域的候选像素点。这些候选像素点中，仍然包含一些虽然在距离阈值范围内，但不属于手势区域的像素点，比如一些背景中冗杂的像素点、噪声点等等。

我们通过K-Means算法，对属于手势区域的像素点进行聚类。K-Means算法有如下优点：

1. 通过设置参数，可以支持深度图像中出现多个手势。
2. 可以去除背景中冗杂的像素点，只保留区域内连通、深度值相近的像素点。

K-Means算法实现的过程中，需要实验确定的参数如下：

1. Cluster Count：需要聚合的类的总数量。当只深度图像中只有一个手势时，设置成1即可。二当出现多个手势时，只需要设置成对应的手势数量，就能按照设置生成同等数量的类。
2. Minimum/Maximum Points For Valid Cluster：每个类所拥有的最小/最大像素点的数量。因为每个用户的手势区域大小是有上下界的。通过实验设定合理的最大、最小值，我们可以去除掉背景中一些明显不属于手势区域的冗杂区域。
3. Maximum Cluster Center Distances：类中所有像素点到类中心的最大空间距离。由于手上各点到手势区域中心的距离具有确定的最大值，通过实验设定合理的最大距离值，我们可以除掉一些不属于手势区域的像素点。

通过实验，我们确定各个参数的取值如表所示：

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster Count | Points For Valid Cluster | | Maximum Cluster Center Distances |
| Minimum | Maximum |
| 1 | 100 | 1000 | 120 |

1. 手势轮廓预处理

在手势分割完成之后，为了迎合后续的手指识别算法的要求，需要对手势轮廓进行预处理。手势轮廓是手势区域与背景区域的边界点组合起来形成的一条封闭曲线。通过预处理，可以得到手指识别算法要求的凸包；同时，降低手指识别算法的计算量，提高算法的实时性。

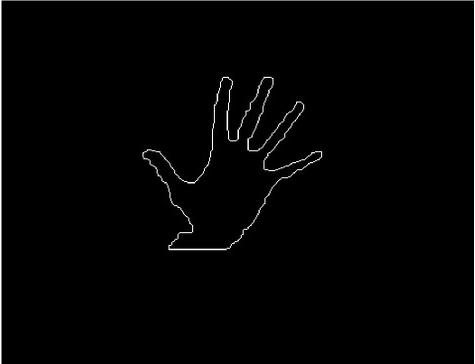
1. 手势轮廓提取

轮廓提取的主要原理是设定严格的检测标准查找符合手势边缘点条件的像素点,然后根据这些边缘点的特征寻找其同类,最后把它们串起來,形成轮廓图像。

本文使用OpenCV的cvFindContours函数来查找手的轮廓。

【FindContours的作用】

通过实验发现，手势轮廓提取的最优结果是，只提取最外层的轮廓和构成这个轮廓的水平、垂直、对角线直线段的端点。这样，构成轮廓的点的数量被压缩，更方便后续的处理。提取得到的手势轮廓如图所示。

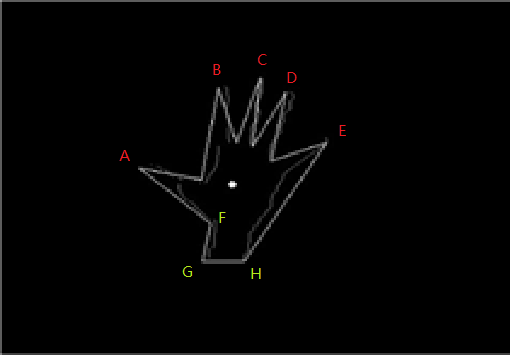


1. 手势轮廓近似

实验统计表明，在手势分割之后，手势区域边界点的平均数量在【伪造60~100】。后续的手指识别算法对边界点逐个判定时，需要耗费大量时间。在保证手指识别准确

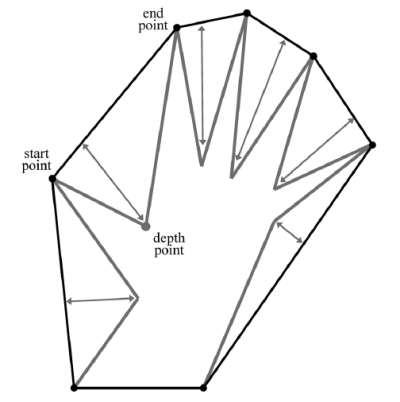
率的前提下，为了进一步减少计算量，本文使用OpenCV库中的cvApproxPoly函数对手势轮廓进行处理， 使用近似多边形描述手势轮廓，如图所示。通过观察可知，【图】中近似多边形的角点ABCDE可以近似地认为是手指点，而FGH则不是。

经过这样的处理，使用手指点检测算法判断手指点时，只需要对构成轮廓的所有角点进行判断。这种方法的另一个好处是，在使用手指识别算法对边界点逐个判定时，可以减少少数错误边界点所带来的干扰。



1. 手势轮廓的最小凸包

通过3.2.2中的处理方法，得到了手势轮廓的近似多边形。借助【cvConvexHul的算法】，可以得到近似多边形的最小凸包。



1. FT-GB算法提取手指

在对手指轮廓进行预处理之后，可以得到了手势区域的最小凸包。FT-GB算法正是利用了这一个特点，通过手型的几何特征，判断最小凸包上的各个点是否属于手指尖。

3.3.1 FT-GB算法描述

假设H={H1，H2...Hn}是构成最小凸包的N个点，C={C1，C2...Cm}是近似多变形上的M个点。FT-GB算法的具体描述如下：

（1）对于所有的Hi H，我们在C中找到Hi相邻的两个点{Di，Di-1}

（2）我们计算HiDi 和 HiDi-1之间的夹角α。

（3）如果α 小于阈值A\_Thres，那么，就判定Hi为指尖。

按照以上所描述的方法步骤，可以得到所有的手指点。这种方法的特点在于：

1. 即使手掌平行于摄像头平面，发生旋转，指尖角度的几何性质也能够保持。
2. 对于每根手指，α参数可以认为保持不变（经过实验验证，也确实是这样）。
3. 相比于其他的手指检测方法，这种方法只有α参数需要假设或者实验设定。

3.3.2 FT-GB算法的参数实验

FT-GB算法中有一个很重要的变量α。直观地来讲，α是手指根两端，与指尖中心的连线的夹角。两个可能因素会影响α：

1. 对于不同用户，手指的粗细和长度会发生变化。
2. 对于同一个用户，手指发生旋转。

对于第1个因素，本文采集和测量了5位用户的α值，发现5位用户的α值并没有明显的差异。因此，本文假设第1个因素不影响α的选取。

由于第2个因素涉及的情况复杂，本文没有对这个因素进行讨论，而是假设手掌所在平面始终与摄像头平面平行。

一方面，如若阈值参数α\_thres设置得太大,可能会将手腕处那一段（可以构成一个钝角）误认为为指尖；另一方面，当手指较短时或者由于离Kinect较远，造成截取图片中手指较短时，测出的角度会很大，如若a\_thres度太小，就无法识别出指尖。

通过实验，我们最终确定α\_thres的取值为45。

1. 静态手势的识别

通过FT-GB算法

1. 本章小结

本章首先通过实验，确定了能够为FT-GB算法提供有效手势区域的深度值阈值；接着，利用深度值阈值对深度图像信息进行手势分割；其次，使用OpenCV的轮廓提取和近似算法，对手势轮廓使用近似多边形进行描述，并使用最小凸包算法，求出近似多变形的最小凸包。最后，借助FT-GB算法，对最小凸包上的边界点，逐个判断是否为指尖，然后根据指尖的数量识别静态手势。

1. 手部追踪及特征提取研究
2. 手语识别系统架构

基于Kinect的手语识别系统的实现主要分为三个阶段：手部检测和跟踪、手部特征提取和手语识别。其中，手部的跟踪和特征提取将在下面两个小节中具体阐述。具体系统的组成如图4-1所示[35]。



图4-1 基于Kinect的手语识别系统

如图4-2所示，本手语识别系统由动态手语识别和静态手语识别两部分组成。其中，动态手语识别部分又包括手掌动态手语识别和手指动态手语识别。

1. 动态手语识别

手掌动态手语识别指通过跟踪手掌心来提取手掌的运动信息，进而对所表示的手语进行识别。但是有些动态手语在表达时，打手语者的手掌位置并没有发生变化，或者手掌的位置变化很小不容易探测到，只能通过其手指的运动来判断手语所表示的意思，因此在动态手语识别中加入了手指动态手语识别部分。

1. 静态手语识别

静态手语识别是指打手语者在表达过程中，手的位置和形状并没有发生变化，因此不能通过手的变化来判断手语，要使用与动态手语识别中探测手部动作相不同的方法来对静态手语识别。



图4-2 系统实现结构

1. 手部检测与追踪

一般情况下通过彩色视频序列分析，来对人动作的检测与追踪。相关的算法有很多，例如，伊尔马兹(Yilmaz)等使用目标的轮廓变化过程描述动作内容；拉普捷夫(Laptev)等通过检测器检测空间兴趣点，根据兴趣点表示相关动作；王亮(Wang Liang)等使用流行分析人的运动形态，并对得到的信息进一步分析达到识别效果。但是当存在光照变化、反射、阴影和背景复杂等因素时，以上算法不能准则的捕捉到目标。使用Kinect获取深度图像能够很好的避免上述这些情况，因为深度图像不会受到环境、背景和光照等的变化，它的像素值只与摄像设备到物体表面的距离有关。因此当确定目标距离摄像设备的距离时，很容易根据距离将目标分割出来[36]。

Kinect摄像头的分辨率为640×480，其深度传感器的图像采集频率为30帧每秒，能够捕捉到可视范围内所有的人体动作，因此本文使用Kinect体感设备提取手部的深度图像，并通过使用OpenNI+NITE来得到手的质心位置。NITE是用于OpenNI的中间件，能够实现手部及其全身的骨骼追踪。本文通过使用NITE的人体动作跟踪与检测算法，找到手部在深度图像中每一帧的手部质心位置，根据获得的位置坐标及不同情况来提取手部的特征向量。

1. 手部特征提取

本系统具体由手掌运动识别、手指运动识别和静态手语识别三部分。根据不同的情况，使用的特征提取方法也不同。下面对本系统所使用的特征提取算法做具体介绍。

1. 手掌运动特征提取

手部的位置信息、运动速率和运动方向角等方法为手部动作识别比较常用的方法。文献[37]对这三种特征提取方法做比较，发现使用方向角作为特征向量时，手部动作的识别率最高，因此本系统使用方向角作为手掌运动的特征向量。

方向角的定义如下：设P1(*x*1,*y*1)和P2(*x*2,*y*2)分别为图像中的两点。那么*P*1和*P*2的方向角*φ*为：

 （4-1）



式4-1中。由上式可得到方向角的范围是0~2π，由此可以按方向角

的大小来将其划为不同的区域。通过方向角所在区域编号，提取特征值组成特征向量，从而作为后续识别算法的输入。本系统中识别手掌运动的算法是隐马尔可夫模型(HMM,Hidden Markov Model)，本文将在后面的章节做具体的介绍。如果将单只手的特征值数量设为*N*，则双手组合的特征值数量为*N*2。因为打手语者打同一个手语时，不可能每次手的运动方向角相同。当方向角区域划分过细，导致同一个动作得到不一样的特征值，降低了特征值表达的准确性。同时，若特征值数量过多，使HMM模型中N值过大，影响识别效果。

经过试验，发现将方向角区域划分为4部分识别效果较好。一般情况下可以通过

的值作为区域编号，但是经测试效果并不是很好。以右手举例，当实验者向右上

方挥手时，手部并不是做的直线运动，而是以手肘为轴做曲线运动，因此，在向右挥手动作接近结束时，手的运动方向往往是偏向右下方的。由于Kinect采集图像的速率是每秒30帧，帧之间的间隔较短，因此可以将手部运动轨迹的切线与水平方向的夹角认为是两帧间夹角，切线如图4-4所示。 切线所在区域如图4-5所示。

手肘位置

*l*3

*l2*

*l*1

***O***

*l*3

*l*1

*l2*

图4-4 右手挥动轨迹 图4-5 切线所在象限位置图

由图4-5可知，切线*l*1,*l*2均落在坐标系的第一象限内，但是当右手在挥手结束时，它的方向角并没有在0~之间，如图4-5中的*l*3落在第四象限内，产生了误差，使特征提取不准确。若右手向右平移时，同样不能保证轨迹与水平面平行的，一般情况下会落入第一或第四象限，影响特征值的准确性。

因此，本系统将代替作为方向角。特征值与方向角对应关系如表4-1所示。

表 4-1方向角与特征值对应关系

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **方向角** |  |  |  |  |
| **特征值** | 0 | 1 | 2 | 3 |

左手的特征值提取方式与右手一样。为了使其特征更加精确，在计算特征向量时加入了手部深度的值，手部坐标用(*x,y,z*)表示。若简单的将*z*值按正负分类，按上述特征值提取方法，单只手的特征值数量为8(即4×2)。设特征值数量为N，左右手特征值分别为，那么所得最终特征值为,当左手的特征值误差为*ξ*时，最终的特征值误差则会扩至*ξ*×N，为了避免最终误差过大，本系统在提取特征值过程中，将三维坐标分割成为两个二维坐标(*x,y*)和(*y,z*)，分别提取特征向量，(*x,y*)平面和(*y,z*)平面特征提取方式如图4-5所示。

***O***

***O***

*x*

3

2

1

0

*y*

*z*

3

2

1

0

*y*

图 4-5 (*x,y*)、(*y,z*)平面特征提取

即一个手语由左右手两个特征向量组成，它与最终特征值对应关系如表4-2所示。而将这两个特征向量分别代入HMM模型训练识别，具体训练识别方法将在后面介绍。

表 4-2特征值分布

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **左手/右手** | **0** | **1** | **2** | **3** |
| **0** | 0 | 1 | 2 | 3 |
| **1** | 4 | 5 | 6 | 7 |
| **2** | 8 | 9 | 10 | 11 |
| **3** | 12 | 13 | 14 | 15 |

本系统针对10种动态手语提取了它们的手掌运动特征向量。基于机器学习的模式识别准确率和样本数有很大关系，相同情况下，样本数量越多模型准确率越高。本系统中的手语均由作者本人输入，正常输入一个动态手语需要1.5s的时间。所以，系统在定位手部位置时，相隔3帧记录一次，一共记录15次，产生15个特征值。观察提取的15个特征值，发现当输入同一个手语时，前面大部分特征值基本一致，后面的特征值由于处于手语结束部分，使其表达出来的差异较大。因此为了提高特征向量的准确性，本文提取前10个特征值组成手语的特征向量。对每个手语分别提取100个训练样本和30个测试样本。一共1300个样本。

1. 手指运动特征提取

根据手掌运动方向角，可较好的提取手部运动信息。但是某些手语在表示时，其手掌的运动幅度并不大，甚至有些手语只是手指在变化。这时，根据手掌位置很难判断出手部运动的方向角，无法提取特征向量。而对于手指运动的检测，一般使用的方法为指尖标记、轮廓分析和模板匹配法。对于这三种方法的比较，不难发现基于指尖曲率的指尖检测算法适用性相对较好[38]。但是基于指尖曲率的指尖检测算法是根据获取手部轮廓，计算轮廓点的曲率，若符合手指点曲率范围则把该点设为指尖。因此它更适合于单手手指语，但不适合本系统中对手指运动的特征提取。

由于Kinect可以根据目标到摄像头之间的距离来生成深度图像，当手指运动时深度图像也会随之改变，因此，本系统根据深度图像的变化信息来对手指运动进行分析。

本系统使用Kinect提取人体深度信息，通过OpenNI+NITE人体跟踪与检测算法定位手部位置，设置并提取手部感兴趣区(ROI,Region Of Interest)。由于深度图像根据目标距摄像机之间的距离生成，在提取出ROI之后，得到手掌位置的深度值*z*，根据深度值设定阀值THRESH，当图像中某点深度值*z’*∈(z – THRESH,z + THRESH)时，认定其为手的一部分，将其提取出来，此时提取的是手部的二值化图像。为了使提取更加准确，这里使用中值滤波对提取的手部二值化图像做处理，得到手的二值化矩阵。[39]设手部ROI原始深度图像矩阵为matsrc，经过处理后的二值化矩阵为matbin，由于需要手部去除背景干扰的深度图像，因此最终得到的手部深度矩阵mat = mtxbin & mtxSrc，部分动态手语的二值化图像和最终手部深度图像如图4-6所示。因此只需对图4-6中手部深度图像做处理即可。



(a)三十 (b)阿姨 (c)职权 (d)工人 (e)导师

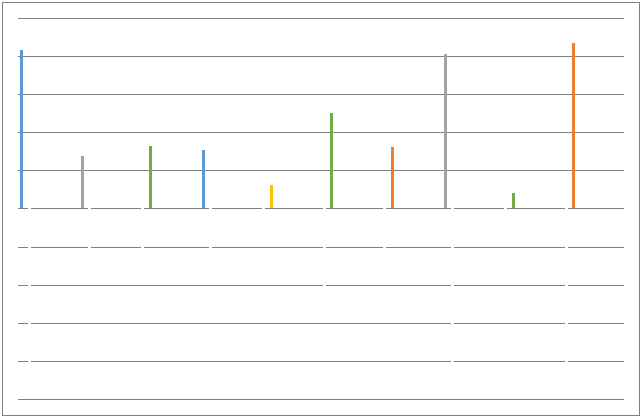
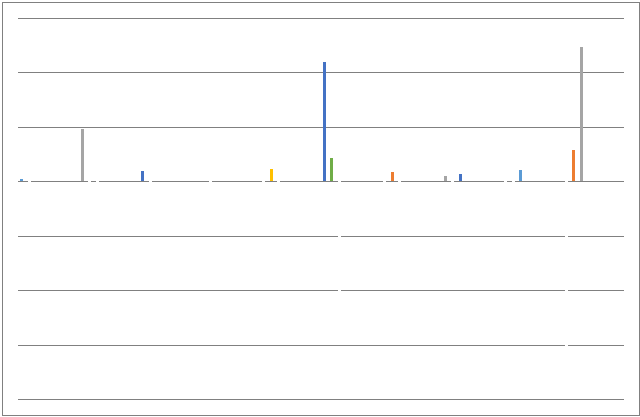
图4-6手部二值化及深度图像

手部深度图像提取流程图如图4-7所示。

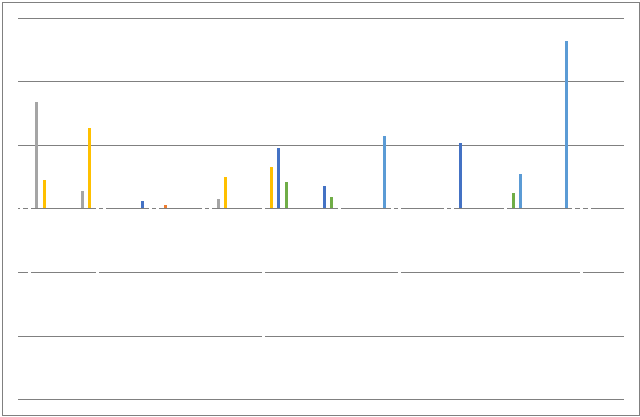
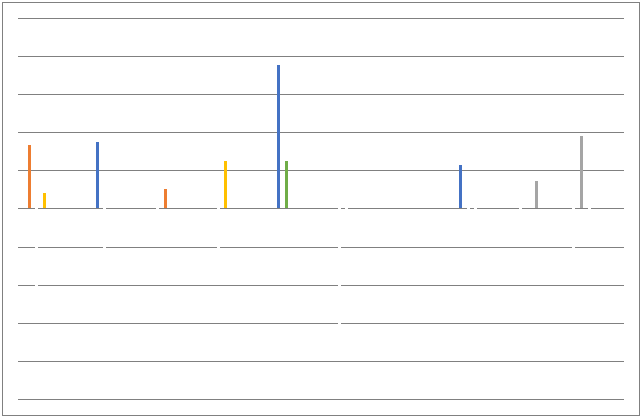
本系统通过提取当前关键帧的中深度直方图，统计具有该种灰度级的像素在图像中出现的频率，根据直方图的分布情况来作为特征向量。本系统中，根据作者本人手语表达习惯，每隔6帧提取一次深度图像，生成当前帧的深度图像的一维直方图，计算设定的灰度级范围内像素的分布概率，一个手语共提取10幅深度图像。根据实验发现，手部的灰度值分布大致在105~125之间，因此将一维直方图中的灰度级范围设为(105，125)，获取图像的均匀分布直方图。将直方图设为8维，即每幅深度图像提取的特征向量长度为8。因为该特征向量仅表示当前帧的深度特征，因此一个手指动态手语生成10幅深度图像。为了更好的显示手指运动情况，计算手语上一关键帧时提取的深度图像特征向量与当前关键帧特征向量的一阶差分，若当前关键为最后第10帧，则计算其与第一个关键帧的特征向量一阶差分。通过这种方法，每个手指动态手语得到了10个长度为8的特征向量。系统中对10个手语分别提取了100个训练样本和30个测试样本，一共1300个。图4-8显示统计的部分手语样本的特征向量信息。其中横轴表示特征向量中80（8×10）个元素，纵轴表示特征向量中每个元素的值。



图4-7手部深度图像提取流程图

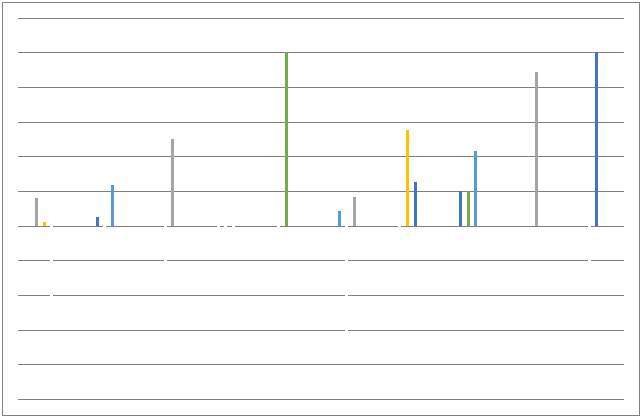
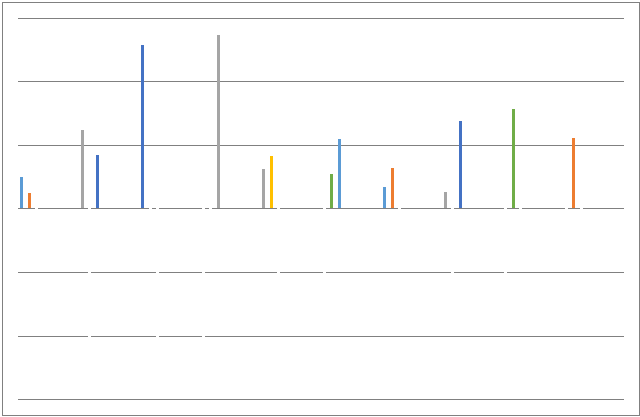


(a) 三十 (b) 六十



(c) 阿姨 (d) 职权

图4-8 手语样本深度直方图（1）



(e) 工人 (f) 导师

图4-8 手语样本深度直方图（2）

1. 静态手语特征提取

静态手语常用的特征主要包括颜色、纹理和形状等特征。其中目标的形状特征对光照、颜色等的干扰具有很好的鲁棒性，被广泛应用于图像处理中。本系统使用手部轮廓作为静态手语特征，根据手部提取轮廓获取Freeman链码，并将得到的Freeman链码做归一化、旋转不变性处理，最终处理结果作为静态手语的特征向量。提取过程如图4-9所示。

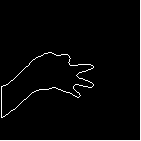
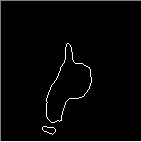


图4-9 静态手语特征向量提取流程

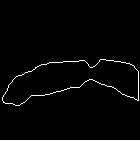
本系统使用Kinect体感摄像机提取手部ROI，并将其二值化，手部二值化图像的获取与4.2.2中提取方法类似。根据获取的手部二值化图像，使用OpenCV提取手部轮廓。本系统中分别对10种静态手语进行处理识别，其轮廓如图4-10所示。

1. Freeman链码定义及提取

Freeman链码(FFC,Freeman Chain Code)简称链码，用于描述图像轮廓特征，它是在1961年由Freeman最早提出的。Freeman链码常用于表示区域边界和曲线，在模式识别(Pattern Recognition)、计算机图形学(CG,Computer Graphics)、图像处理(Image Processing)等领域中应用十分广泛。它是一种边界的编码表示方法，描述了一个多边形序列是如何移动的。它将指定长度且有方向的线段沿曲线边界连接起来，用组成的序列来描述目标的边界。



(a)A (b)E (c)五 (d)十 (e)山



(f)个 (g)饼干 (h)房子 (i)王(姓) (j)教育

表4-10 静态手语轮廓

根据图形中像素点邻接方向选取的不同，将freeman链码表示方法也不同，常用的有4连通链码和8连通链码。4连通链码提取了上、下、左、右四个邻接点，8连通链码在4连通链码的基础上增加了左上、右上、左下、右下四个邻接点。将每个方向线段加以编码，编码方式如图4-11所示。

1. 4连通链码 (b) 8连通链码

图4-11 Freeman链码

以8连通链码为例，如图4-12，将汽车的后保险杠设为起点位置，对该点的8邻域按照顺时针方向试探，当检测到下一点的边缘像素，获取该点的方向数，依次类推，这里忽略了线段长度。每次检测时，判断下一点是否为起点，若是则跟踪结束。最终得到Freeman链码，本例中提出图4-12中的链码值为2110067075444443[40]。

1. Freeman链码的归一化处理

根据Freeman链码的定义和提取方式，可以发现Freeman链码与起点的选取有很大关系。当选取不同的起点，会得到不同的Freeman链码。

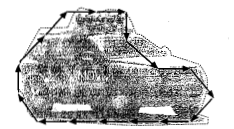


图4-12 提取Freeman链码特征

因此，需要通过一种方法将Freeman链码归一化，使其不受起点选取的影响[41]。具体方法如下：

设原始的Freeman链码长度为P，链码中最小值的个数为M。

1. 开辟一个的空矩阵A；查找出链码中最小值，将最小值的位置存于向量T中。
2. 根据T中存储的位置T[i]，将链码循环向左旋转T[i]格，所得的旋转后的链码存入A[i]中，直至i值为M。至此，矩阵A中存储的是以最小值为起点的链码矩阵。
3. 查找A中第二列最小值，设第二列最小值分别存在于N（N>1）行中，则创建一个矩阵J，将这N个行的行号保存在J中；
4. 根据J中存储的位置J[i]，查找A中第三列最小值，并将其所在行号存入J中进行替换。
5. 以此类推，直至J的长度为1。

假设某目标轮廓的Freeman链码为1-0-2-3-7-6-5-0-2-0-7-0-2-3-4，其最小值为0。根据以上办法得P=15，M=4，T={1,7,9,11}，得矩阵A如表4-3所示。

表4-3 矩阵A

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 2 | 3 | 7 | 6 | 5 | 0 | 2 | 0 | 7 | 0 | 2 | 3 | 4 | 1 |
| 0 | 2 | 0 | 7 | 0 | 2 | 3 | 4 | 1 | 0 | 2 | 3 | 7 | 6 | 5 |
| 0 | 7 | 0 | 2 | 3 | 4 | 1 | 0 | 2 | 3 | 7 | 6 | 5 | 0 | 2 |
| 0 | 2 | 3 | 4 | 1 | 0 | 2 | 3 | 7 | 6 | 5 | 0 | 2 | 0 | 7 |

A第二列中最小值为2，得J={0,1,3}，由于J长度大于1，继续下一轮循环。此时查找A中第0,1,3行中第三列最小值。得到最小值0，记录其行号得J={1}。此时J的长度为1，停止循环，得到最终的Freeman链码0-2-0-7-0-2-3-4-1-0-2-3-7-6-5。系统中十种静态手语的Freeman链码归一化结果在后续章节中介绍。

1. Freeman链码旋转不变性处理[41]

由Freeman链码的定义和提取方式不难看出，当图像发生旋转时图像的Freeman链码同样会随之变化。因此为了使链码在发生旋转时不发生改变，系统中将链码做一阶差分，将归一化后链码的一阶差分作为图像最后的特征向量[41]。

这里的一阶差分即Freeman链码中相邻两个元素之间按顺时针方向变化的次数。

具体如图4-13中所示。



图4-13 Freeman链码一阶差分

如上例中，得到最终的归一化链码为0-2-0-7-0-2-3-4-1-0-2-3-7-6-5，那么经过处理得到最后的特征向量为6-2-1-7-6-7-7-3-1-6-7-4-1-1-5。系统最终特征将在后续章节介绍。

1. 本章小结

本章首先介绍手语识别的流程，接着讲述本系统中使用的手部跟踪与检测方法，最后着重介绍手语的特征提取方法。其中动态手语部分，分别详细的介绍了基于手掌运动和手指运动的特征提取方法，最后又详细阐述了静态手语特征提取方法，为后文的实现奠定基础。

1. 基于HMM和SVM模型的手语识别

手语的特征提取和手语识别是系统中比较关键的两部分。本系统的手语识别与特征提取一样分为手掌动态手语识别、手指动态手语识别和静态手语识别。其中手掌动态手语识别使用的是改进的双层隐马尔科夫模型(HMM,Hidden Markov Model)，手指动态手语识别使用的是改进的SVM/HMM模型，静态手语的识别使用的是支持向量机(SVM,Support Vector Machine)。下面对HMM和SVM模型及手语的识别方法进行详细的阐述。

1. HMM模型
2. HMM基本原理

隐马尔可夫模型(HMM)[42]是一种统计分析模型。它将事物的变化分为两种，一种是可以通过某个函数，以时间t为变量来确定；另一种会由于某些条件出现多个结果，是随机的，它的结果可以由某个随机变量描述。前一种变化过程，称之为确定性过程；而后一种则称之为随机过程。

马尔可夫链是由安德雷·安德耶维奇·马尔可夫(Andrei Andreyevich Markov)于1870年提出的，它是具有马尔可夫性质的离散时间随机过程。隐马尔可夫模型是对马尔科夫链的扩展，它分为状态量和观察变量两个部分，是一个双重的随机过程。它的状态量和状态之间的转换不能直接观察到，称之为隐含层。但是它的状态量可以通过观测到的观察变量，根据一定的概率得到。由于其状态隐藏在观察序列中，因此称之为隐马尔科夫模型[42]-[45]。

HMM由观测随机过程和一阶马尔科夫过程组成。一阶马尔科夫过程由状态转移矩阵A和初始状态分布组成，其中A用来表达各个状态之间转移的概率。观测随机过程由状态输出概率矩阵B表示，B中的值用来描述状态和观测序列之间的对应关系。

具体构成如5-1所示[35]。



图5-1 HMM构成

HMM可用表示，一般情况下，可简单表示为。

具体参数说明如下[35]：设T表示观测值序列长度

1. N表示HMM模型中存在状态的个数

设Qt为t时刻所处的状态，观测值序列Q={q1,q2,…qT}，其中则Qt的取值范围为{s1,s2,…,sN}。

1. M表示模型中观测值的个数

设Ot表示t时刻的观察值，观测序列O={o1,o2,o3,…,oT}，v*i*为t时刻的观察值，则Ot的取值范围为{v1,v2,…,vM}。

1. A表示模型的状态转移矩阵

A={aij|i,j=1,2,…,N}，aij=P(qt=sj|qt-1=si)表示在t时刻状态由si转移到sj的概率。

1. B为观察概率分布矩阵

B={bjk,j=1,2,…N;k = 1,2,…M}, bjk=P(ot= vk|q=sj)，表示t时刻状态值为sj时观察值为vk的概率。

1. 表示初始状态分布

，，表示序列的初始状态为。

1. HMM基本问题

HMM模型主要解决了三个基本问题：估值问题，解码问题，学习问题[46]。

1. 估值问题

假设，已经知道一个HMM模型λ，其转移概率aij和bjk均已知。估值问题就是计算在当前模型下产生某一给定的观察序列O={o1,o2,o3,…,oT}的概率。

估值问题一般用于识别算法中，每个类别有对应一个HMM模型，计算在某个HMM模型下产生观测序列O的概率，由此计算出来产生这个观测序列O的最大可能性的HMM，从而进行识别，一般使用前向后算法和后向算法来解决估值问题[47]。

1. 解码问题

假设，有一个HMM模型λ，并利用该模型产生观测序列O。解码问题就是根据给定的HMM模型和观测序列的条件下，找到可能性最大的隐状态序列。

即要找产生观测序列O时的λ的状态次序，一般使用Viterbi算法来解决解码问题，其算法的核心思想就是找到最有可能产生观测序列的Viterbi路径，用于解释产生的观测序列[47]。

1. 学习问题

学习问题也称训练问题。假设，只知道HMM模型的N和M值，aij和bjk均未知。学习问题就是如何根据一个或多个观测序列，调整中aij和bjk的值，使其在当前模型中产生本观测序列的概率最大，即使的值最大。

在HMM的三个基本问题中，学习问题是最困难的一个。到目前为止，学习问题还不能得到真实的模型参数，一般使用Baum-Welch算法来解决学习问题。Baum-Welch算法也称作前向-后向算法，它是一种极大似然估计过程。其核心是想是通过迭代方式更新权重，使得模型能够更好的解释训练样本序列[47]。

1. SVM模型
2. SVM基本原理

SVM模型是V·Vapnik等在20世纪六、七时年代提出，九十年代中期发展起来的一种机器学习方法。SVM的方法是建在结构VC维理论和风险最小化理论基础之上的。SVM将数据通过非线性变换，使其映射到高维甚至无穷维的特征空间中。将原空间中的非线性判别函数转化为高维空间中的线性判别函数，在高维空间中找到最优分类面，从而到达分类目的[48]。

1. VC维

VC维用来描述机器学习的复杂程度，它是对函数类的一种度量。函数集的VC维就是当VC维越高时，问题的复杂程度越高。如对于只有0，1两种取值的函数，存在n个样本，如果该函数可以进行2n种分类，那么这个函数的VC维就是它能够分类出的最大的样本集的样本数目n。若存在函数可以把任意数目的样本分开，则这个函数簇的VC维是无穷大，但目前并不存在通用的关于任意函数簇的VC维理论[48]-[50]。

1. 经验风险

由于机器学习实质是通过建立与真实世界的事物相逼近的一种模型，通过模型来计算、还原和推断事物的运行状态。但是真实事物的运行状态和得到的模型之间的差距往往无法预知，这种差距称之为风险。在SVM模型中，将分类器产生的分类结果与真实结果之间的差距称为经验风险(Remp) [49]。

1. 置信风险

置信风险也可称为置信范围，它代表了给定的样本在分类器中的分类结果的可信程度，即分类器在给定样本上的误差。显然，置信风险无法准确的计算出来，只能给出一个估计的区间[38]。一般置信风险和学习机器的VC维n及样本数m有关，置信风险φ(m/n)与m/n之间的关系如图5-2中所示[49]。



图5-2 置信风险φ与m/n关系

1. 结构风险最小化理论

真实风险R(w)由两部分组成：一个是经验风险Remp(w)，另一个是置信风险φ(m/n),即 [50]。因此，在机器学习中，不但要减少经验风险， 同样需要考虑置信风险的影响因素，即使VC维尽量小来降低置信风险。

结构风险最小化思想就是将一个函数集分为若干子集，并按其VC维大小排列。在每个子集中寻找其最小的经验风险，将R(w)最小的子集作为期望风险最小的函数模型[49]。

1. 最优分类面
2. 线性最优分类面

线性最优面以二维空间中的两类样本为例，具体如图5-2所示[50]。图中的方块和圆形分别代表两类不同的样本，H为其分类线，H1和H2分别是与H平行且穿过离H最近的两类样本的直线，H1到H2的距离就是分类间隔(margin)。为了使风险R(w)最小，最优分类线的选取标准就是减少经验风险和置信风险。为了减少经验风险，需要分类线能够把两类尊亲的分开；为了减少置信风险，需要使两类之间的分类间隔尽量大。将最优分类线概念扩展到更高维的空间，就形成了最优分类面的概念了[51]。



图 5-2 线性最优分类面

1. 高维空间中的最优分类面

线性最优分类面很少能解决现实生活中的问题，因此为了解决更复杂的问题，需要将许多复杂的、非线性的问题转化为高维空间中的线性问题，得出其在高维空间中的最优分类面[52]。

1. SVM核函数

若要解决高维空间中的最优面问题，需要将低维空间中的样本，通过转换映射到高维空间中。由于在转换过程中只涉及内积的计算，因此，可以使用一个函数K(*x*i,*x*j)，使其低维空间中的输入值计算出高维空间中的内积值。K(*x*i,*x*j)如式5-1所示。

 (5-1)

式5-1即为核函数[50][52]。

能够满足条件的核函数有很多种，但是缺少选择核函数的理论依据。[49]如今比较常用的核函数有线性核函数、多项式核函数、径向基核函数和Sigmoid核函数四种，具体形式如式5-2至5-5[48]。

线性核函数： (5-2)

多项式核函数： (5-3)

径向基核函数： (5-4)

Sigmoid核函数： (5-5)

1. 基于HMM的手掌动态手语识别

在手掌动态手语识别部分，本系统使用的是HMM算法。将采集到的样本通过Baum Welch算法进行训练，生成模型。最后将测试集代入模型中，使用Viterbi算法进行测试。具体过程如图5-3所示。



图 5-3 HMM模型识别流程

1. 模型参数设置

在模型训练之前，需要设置模型中各个元素的值。由于在训练样本数比较少的情况下，模型中状态数N过多会影响模型的准确性[44]。由于本系统中对“物质”、“超级”、“调整”、“奋斗”、“比赛”、“举行”、“壮丽”、“出版”、“朋友”、“教学”这10种手掌动态手语进行识别，因此本文中将每个手语的状态范围设置为2~9，依次测试每个手语处于最佳识别率时的状态值。图5-4展示了手语“壮丽”在30个(x,y)平面输入样本的情况下，取不同状态值时的识别结果分布。从图中易得手语“壮丽”的HMM模型的状态值为7。表5-1展示了个手语与其最佳状态值之间的对应关系。

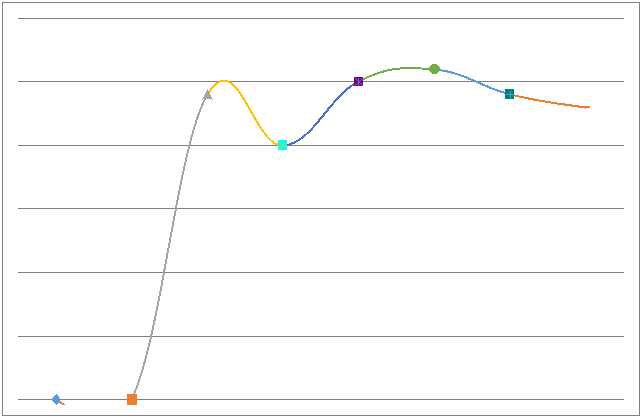


图 5-4 状态与识别结果关系图

表 5-1 各手语最佳状态对应表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **手语** | 物质 | 超级 | 调整 | 奋斗 | 比赛 |
| **状态值** | 4 | 5 | 9 | 8 | 4 |
| **手语** | 举行 | 壮丽 | 出版 | 朋友 | 教学 |
| **状态值** | 7 | 7 | 8 | 4 | 3 |

因为在提取特征向量时，单手在平面中的特征值存在4个，双手则存在16个特征值，即N值为16。

模型中的状态转移矩阵A的初始值根据设置的随机种子，由系统随机获得。系统中将种子设置为15，以N=7为例，那么T[7][7]为获得的随机矩阵，sum[i]为T中第i行所有值的和，则最终获取的矩阵A中的值aij=t[i][j]/sum[i]，如式5-6所示。

 (5-6)

同理得到观察矩阵B，如式5-7所示。

 (5-7)

初始状态分布π如式5-8所示。

 (5-8)

1. 模型的训练

模型的训练部分本系统做了一部分改进，即通过其识别结果训练双层的HMM模型。

系统首先通过Baum Welch算法在针对(x,y)平面的特征进行训练，得到10个HMM模型，并将测试集代入这10个模型中，计算其识别率；当个别手语的识别率较低时，则使用样本对HMM模型重新训练，直至该模型的识别率趋于稳定；最后再用测试集来计算生成的10个HMM模型的识别率，且当该模型的识别率低于85%时，认为该模型的识别效果欠佳，则对该模型对应的深度特征进行训练，生成其深度特征的HMM模型。

本系统分别对十个手语在(x,y)平面进行训练和识别。其中，物质、奋斗、出版、和朋友这4个手语识别率低于85%，因此提取它们的深度特征向量样本进行训练，生成其深度特征的HMM模型。

1. 手语的识别

系统首先在(x,y)平面内对10种手语进行识别，将10种手语产生的测试集分别代入已训练的10个HMM模型中，通过Viterbi算法计算手语的识别率。

当某些手语在(x,y)平面内识别率低于85%时，为了使识别结果的准确性，使用了双层的HMM模型。将识别率低于85%的手语通过训练好的深度HMM再次识别。

以本系统为例，由于系统训练的(x,y)平面的HMM模型对手语“超级”、“调整”、“比赛”、“举行”、“壮丽”和“教学”这6种手语识别效果较好，因此不需要训练其在深度平面内的HMM模型。

在(x,y)平面的识别过程中，发现手语误识的原因大多是将一种手语识别成其它某个手语，或者某种手语模型无法识别。如对于手语“物质”，识别成“朋友”的概率达10%，手语“奋斗”中不识别样本占13%。具体4种手语的误识对应关系如表5-2所示。

表 5-2 正确手语和误识手语对应关系

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **正确手语** | 物质 | 奋斗 | 出版 | 朋友 |
| **误识** | 调整 | 无法识别 | 物质 | 调整 |

因此，本系统中，当测试手语通过(x,y)平面的HMM模型识别结果在误识手语列表中时，则可以假设其存在误识的可能性。因此，为了排除这种可能，提高识别的正确率，需要对这个手语通过深度平面的HMM模型再次识别。具体的识别流程如图5-5所示。

通过上述方法，得出10种手语使用改进的双层HMM模型的最终识别结果。



图5-5 手语识别流程

1. 基于SVM/HMM的手指动态手语识别

训练HMM模型时，需要得到样本特征向量中的观测值个数。对于手指动态手语的特征向量，选取的是手部深度直方图信息的变化值，对于不同的手语其深度图像变化较大，不容易确定观测值个数。因此不能直接使用HMM模型进行训练和识别。

由于SVM模型对小样本、高维度及非线性等问题有很好的而解决能力[25]，结合HMM较强的处理时间序列能力，本文将这两种模型相结合，使用SVM/HMM模型，来进行手指动态手语的识别。

由于系统中一共选取了10种手指动态手语，因此使用SVM模型的输出结果有10种。首先，将每个手语的10个特征向量依次输入SVM模型中，通过SVM模型将手语特征向量进行分类。每个手语输出一个长度为10的识别序列，将这个识别序列作为HMM模型的输入，再次进行识别，从而得到最终的识别结果。SVM/HMM模型的结构如图5-6所示。

1. SVM模型的识别
2. 模型的选择

本系统主要通过SVM开发软件包LIBSVM，它包含的模式一共有5种：C-SVM分类，nu-SVM分类，one-class-SVM单分类器，epsilion-SVM回归，nu-SVM回归。这些方法主要通过变换函数项、参数等方法改变公式，使其产生出的算法在解决某类或多类问题时更有优势。本系统中使用的手语样本相对较少，这里选用C-SVM模式。



图5-6 SVM/HMM模型结构

1. 核函数的选择

径向基核函数核函数在SVM模型中使用比较广泛，本系统中选择的核函数是径向基核函数(RBF)，如式5-4所示，RBF核函数类似高斯分布，也称高斯核函数。它将样本映射到更高维的空间，线性核函数是径向基核函数的一个特例，因此可以用RBF代替线性核函数。由于核函数参数的个数对函数的复杂程度影响很大，参数越多函数越复杂，因为径向基核函数比多项式核函数需要更少的参数，并且当多项式的阶数较高时，可能会导致溢出的情况，所以相对来说使用径向基核函数更加简单。

1. 参数选择

使用RBF核函数，需要设置其参数和其惩罚因子的值。即需要输入惩罚因子C到目标函数中，得出SVM对应的测试集的识别率。本系统通过使用LIBSVM的grid.py工具，训练得到惩罚因子C的值为13。

1. 分类方法选择

常用的SVM算法主要有一对一、一对多和DAG-SVM算法。

对于有M类样本，一对一SVM算法使用不同的两类样本训练成一个SVM分类器，需要构建M(M-1)/2个分类器。一对多SVM算法即将某个类的样本与其他所有类的区分，来训练SVM模型，共需要训练M个分类器。DAG-SVM算法是一个二叉树结构，二叉树一共有M个叶子节点，代表着M个类。测试样本通过每层的判别，最终落到的叶子节点所代表的类别即为该测试样本的类别。本文使用一对一SVM算法作为分类算法[51]。

1. SVM/HMM模型的识别

通过SVM模型对手指动态手语特征向量的训练，使每个手语产生了长度为10，且取值范围为1到10的识别序列。将产生的识别序列作为HMM模型的特征向量再次进行训练和识别，最终的到识别结果。其中，HMM模型的训练和识别过程与本文5.3中相近。

本系统共针对10种手指动态手语进行识别，其中单手数字动态手语有3个，分别是“三十”、“六十”和“九十”。单手动态词语有3个，分别是“阿姨”、“职权”和“组长”。双手动态词语包含4个，分别是“工人”、“导师”、“家属”和“亲戚”。

1. 基于SVM的静态手语识别

本文中对静态手语的识别是根据提取手部的轮廓信息作为其特征向量，其手部轮廓信息包含手掌与手指的轮廓特征，因此系统将其统一进行处理和识别。

1. SVM模型的参数选择及训练

本系统中静态手语的识别方式与动态手指语类似，均采用SVM模型。选用C-SVM模式，设其核函数为RBF，通过使用LIBSVM的grid.py工具，训练得到惩罚因子C的值为12。

1. 手语的识别

本系统分别对10个静态手语进行识别，其中单手静态手语5个，静态手语分字母、数字和词句，分别为“A”、“E”、“五”、“十”和“山”。双手静态手语5个，分别为“个”，“饼干”，“房子”，“王（姓）”和“教育”。

与手指运动识别一样，静态手语的识别选用一对一SVM算法。因为系统中一共对10个静态手语进行识别，所以共产生45个SVM模型。

1. 本章小结

本章介绍了系统中所使用的HMM和SVM算法，并详细阐述了每种手语所使用的算法及其参数的配置。详细阐述了基于双层HMM算法的动态手掌识别的参数配置和识别流程，同时讲述了基于SVM模型的动态手指语与静态手语的识别。说明了模型的配置和参数的选择，给出了每种手语的识别结果并对结果进行分析。

1. 手语识别系统的设计与实现

通过手语识别系统，减少了聋哑人和健全人之间的沟通障碍，促进他们之间的交流。将手语识别系统应用于聋哑学校，帮助聋哑学生与健全人沟通，扩大聋哑学生的交友圈，使他们更多的接触健全人的世界，帮助他们更加健康的学习和生活。同时也为人机交互技术的研究提供了一种新的实现方式。

本章结合前文对手部跟踪、特征提取和手语识别算法等的研究和阐述，实现了基于手掌运动和手指运动的动态手语识别和基于手部轮廓的静态手语识别，将手语翻译成人们易懂的词语显示出来。本章详细介绍了本系统中主要功能的设计与实现，系统的整体流程如图6-1所示。



图6-1 系统流程图

1. 系统硬件及软件环境
2. 硬件环境

（1）PC主机一台

配置信息：华硕笔记本电脑，其型号为X42JZ，CPU为Intel(R) Core(TM) i3 480M，主频为2.13 GHz。内存为2.00 GB。显卡为ADM Radeon HD 6470M，显存为512M，支持DirectX11。

（2）Kinect体感摄像机

配置信息：类别为Xbox360，水平视角为57度，垂直视角为43度，机身转动范围为正负27度，传感深度范围为1.2~3.5米，有效工作距离为5米[28]。

1. 软件环境

操作系统为Windows7，32位，使用Microsoft Visual Studio2010工具开发。Kinect驱动为SensorKinect 5.0.0，Kinect开发包为OPENNI 1.0.0.23，NITE为NITE 1.3.0.17，图像处理软件使用OpenCV 2.4.4。

1. 手部检测与追踪

本研究主要使用Kinect体感设备和OpenNI/NITE平台，实现人体手语的识别。手部的检测与追踪是系统对手语处理和识别的前提，本系统中使用OpenNI/NITE自带的手部追踪算法。

OpenNI骨架由关节点(Joint)构成，每一个关节点都包含位置信息。OpenNI首先通过Kinect获取深度图像，进而提取人物信息，这里使用了OpenNI中DepthGenerator和UserGenerator生产节点，用来产生深度图像和人物信息。根据获取的人物信息生成骨骼信息，为了确定关节点的位置，使用了NITE中预先定义的矫正用姿势“Psi”，如图2所示。

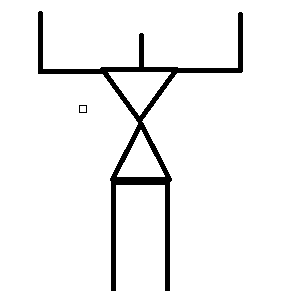


图6-2 “Psi”姿势图

当人物摆出的“Psi”姿势且NITE正确识别后，调用Skeleton进行人体骨骼的校正和分析。当骨骼校正成功后，便可对骨骼关节点追踪，通过GetSkeleton函数获取骨骼点信息。如果骨骼校正失败，则重新矫正。当骨骼追踪成功后，可通过GetSkeletonJoint()来读取骨骼的信息。函数具体声明如下：XnStatus xn::SkeletonCapability::GetSkeletonJoint(XnUserID user,XnSkeletonJoint eJoint, XnSkeletonJointTransformation & joint);其中，user为追踪的用户的ID；eJoint指定系统要追踪哪个关节，这里追踪的是用户的手部信息，因此可设为XN\_SKEL\_LEFT\_HAND和XN\_SKEL\_RIGHT\_HAND；joint则用来存储要追踪的关节点信息。具体识别与追踪流程如图6-3所示。



图6-3手部追踪与识别流程

手部追踪与识别的实现部分代码如下[52]：

//发现新用户

void XN\_CALLBACK\_TYPE NewUser( xn::UserGenerator& generator, XnUserID user,

void\* pCookie ){

generator.GetPoseDetectionCap().StartPoseDetection("Psi", user);

}

//骨骼校正

void XN\_CALLBACK\_TYPE CalibrationEnd( xn::SkeletonCapability& skeleton,XnUserID user,

XnBool bSuccess, void\* pCookie ){

if( bSuccess ) {

skeleton.StartTracking( user );

} else

{

((xn::UserGenerator\*)pCookie)->GetPoseDetectionCap().StartPoseDetection( "Psi", user );

}

}

//姿势检测

void XN\_CALLBACK\_TYPE PoseDetected( xn::PoseDetectionCapability& poseDetection,const XnChar\* strPose, XnUserID user, oid\* pCookie){

((xn::UserGenerator\*)pCookie)->GetSkeletonCap().RequestCalibration( user, FALSE );

poseDetection.StopPoseDetection( user );

}

1. 手语特征提取的实现
2. 手掌运动特征提取

手掌运动特征提取方法在章节4.2.1详细介绍，系统每隔3帧记录一次手部位置，并计算特征向量。当将手部看似静止，其实位置在某种程度上都会发生微小的变化，由于Kinect摄像机捕捉位置较敏感，当发生这些变化时它都会捕捉到，计算特征向量时就会产生误差，因此在提取前后手部的位置时设置阀值，本系统设相隔两次提取的位置距离相隔小于10mm时，认为手部位置并未发生变化。十种手语在(x,y)平面内的部分样本特征向量如表6-1所示。

表6-1 手掌运动特征向量

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **名称** | **特征向量** | | | | | | | | | |
| 物质 | 8 | 12 | 6 | 2 | 2 | 2 | 9 | 8 | 8 | 12 |
| 超级 | 9 | 1 | 13 | 7 | 7 | 7 | 1 | 15 | 3 | 3 |
| 调整 | 2 | 8 | 8 | 2 | 11 | 8 | 5 | 5 | 5 | 15 |
| 奋斗 | 7 | 11 | 10 | 8 | 13 | 13 | 4 | 14 | 11 | 11 |
| 比赛 | 7 | 13 | 13 | 7 | 7 | 13 | 13 | 13 | 13 | 11 |
| 举行 | 15 | 15 | 8 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 壮丽 | 6 | 5 | 8 | 15 | 14 | 9 | 13 | 8 | 3 | 11 |
| 出版 | 8 | 8 | 13 | 3 | 2 | 2 | 2 | 2 | 2 | 2 |
| 朋友 | 4 | 11 | 11 | 11 | 1 | 1 | 14 | 14 | 15 | 10 |
| 教学 | 5 | 15 | 15 | 2 | 5 | 4 | 8 | 15 | 15 | 15 |

1. 手指运动特征提取

通过OpenNI提取深度图像，根据深度图像追踪到手部得到手部深度ROI，具体方法在章节4.2.2中详细介绍。系统中，将深度图像大小设为640×480，ROI大小设为120×120。提取手部深度的阀值设为5，提取手部深度ROI主要代码实现如下[39]：

mDepthGenerator.ConvertRealWorldToProjective(1,&mJointTranRight.position.position,&handPoint);

handPosition.x = (float)handPoint.X;

handPosition.y = (float)handPoint.Y;

int handDepth = handPoint.Z \* (DEPTH\_SCALE\_FACTOR);

roi.x = handPosition.x - 60.0;

roi.y = handPosition.y - 60.0;

cvSetImageROI(depthShow,cvRect(roi.x,roi.y,120,120));

…

Mat mtx(pDest);

Mat mtxSrc(pSrc);

mtx = (mtx > (handDepth - BIN\_THRESH\_OFFSET)) & (mtx < (handDepth + BIN\_THRESH\_OFFSET));

medianBlur(mtx, mtx, MEDIAN\_BLUR\_K);

Mat mtxShow = mtx & mtxSrc;

提取的部分样本的特征向量如图6-4所示。

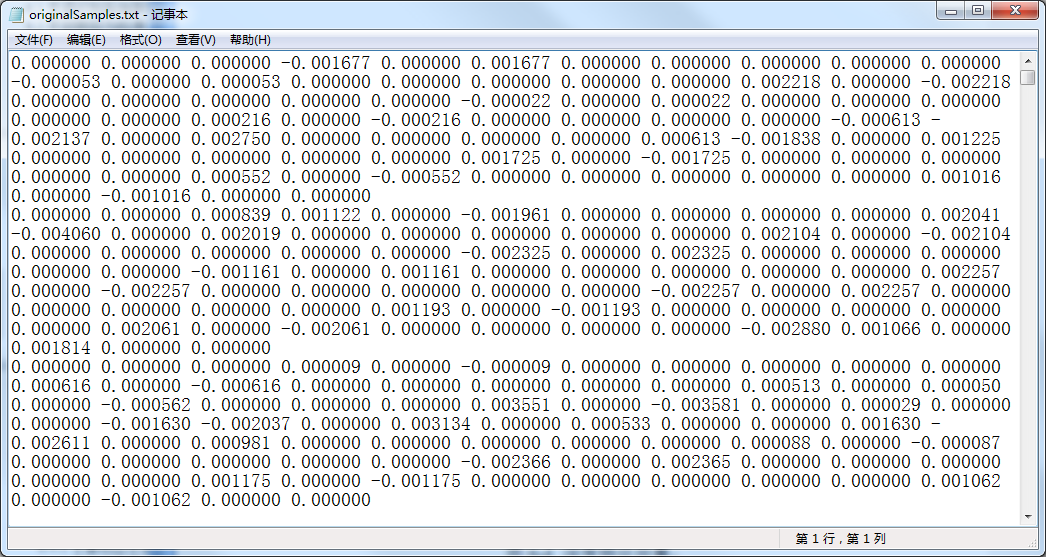


图6-4 深度特征向量

1. 静态手语特征提取

静态手语特征提取方法在章节4.2.3中详细介绍，本系统中一共对10种静态手语进行识别。通过提取它们的Freeman链码，并进行了归一化和旋转不变性处理，其中静态手语“王”处理后的归一化Freeman链码和旋转不变性处理后的Freeman链码如图6-5,6-6所示。

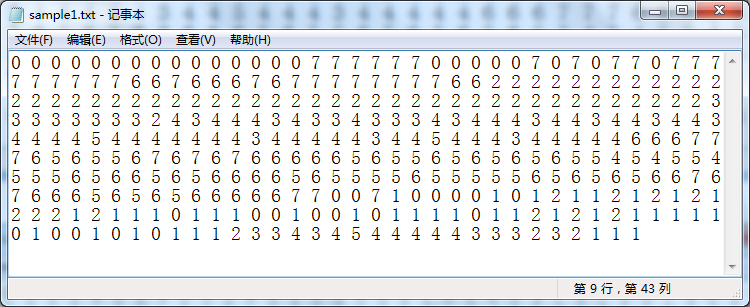


图 6-5 归一化Freeman链表

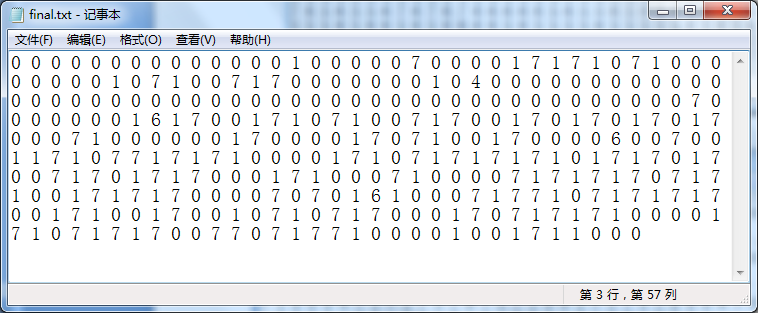


图 6-6 手语特征向量

1. 手语识别的实现及运行效果

将上小节中的特征向量代入到模型中，得到了30种手语的HMM和SVM模型。使用Kinect捕捉并读取聋哑学生的手部信息，由于本系统将手语分为手掌动态手语识别、手指动态手语识别和静态手语识别三种情况，聋哑学生根据打出手语所属的情况，使用鼠标选择对应的识别方式。系统根据选择的类型提取手部特征信息，并代入相应的训练好的模型中识别。最终，系统对识别结果的响应方式是通过文本将结果显示在“信息输出区”中，也可以将系统设定为鼠标或键盘响应方式，进而控制电脑操作，如PPT操作、多媒体播放器播放等，改变教师与计算机的交互方式，增强老师与聋哑学生的互动性，提高教学质量。

本系统使用MFC框架，系统的初始化界面如图6-7所示。

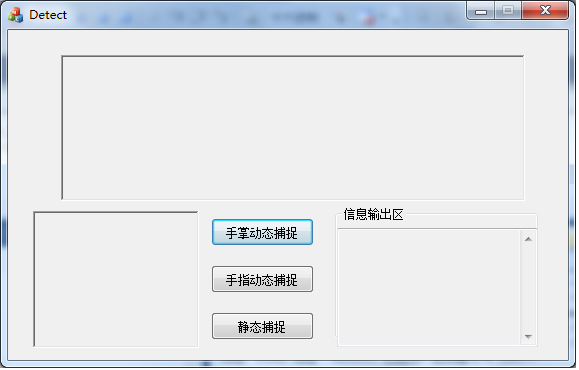


图 6-7 系统初始化界面

1. 手掌动态手语识别

当鼠标点击“手掌动态捕捉”时，Kinect获取深度信息，OpenNI探测人体位置并检测预设的“Psi”姿势定位手部位置。若定位成功，在信息输出去输出“定位成功，请开始识别”的信息，若定位失败，则输出“定位失败，请重新定位”的信息。若定位成功，则开始提取手掌运动特征向量，并根据得到的特征向量使用训练好的10个HMM模型识别。手掌动态手语识别流程如图6-6所示。

在界面中心区实时输出深度图像，系统读取人手掌位置变化信息，计算特征向量，识别结果将在“信息输出区”输出。

部分手掌动态手语识别效果如图6-8所示。



(a)比赛 (b)调整

图 6-8 手掌动态手语识别---(a)、(b)



(c) 奋斗 (d)出版

图 6-8 手掌动态手语识别---(c)、(d)

本文分别对10种手掌运动手语进行识别，每种手语使用100个训练样本和30个测试样本。表6-2和6-3统计了10种手语分别在(x,y)平面的HMM模型和改进HMM模型中的识别结果。

表 6-2 (x,y)平面HMM模型识别结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **手语** | 物质 | 超级 | 调整 | 奋斗 | 比赛 |
| **识别率** | 70% | 86.7% | 90% | 73.3% | 100% |
| **手语** | 举行 | 壮丽 | 出版 | 朋友 | 教学 |
| **识别率** | 100% | 86.7% | 73.3% | 73.3% | 100% |
| **平均识别率** | | 85.33% | | | |

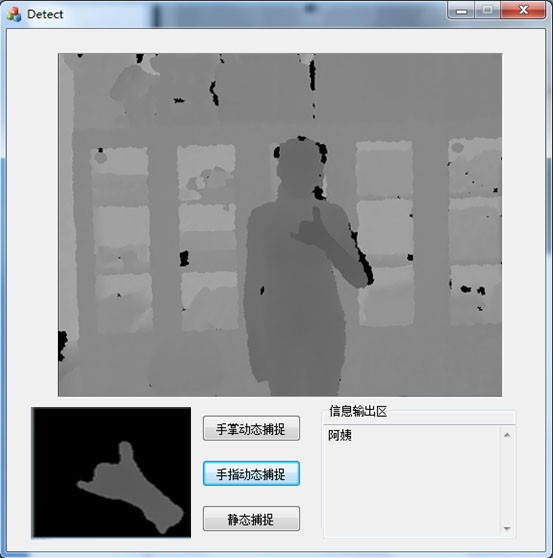
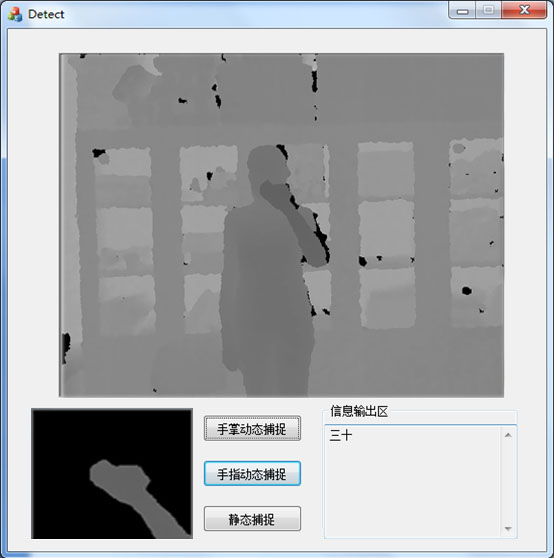
表 6-3 改进的双层HMM模型识别结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **手语** | 物质 | 超级 | 调整 | 奋斗 | 比赛 |
| **识别率** | 90% | 96.7% | 86.7% | 86.7% | 100% |
| **手语** | 举行 | 壮丽 | 出版 | 朋友 | 教学 |
| **识别率** | 100% | 86.7% | 80% | 83.3% | 100% |
| **平均识别率** | | 91.01% | | | |

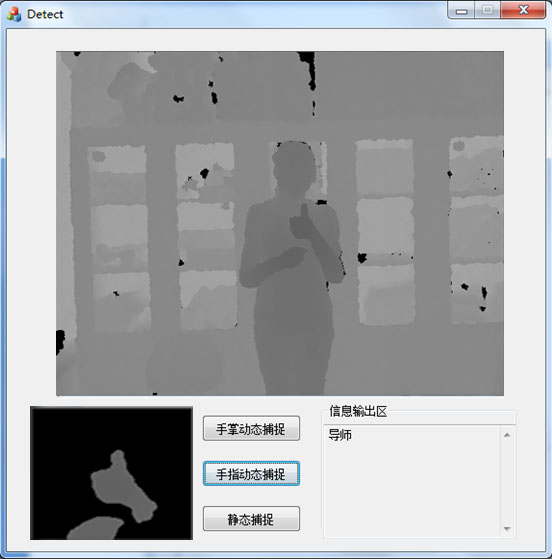
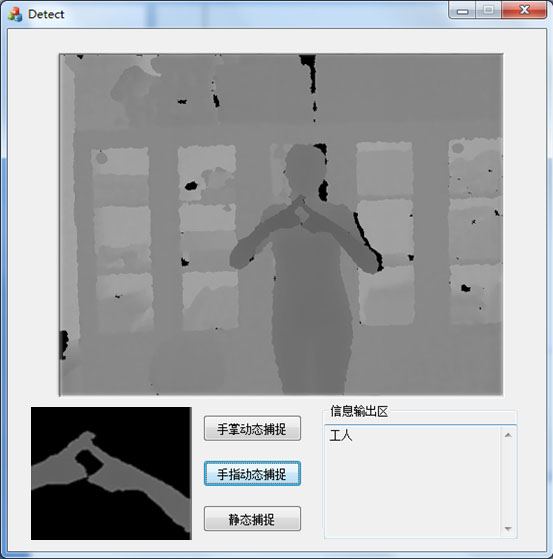
对比表6-2与6-3不难发现，使用改进的双层HMM模型进行识别的方法，提高了手语的识别效率。通过这种方法，将需要深度识别的手语由10种减至4种，减少了识别的工作量。同时，由于这4种手语有许多是被误识为“物质”，“调整”，在进行第二次识别时排除了为“调整”的可能性，提高了识别效率。

1. 手指动态手语识别

当点击“手指动态捕捉”，定位流程与手掌动态捕捉类似，此时系统捕捉手指深度图像的变化情况，生成深度特征向量，代入SVM模型中识别。在中间区域同样显示整体的深度图像，同时在左下方图像区域实时显示手部深度ROI图像，并在“信息输出区”输出识别结果。部分手指动态手语识别效果如图6-9所示。



(a)三十 (b)阿姨



(c)工人 (d)导师

图 6-9 手指动态手语识别

本系统对10个动态手指语进行识别，为每个手指语提取100个训练样本，30个测试样本。10种动态手指语的识别结果如表6-4所示。

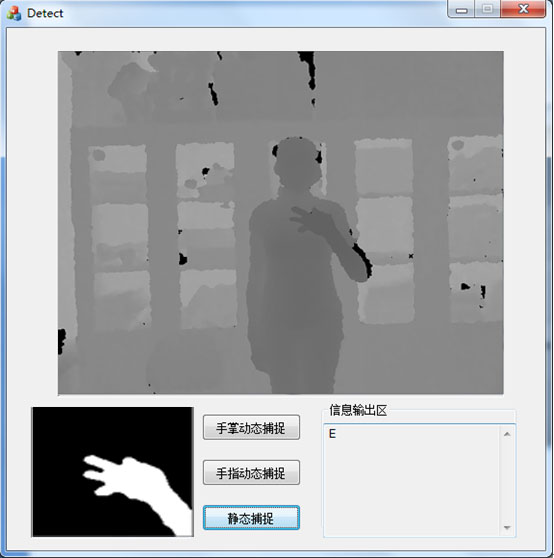
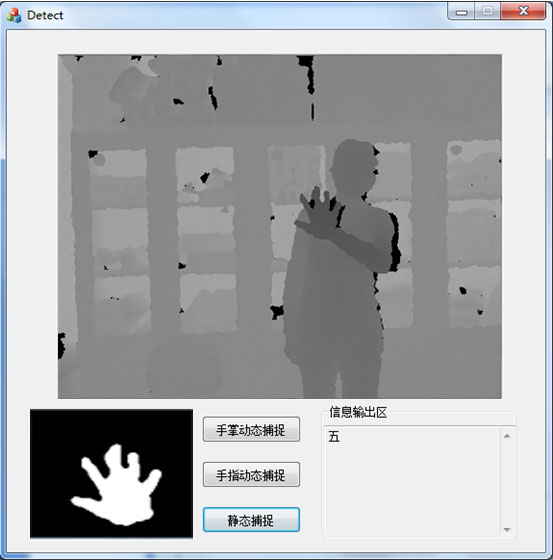
表6-4 手指动态手语识别率

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **手语** | 三十 | 六十 | 九十 | 阿姨 | 职权 |
| **识别率** | 100% | 93.3% | 96.7% | 93.3% | 93.3% |
| **手语** | 组长 | 工人 | 导师 | 家属 | 亲戚 |
| **识别率** | 86.7% | 96.7% | 83.3% | 86.7% | 83.3% |

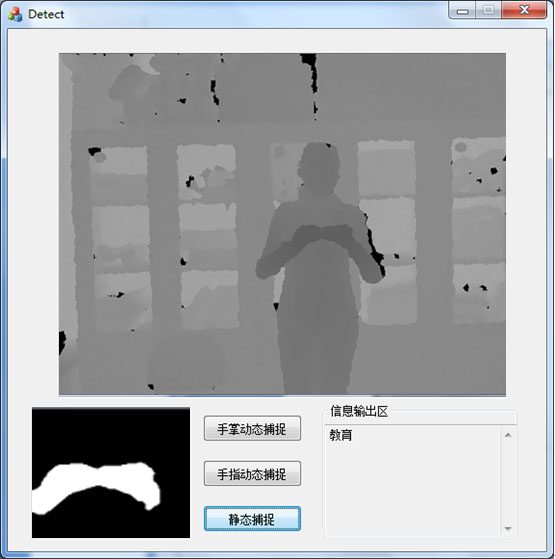
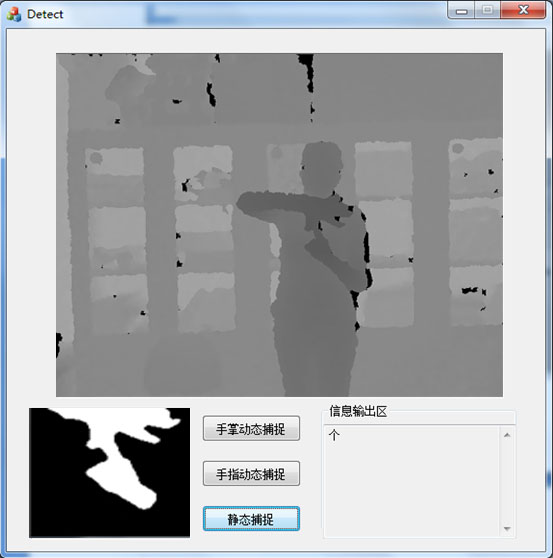
分析表6-4发现总体识别率还是可观的。其中影响识别率的主要是原因是单手与双手手指动态手语彼此误识，这是由于双手动态手语是由单手和双手动作组合而成，当双手距离过远或双手与单手动作切换过快等情况时，双手动态手语会误识为单手动态手语，从而影响了手语的识别率。

1. 静态手语识别

点击“静态捕捉”显示方式类似。提取手部轮廓，提取并处理轮廓的Freeman链码，并代入训练好的SVM模型中进行识别。如图6-10所示，在界面的中间显示整体的深度图像，左下角图像区域选择二值化的手部图像，在右边的“信息输出区”输出识别结果。



(a)五 (b) E



(c)个 (d)教育

图 6-10 静态手语识别

同样每个静态手语选取100个训练样本和30个测试样本。将样本的特征向量和参数及惩罚因子代入SVM模型中训练。得出手语的识别率如表6-5所示。

表6-5 静态手语识别率

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **手语** | A | E | 五 | 十 | 山 |
| **识别率** | 96.7% | 86.7% | 93.3% | 100% | 93.3% |
| **手语** | 个 | 饼干 | 房子 | 王（姓） | 教育 |
| **识别率** | 83.3% | 80 % | 86.7% | 83.3 % | 100% |

由表6-5可知，对于手部外轮廓变化比较平缓的手语，如“十”、“教育”等识别率较高。对于手部外轮廓变化幅度比较大的手语，如“个”、“王”等识别率较低。

1. 本章小结

本章首先阐述了系统的具体实现流程，并给出系统整体流程图，同时叙述了系统的开发环境与部署情况。针对系统实现的不同流程，详细的进行了描述，并给出了关键部分的实现代码，同时列出了系统中不同情况下提取的手语特征。最后，对系统的识别结果进行展示与分析。

结论和展望

本文介绍了手语对聋哑学生的重要程度，以及手语识别在人机交互中的意义和前景。介绍了当今手语识别技术所使用到的常用特征提取方法和模式识别算法，及与本文相关的概念、平台和设备。同时分析了手语识别系统为聋哑儿童所带来的的积极的影响，并阐述了手语识别系统对教学理论的支持。讲述了使用Kinect体感设备实现的手语识别系统流程，讲述了针对手掌运动、手指运动和静态手语这三种情况下特征向量的提取方法。并针对这三种特征向量，介绍了基于HMM和SVM算法的识别方法，并对识别结果进行验证。最后，在上述研究的基础上对系统进行设计与开发，实现了基于Kinect的手语识别系统。

本文研究成果主要体现在以下几个方面：

* 1. 分析了手语对聋哑儿童的重要性，为了促进聋哑儿童与健全人之间的互动及改善人机交互方式，提出了手语识别系统。
  2. 学习了相关的教学理论，并研究了手语识别系统对教学理论的支持。
  3. 阐述了手语识别的流程和目前常用的算法，在此基础上提出了将Kinect应用于手语识别系统的优势与可行性。
  4. 研究了使用Kinect和OpenNI/NITE进行手部跟踪的流程，及手部特征提取方法。本文针对手语的不同形式分三种情况进行了研究。对于手掌动态手语，使用了针对手掌运动方向角来提取特征向量的方法；针对手掌运动较小的情况，提出了根据手指运动，提取深度变化直方图的一阶差分的办法；对于静态手语，使用其轮廓的Freeman链码做为特征向量，同时对Freeman链码进行归一化和旋转不变性的处理。
  5. 在识别阶段，本文针对手掌运动识别提出了改进的HMM模型，即分别提取手掌运动的平面和深度特征向量来训练HMM模型。针对手指动态手语识别，本文使用的是SVM/HMM模型。对于静态手语的识别，本文使用的是SVM模型，并对识别结果进行分析和统计。
  6. 最后针对上述研究进行系统设计，实现了基于Kinect的手语识别系统。

本文取得了一定的研究成果，但是尚存在一些不足及有待改进处。在手指语的特征提取方面，深度信息提取的有些粗糙。在手掌动态手语识别部分，对误识列表中手语在第二层HMM模型中误识的情况没有进一步的分析和研究。忽略了系统的性能方面考虑，没有详细的记录系统的处理时间，只是针对Kinect读取深度的频率进行提取和处理。目前系统只能识别词语，并不能识别整个句子，并不是很完美。

对于这些不足之处，在接下来的继续研究的过程中，将进一步修改和完善，以及对手指深度特征的提取进一步细化和处理，提高特征向量的准确性。使用SVM模型训练和识别时，对于参数和惩罚因子可设置成不同的值，并作进一步对比。在性能方便，可以对系统的处理和识别时间进行统计，争取最大程度上保证其实时性要求。可以将不同的手语整合起来，组合成一个句子，使人们更容易理解打手语人所表示的意思，使系统更加人性化。

# 致 谢

历时一年多的论文终于完成了，在这一年里，从学习资料的搜集到具体的分析与实现，这期间遇到了许多困难和障碍，在完成之际，本人向所有关心我、帮助我、支持我的老师和朋友们道一声最诚挚的感谢。

尤其要感谢我研究生阶段的导师景红教授，景老师在我完成论文阶段，为我提供了良好的科研氛围，每当研究过程中遇到问题，老师都尽力帮我解决。对于研究中出现的不足之处，老师会及时的指出并帮助我改正，保证了我顺利的完成了课堂研究工作。除了在学业上，在生活中老师也给及我无微不至的关怀。每当我遇到挫折时，老师都如母亲般的给予我关爱和鼓励，同时为我提出了宝贵的意见，帮助我克服困难。在此，向景红老师表达我最崇高的敬意。

感谢我的周围的同学，在我写论文遇到困难时，你们给予我无私的帮助，帮助我克服一个又一个阻碍。同时，当我在生活中遇到挫折，需要帮助的时候，你们也无时无刻不在我身边，帮助我安慰我。感觉你们在我研究生阶段一直陪在我身边。

感谢我的父母，感谢父母辛勤把我养育成人，默默的无私的为我付出，陪伴着我从一个幼儿长大成人。我能从一个懵懂无知的少年，成长成为一名研究生，这些与父母对我的爱是分不开的，他们陪伴这我面对人生一个又抉择，度过一个又一个困难，分享一个又一个喜悦，是他们给予我不断学习进取的动力。

最后，衷心感谢对我论文进行评阅和指导的各位专家和教授。

# 参考文献

1. 张亚新 原魁 杜清秀 邹伟. 一种用于手语识别的中国手语分类方法. 北京科技大学学报 .2001年23卷3期
2. 王春立. 面向大词汇量的连续中国手语识别系统的研究与实现. 大连理工大学 2003年4月
3. 陈少毅 兰继军. 手语研究与聋人康复贺聋校教学. 中国特殊教育 2003年第5期
4. 刘信. 基于视频的中国手语识别算法研究. 中国科学技术大学. 2008年5月
5. 王凯. 基于OpenCV的手语识别研究与实现.电子科技大学.2012年6月
6. 甄素娟.浅谈聋哑儿童的人格行为特点与特殊教育. 大连教育学院学报.2011年27卷1期
7. 张中甫. 基于深度信息的手势识别研究及应用.浙江大学.2013年1月
8. G.J.Grimes. Digital Data Entry Glove Interface Device. US Patent 4414537, AT&T Bell Telephone Laboratories, Nov.8,1983
9. M.Fucumoto, K.Mase, Y.Suensgs. Real-time detection of pointing actions for glove free interface. In IAPR workshop on machine vision application, 1992: 473-476
10. Fels, Hinton G.Glove Talk: A neural network interface between a data glove and a speech synthesizer.IEEE Trans Neural Networks, 1993,4(1):2-8
11. Fels. Glove talkⅡ:Mapping hand gestures to speech using neural networks an approach to building adaptive interface[Ph.D dissertation]. Computer Science Department, University of Torono,1994
12. Y Cui, J Weng. Hand segmentation using learning-based prediction and verification for hand sign language[C]. In Proc of ICCV'96, 1996
13. Mohammed Waleed Kadous. Machine recognition of Auslan signs using Power Gloves:Towards large-lexicon recognition of sign language. Lynn Messing, editor, Proceedings of the Workshop on the Integration of Gesture in Language and Speech, pages 165-174,Applied Science and Engineering Laboratories Newark, Delaware and Wilmingtion, Delaware, October 1996
14. B.Bauer, H.Hienz. Relevant features for video-based continuous sign language recognition.Proceedings of the Fourth International Conference on Automatic Face and Gesture Recognition, 2000
15. J.L.Hermamdez,Rebollar,N.Kyriakopoulos,R.W.Lindeman.A New Instrumented Approach For Translating Ameraican Sign Language Into Sound And Text.Six IEEE Int'I Conf.Automatic Face and Gesture Recognition(FG'04), Korean, 2004.
16. O.Aran, T.Burger, A.Caplier, L.Akarun. Sequential belief based fusion of manual and non-manual signs. Gesture Workshop, 2007
17. 戴金波.基于视觉信息的图像特征提取算法研究.吉林大学.2013年5月
18. 孙启艳.人体寄生虫虫卵识别系统研究与设计.江苏大学.2012年6月
19. 周志峰. 一种正六边形检测方法的研究.应用光学.2012年第33卷1期
20. 左玉龙.基于颜色和形状特征的图像检索.山东大学.2012年10月
21. 王志瑞，闫彩良.图像特征提取方法的综述,吉首大学学报（自然科学版）.2011年第23卷5期
22. 陈健斌.图像特征提取及其相似度的研究和实现.西安电子科技大学.2012年6月
23. 周丽莎.基于模板匹配的视觉定位技术研究与应用.大连理工大学.2012年6月
24. 刘青.人工神经网络遥感影像分类系统的设计与实现.昆明理工大学.2012年4月
25. 苏红刚.基于SVM的中文文本分类系统实现.吉林大学.2012年10月
26. 杨健等.基于隐马尔可夫模型的文本分类算法.计算机应用.2010年.9月
27. 许杏. 基于隐马尔可夫模型的手势识别研究.华南理工大学.2011年5月
28. 柯福全. 基于视频运动捕捉的肢体康复虚拟现实系统的设计.中南大学.2012年5月
29. 侯穆. 基于OPENCV的运动目标检测与跟踪技术研究.西安电子科技大学.2012年1月
30. 林小文.论“罗森塔尔效应”在高中思想政治课教学中的运用.山东师范大学.2012年6月
31. <http://baike.baidu.com/link?url=TEQF35oFO4jfmutMKZwfiaBa87IfROAbPM7OYJb5WOzGuzyQdkDnW-7pL9T4lWWg-PEyUelPSPBnRDhEl5e7FK>
32. 张立勤.基于学习金字塔理论的初中数学试卷讲评课的研究与实践.上海师范大学.2013年5月
33. 雷刚.人本主义学习理论对教育技术的新启示.中国电化教育.2010年6月
34. 杨宝燕.浅谈聋哑儿童学习动机存在的问题、原因及对策.《新课程-中旬》.2011年11月
35. 常亚南.基于HMM的动态手势识别.华南理工大学.2012年5月
36. 申晓霞，张桦等.基于深度信息和RGB图像的行为识别算法.模式识别与人工智能.2013.08
37. Yoon H-S, Soh J, Bae Y.J, et al. Hand gesture recognition using combined features of location, angle and velocity [J]. Pattern Recognition, 2001, 34(7): 1491-501.
38. 南睿.基于单目视觉的手指识别跟踪系统设计.哈尔滨工业大学.2013年6月
39. http://www.cnblogs.com/tornadomeet/archive/2012/10/18/2728896.html
40. 于仕琪，刘瑞祯等.学习OpenCV.清华大学出版社.2009年10月
41. 孔新明.基于形状的图像检索算法研究.哈尔滨工业大学.2010年6月
42. A.B.Poritz.Hidden Markov Models:a guided tour.Proc.of IEEE Int1.Conf. on Acoustic

SPeech and Signal Processing.1988.

1. Elmezain M, Al-Hamadi A. Gesture Recognition for Alphabets from Hand Motion

Trajectory Using Hidden Markov Models[A]. IEEE International Symposium on Signal Processing and Information Technology[C]. USA: IEEE, 2007: 1192-1197

1. 许杏.基于隐马尔可夫模型的手势识别研究.华南理工大学.2011年5月
2. 江超.实时视频中的手势轨迹识别及应用.广西大学.2012年6月
3. Richard O.Duda,Peter E.Hart,David G.Stork.Pattern Classification.机械工业出版社.2013年10月
4. 陈静.基于Kinect的手势识别技术及其在教学中的应用.上海交通大学.2013年1月
5. 王园.基于SVM\_AdaBoost模型的上市公司退市预警研究.华南理工大学.2013年3月
6. 吴迪.基于SVM分类器的分步定位算法研究.哈尔滨工业大学.2010年7月
7. 傅智勇.HOG+SVM行人检测算法在DM6437上的实现与优化.华南理工大学.2012年12月
8. 李玉朵.基于SVM的人脸表情识别研究.河北工程大学.2012年5月
9. http://viml.nchc.org.tw/blog/sub\_class.php?SUB\_ID=1&CLASS\_ID=1

# 攻读硕士学位期间发表的论文

1. 朱明茗，景红. 基于Kinect的PPT控制系统的研究. 中国教育信息化. 2013.11
2. 朱明茗，景红. 基于体感技术的手势追踪与识别研究. 计算机系统应用.

2014.1

**References:**

王松林 (2014). 基于Kinect的手势识别与机器人控制技术研究, 北京交通大学. **硕士**: 82.