中山大學
SUN YAT-SEN UNIVERSITY

本 科 生 毕 业 论 文

题　　目：应用于自闭症儿童诊断的手势识别系统

院　　系：数据与计算机科学学院

专　　业：软件工程（移动信息工程）

学生姓名：刘爽

学　　号：13354222

指导教师：李明（副教授）

二〇一七年四月

附表一、毕业论文开题报告

论文（设计）题目：应用于自闭症儿童诊断的手势识别系统

（简述选题的目的、思路、方法、相关支持条件及进度安排等）

现在对于自闭症诊疗的诊断，基本上是由诊疗师一人完成，越有经验的人越能娴熟快速地分辨出一个小孩子是否患有自闭症。计算机所能起到作用的地方非常少。而随着社会的发展，人们越来越重视对心理疾病的预防和诊治，未来诊疗师可能一天要诊治很多个患者。在这样的前提下，对用计算机来减轻诊疗师负担的需求越来越大，这是一个无法逾越的话题。而现在，对计算机和自闭症诊断交叉领域的研究颇少。如果能做到将自闭症诊断中的非主观因素当作特征提取出来并能让计算机识别，会大大降低诊疗师的工作量。

目的：辅助治疗师对小孩进行的自闭症检测

思路：

在自闭症检测中，有一套标准的检测程序 ADOS。ADOS 分为三个模块，每个模块对小孩的能力要求都不同。第一个模块可能只需要小孩简单的回应，而第三个模块可能需要小孩子更加复杂的回应模式。每个模块下，有许多个小的单元，每个单元都类似一个问答的小游戏，而治疗师则会根据小孩在每个单元内的表现依次打分，打分的标准也是有参考的标准的。测试的最后，治疗师会根据小孩在每个模块中的分数进行整体评估，来决定小孩是否有自闭症。

为了能辅助治疗师进行治疗，我们将整个 ADOS 的评估过程用摄像机和 Kinect 录了下来。通过分析视频中小孩的动作以及对治疗师的提问或动作的回应来判断小孩是否有患有自闭症的特征并将其告知治疗师从而达到减少其进行判断是否为自闭症的时间。

具体到这次选题的部分，在 ADOS 中，治疗师会根据小孩是否有手指指向指定物体，指向物体的时间长短等现象来判断小孩是否具有自闭症。我们需要做的就是判断治疗师发出具体的指令后，小孩是否快速的指向了指定物体，并且判断从发出指令到指向物体的时间长短，并将这些反馈给治疗师。

方法：

首先截取整段视频，将治疗师发出指令后的一小段视频截取出来以提升效

率；

接着处理截取出来的视频，将视频中的小孩子单独圈出来；只提取圈出来的小孩，利用 RGB 信息以及深度信息将小孩子的手分离出来；

将小孩的手单独提取出来之后，可以单独拿出手的 RGB 信息和深度信息。可以通过分析手腕与手肘的角度以及手指与手腕的角度来分析手指的指向位置；

又通过多根手指指向和单根手指的深度信息的不同我们可以识别出小孩子是伸出了一个手指还是多根手指。

对于判断小孩反应时间的要求，可通过 hand tracking 来记录从治疗师发出指令到小孩的手指刚好指向指定物体的时间，从而帮助治疗师进行诊断。

课题的目标有三个阶段：

第一个阶段：将 hand segmentation, hand tracking 和 hand gesture recognition 在理想数据上得到应用；

第二个阶段，将可以用于理想数据的技术应用与现实的自闭症诊断视频中去；

第三个阶段，将已实现的技术整理成一个有 API 接口的 demo，并更加深入地优化这个 demo。

相关支持条件：

此课题在中山大学-卡耐基梅隆联合研究院 SMIIP 实验室的支持下进行。该实验室有充足的设备资源，实验条件和资金，因而可以完全支持这个课题的进行。

研究步骤如下：

2016 年 11 月中旬-12 月中旬：

在理想样本下，完成 hand tracking

2016 年 12 月中旬-1 月中旬：

在理想样本下，完成 hand segmentation

2016 年 1 月中旬-2 月中旬：

在理想样本下，完成手指指向的判断

2016 年 2 月中旬-3 月中旬：

尝试将小孩单独提取出来，将 hand tracking 和 hand segmentation 应用于实际视频中判断效果。

2016 年 3 中旬-4 月中旬：

将手指指向的判断应用于实际视频中并将技术整理成 demo。

Student Signature：　　Date:

指导教师意见

Comments from Supervisor：

1.同意开题　　2.修改后开题　　3.重新开题

1.Approved(√)　2. Approved after Revision (　)　3. Disapproved(　)

Supervisor Signature：　　Date:

附表二、毕业论文过程检查情况记录表

指导教师分阶段检查论文的进展情况（要求过程检查记录不少于 3 次）：

第 1 次检查

学生总结：

After getting the hand segmentation, the further step is to get a more precise hand segmentation and try to recognize the hand gesture.

After modification, this new color model can easily identify the difference between the face and the hand, and therefore always rectangle the hand corrcectly.

• Principle

On the threshold, we get all the skin by using the HSV color space;

Then, we get the contour of all these skins. Assuming that the hand contour is bigger thant the face contour, we can extract the hand exclusively by selecting the maximum congtour.

Finally, we use a red rectangle to indeify the segmented hand.

• Issue

If the face area is larger than the hand area, then the segmentation result will be wrong.

Fingertip recognition

As we need to know the exact number of fingers the children are extending when pointing to the toy, we need to count the fingers of the hand.

• Convec Defects

After we get the contour of the hand, we can get the convex defect of the contour, which can help us identify the fingertip. All the convex defects are annotated by full circles in the frame. We want to get the convex defects between two fingers, and ingore all other convex defects that are useless.

As people's hand can not open too widely, we can recoginze the convex

defects between two fingers by setting the angle threshold. That is to say, if the angle between the fingertip and the convex defect is less than 80 degrees, the convex defects is useful, which is annotated by yellow. And other useless defects are annotated as blue.

• Fingers Counting

Because the convex defects between every two fingers are annotated by yellow circles, a simple and straightfoward idea is to count the number of the yellow circles.

• Issue

The result is not stable, keeping changing along with the frames, because of the raw calculation method and the simple model. The finger counting method needs to be improved.

指导教师意见：

关于 theano，你可以试试用 python 从 keras 搭一个，keras 是基于 theano 的，应该比较容易些。人脸和手的判别，也许不一定要用 contour 的大小，可以结合人脸检测来做。医院里面指物的数据，我去帮你看看，是否 kinect 的数据质量满足要求如果不满足，我们可能需要专门设计一个你最容易处理的小场景，比如在小孩的侧面放一个 kinect。

第 2 次检查

学生总结：

The position of the Kinect is in front of the whole scene which is not what before. Maybe I have another kind of dataset here...

If it is the dataset I need to cope with, I guess what I need to do is to extract the kid alone and start the hand segmentation.

Frame recorded by Kinect:

指导教师意见：

去医院跟治疗师沟通一下，看一下能不能确定一个双方都能接受的范式，即你可以识别，他们也可以进行诊断治疗。

第 3 次检查

学生总结：

PartI：

Using the vector between the index mid point and the index front point, we can recognizewhether the index finger is vertical to the screen, which is the Kinect, or not, by calculating the angle between the index finger and the screen.

We suppose that when the angle is smaller than 30 degrees, then the index finger is vertical to the screen.

The algorithm is applied to the NYU dataset, and the result is showed in the video called'NYU_index_pointing'.When the index finger is vertical to the screen, the'Vertical' text will be showed in the left-up corner.

PartII

The hand of the tester can be successfully segmented, and the result is satisfying.

The result is showed in the video called 'hospital_test_segmentation'.

However, there are some problem when I am trying to recognize the fingertip.

Issues:

The depth frame and the color frame is not synced.

The size of the depth video dose not equal to the size of the color video.

指导教师意见：

你可能需要师兄之前做的基于 Kinect 的深度和 rbg 图像的对齐。

可以找他交流一下。

第 4 次检查

学生总结：

目前已经做完了视频中的指尖垂直摄像机检测，视频是同学的 sample。

指尖垂直于摄像头时，左上角会显示"Pointing！"。

医院的数据上周问了诊疗师，她说这周才有空录新的视频。

就是新的数据。

指导教师意见：

好，你还要把脸和手分开，现在混在一起医生不容易看。

另外，可以考虑开始写你的毕业论文的前面 2 章。

<table>
<tr><td colspan="2">等医院数据到了之后，就可以完善后面的实验结果了。

学生签名：　　年　月　日

指导教师签名：　　年　月　日</td></tr>
<tr><td>总体完成情况</td><td>指导教师意见：

1、按计划完成，完成情况优（ ）
2、按计划完成，完成情况良（ ）
3、基本按计划完成，完成情况合格（ ）
4、完成情况不合格（ ）
指导教师签名：　　年　月　日</td></tr>
</table>

附表三、毕业论文答辩情况

答辩人		专 业	
论文（设计）题目			
答辩小组成员			
答辩记录： 记录人签名：　　年　　月　　日			

学术诚信声明

本人所呈交的毕业论文，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料均真实可靠。除文中已经注明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的作品或成果。对本论文的研究作出重要贡献的个人和集体，均已在文中以明确的方式标明。本毕业论文的知识产权归属于培养单位。本人完全意识到本声明的法律结果由本人承担。

本人签名：　　日期：

【摘　要】

在以往自闭症检测中，诊疗师的主观检测成分占比很大，极少有机器的参与。虽然有标准化和模式化的 ADOS 检测流程，但是整段过程的评分都是靠诊疗师的主观打分。一方面诊疗师的负担很大，另一方面，诊断的结果可能因诊疗师不同而有很大不同。这篇文章首先提出了手势识别的方法，利用肤色模型，形态学操作，其模糊化和二值化，轮廓识别，以及脸部识别帮助计算机自主识别视频中的人是否有指向摄像机，提取出与人手指向发声物体有关的指标。随后我们将其用于自闭症检测中，自闭症检测全程由 Kinect 录制。在自闭症检测中，被测者是否在受到声音刺激时有指向发声目标是评判一个人患有自闭症概率的重要指标。我们希望将这个客观指标用计算机提取出来，以减轻诊疗师的诊断负担，并加大诊断中的客观成分。

【关键词】手势识别；自闭症诊断；Kinect

【ABSTRACT】

In the autism diagnosis, the diagnosis result is determined by doctor in a subject way, with little machine interpretation. Although there is a standard and structured ADOS diagnosis process, the score of every activity or module is also totally determined by the doctor. This leads to much burden on psychotherapist, and diverse diagnosis analysis by different therapists. This paper proposes a system that can identify whether the finger is pointing to the camera or not utilizing the color model, morphological operation, blur, binary operation, contour detection and face recognition. Then, we apply this system on autism diagnosis, the dataset of which is recorded by Kinect, which is often used in recognition area. In the autism diagnosis, whether the individual tested points to the object which is making noise when he or she is stimulated by that sound is a very important indicator of the percentage of having autism issue. We hope this kind of object indicator can facilitate the diagnosis process, and increase the ratio of object element in the diagnosis.

【**Keywords**】Hand Recognition; Autism Diagnosis; Kinect

目录

应用于自闭症儿童诊断的手势识别系统

第一章 概述/引言

1.1 背景和意义

2016 年《中国自闭症教育康复行业发展状况报告》蓝皮书显示，自闭症的发病率增长速度快，数量惊人。从万分之一、千分之一、百分之一，到 2016 年的 1/45。在我国，依据已有调查数据保守估计，发生率也在 1%--也就是说，在我国 13 亿人口中，可能有超过 1000 万的自闭症人群、200 万的自闭症儿童，并以每年将近 20 万的速度增长。研究表明，进过早期科学的干预，不同病理程度的自闭症人群能相应地减轻社会照顾、拥有一技之长、甚至为社会做出贡献。很多研究发现，最早于 6 个月时即可发现患儿在情感交流、社交互动及运动技能上与正常儿童存在差异。儿童早期发展的可塑性更强，越早开展高质量的干预，越能最大程度优化自闭症儿童的预后。所以，早怀疑，早评估，早诊断，非常重要。而我国自闭症谱系障碍儿童，在接受早期干预前的几个关键节点上都存在“时间延迟”。另一方面，自闭症从国外引入国内，从科学确认到大规模治疗干预仅有二十年的历程。由于没有建立完整的鉴定体系，特别是由于医学基础科学对其没有建立完整、系统的生物医学机制理论，所以自闭症还不能通过医疗仪器检测出来，主要依赖医生通过国际标准行为量表和个人经历进行判断。所以，尽快解决当前早期筛查诊断专业人员缺乏的问题，对于及时诊断出自闭症儿童，干预治疗并令其过上正常的生活有重要意义。[1]

ADOS 是评估疑似自闭症或其他广泛性发展障碍人群的交流，社交和娱乐能力的半结构化方法。ADOS 由四个模块组成，每一个分别适应不同的儿童和成人的发展和语言水平，包括能流利交谈的和无交谈能力的。ADOS 由标准化的活动组成，测试者需要观察对于诊断自闭症或其他广泛性发展障碍的重要行为的出现与否。测试者基于被试者的表达语言能力和年纪选择最适合他们的模块进行测试。结构化的活动，材料以及半结构化的交流提供了能够观察到与广泛性发展障碍有关的社交，交流等行为的标准化环境。在每个模块中，参与者对于每个活动的反应都会被记录下来。在测试的最后，会有一个总体的打分评判。通过使用每个模块中

的诊断算法，这些评判可以较清晰地表达出诊断结果。事实上，ADOS 提供 30 至 45 分钟的观察期限，在这段时间内，测试者会通过标准的交流和社交类的压力测试给被试者多个机会展示出与自闭症诊断有关的行为。压力测试由一些计划哈的社交场合组成，而那些有可能发生的特定行为是已经被提前决定好了的。模块都提供一系列的结构化场景，每个场景为特定的社会行为提供不同的压力测试。模块一位无法使用短句（如无法回应他人对话，经常出现自发而无意义的字词组合）的人而设。模块二为可以使用少量短句但是无法流利交谈的人而设。模块三为有能力流利交流的儿童而设（一般是 12-16 岁）。交流流利的广泛定义为能说出一系列带有语法规则的长句子，并且句子之间具有逻辑关系（如虽然，但是）。模块四包括了模块三的很多任务，以及一些关于日常生活的采访式元素。它是专门为能流利交流的青少年和成人所设。模块三和模块四中的区别主要在于有关社交的信息是否更加恰当和容易的在玩乐和交流采访中获取到。

四个模块有活动上的交叉，这些活动小至模块一的观察儿童测试者如何要求测试者持续冲大模块一的气球，大至模块四的谈论关于学校和工作社会关系。模块一和模块二的活动会在一个固定房间中的不同位置进行，这样才能在有限的语言下反应出儿童的兴趣以及互动水平。模块三和模块四更多的是坐在一张桌子前，在没有外界固有环境下的言语交谈。虽然表面上看每个模块都是很不一样的，但是它们都有着相同的原理，即通过结构化和非结构化的社会行为来使小孩有特定的反应。

虽说最终的诊断结果是依据医生的主观判断，但 ADOS 提供的一系列标准化活动使得我们能从这些诊断过程中提取出一些客观的指标，令这些指标在非人工的情况下也能被提取出来，并用其辅助医生进行诊断，减少诊断的时间，缓解诊断的压力。模块二中，有一个特定的单元，考验儿童的反应能力和表达能力，名字叫指兔子测试。具体的情境是，医生在儿童背后放置一个会发出声音的兔子，当兔子发出声音后，儿童面前的医生要用手指向兔子，并在口中喊着兔子，以此来吸引小孩子的注意力。医生需要观察小孩是否有转头看向兔子，并且是否有用手指指向兔子。如果要将这一系列过程转化成可以为医生所用的客观指标，有几点需要我们注意。首先，这一系列的过程非常长，在这很长的一段时间内，对分析有用的其实只有从小孩转头那一刻到小孩指完兔子并放下手指那一刻。在这段

时间之前的只是医生的个人行为，对于诊断并没有什么用。其次，判断小孩有没有指兔子并不能给诊断带来太大的帮助。虽然由分析数据发现，自闭症小孩大部分都不会主动指兔子，但是有一些正常的孩子也不会主动指兔子。那么，我们需要提取一些关键的指标，使得自闭症小孩的指兔子过程具有和正常小孩子不同的特征。这里我们考虑几个关键指标，从小孩子转头到小孩子举起手指的时间，从举起手指并指到兔子的时间，指完兔子并放下来的时间。我们认为自闭症小孩和正常小孩所拥有的这些指标是不同的。[2,3]

随着微软推出了 Kinect，越来越多的研究人员将 Kinect 应用于他们的研究领域当中。Kinect 是微软开发出来的集深度摄像机，色彩摄像机和红外摄像机一体的设备。Kinect 同时获取色彩信息和深度信息并提供实时，可靠的 3D 重建来使得人体或人体的一部分可以当控制的工具。Kinect 为高级别以及自然的人机交互领域开了新的篇章。从游戏到医疗领域，许多领域都在逐步发展中。我们的实验采用 Kinect 为我们采集数据。不同于以往单纯的色彩摄像机，Kinect 提供同步的深度图像。深度帧提供物体离摄像机之间的距离，从而可以给我们额外的信息。在这篇文章中，我们提出了一个手势识别系统，识别小孩从转头到抬起手指，指向目标物的和收回手指的时间等关键指标，来反馈给医生。这个系统可以大大减少医生进行诊断的时间。

1.2 识别问题的描述

从动作识别的角度看，分析小孩的社交行为带来了很多现存的数据集中不存在的问题。首先，交流的二元性我们需要清晰地定义出互动的模型，比如说，每个环节的时间，以及每个环节的内容等。其次，社交性行为本质上是多模态的，需要综合视频，音频和其他的模态来完成对一个行为的刻画。另外，社会互动经常被定义为参与者的参与度，而不是以他们所完成的具体的任务来定的。最后，因为那些需要被测试者做的环节并不是强制性的，参与者完成任务需要的时间可能会超出预料之外。

在评估和诊疗的目的下，成人与小孩之间的互动分析给心理治疗师和计算机科学家提供了一个合作的机会，去解答关于年轻小孩的早期发育的一些问题。例如，通过检测小孩的姿势，面部表情，言语以及注视成人的视线可以让诊疗师判

断小孩的行为是否是具有社交性和目的性的。另一个挑战是，心里治疗师和计算机科学家需要判断小孩交流性行为的目的或者功用是什么。当一个小孩说话或者做手势时，他们的目的是要求诊疗师给他们一个物体，展示一个行为，还是引导诊疗师的视线至他们认为有趣的东西上，又或者是仅仅是想继续维持一份社交活动而已，因为小孩可能觉得游戏很好玩，不想停下来。回答这些问题需要我们想出一个新的方法来解答视频中的那些行为。[4]

为了实现这个系统，有几个问题需要解决。首先，整个实验场景中的干扰因素太多，视频中除了小孩之外，还会有小孩的家长，以及医生。所以我们要先将小孩单独提取出来，然后再进行手势识别和特征提取。其次，小孩本身的干扰因素也需要考虑进去。我们需要专门识别出小孩用于指的那只手，并且定位其位置和形状，从而达到识别的目的。再者，我们仍需要分辨出小孩的手是否指向了兔子。在这个实验场景中，兔子放在了 Kinect 的正下方，所以我们可以将小孩指向兔子的时刻看作小孩指向 Kinect 的时刻。而需要解决的问题就变成了，如何判断小孩在某一帧上是否指向了 Kinect。

1.3 文章的工作

在这篇文章中，我们主要做了下列这些事：

• 通过手部检测，脸部识别和物体识别，我们将小孩从复杂的环境中单独提取了出来，大大降低了噪音，保障了后面的手部分割和手势识别的正确性，避免了检测到其他人体的部位上去的情况。

• 通过手部分割，我们将小孩的手部单独的分离出来，有利于单独判断手部的形状和指尖指向。小孩其他部位的动作也会极大的影响识别的结果，如小孩没有动的另一只手，如果不单独将目标的首部提取出来，识别结果会不尽人意。

•通过手部的轮廓特征和相关的拟合图形，我们可以判断出手指的具体朝向，判断它是否指向了 Kinect，并将具体的数据显示出来。再结合脸部识别，通过这些帧分析提取小孩在收到医生提示之后到举起手指指向 Kinect 的时间，从手指一直指向 Kinect 到手指放下来的时间等关键指标。

1.4 论文结构

之后的内容如下，第二部分介绍了在手势识别领域和自闭症领域中行为识别相关的国内外相关的研究方法和成果。第三部分说明了这篇文章的具体实现方法。第四部展示并分析了实验结果。第五部分为结论与对未来的展望。

第二章 综述

手势是一个本能就能判断并简单的信息，它能有效的用于控制各种各样的东西[5,6]因此，关于手势识别有很多研究。在这些研究中，输入设备是 Kinect 和照相机，并且用肤色，深度信息，或者同时使用两者来进行手部区域的识别。

Han 使用肤色来检测手部区域。第一步是通过肤色模型来检测用照相机拍下的一张照片[7]。然而，如果被检测的手部区域包括了手臂，通常检测结果会出现错误。手臂区域一般可以通过手部的几何特征而分离出来。当手部区域被找到之后，不管有没有包含手臂，找到的手部区域的中心点都要显示出来。在由距离转化过来的像素值中，中心点是有着最大的像素值的。当中心点被识别出来之后，需要画一个圆来表示手部区域的手指个数。而这个圆的半径通常是中间点到手部区域的 1.5 倍大小。当圆画好之后，圆和手部区域重叠的部分由顺时针方向一一检测出来。一一计算不重叠区域的大致夹角，如果夹角小于 10 度，则视为检测到的手指增加了一个，如果检测到的夹角大于 25 度，则认为检测到了手腕区域。但是 Han 的方法没有解决会检测到脸部的问题。在实验数据中，人脸也会出现在视频当中，Han 并没有解决这个问题。

Jagdish 使用 Kinect 的深度图像以及使用 OpenNI 模块来检测手指的个数[8]。他利用深度图像的深度值将背景和手部分离开来。一般来说，当你展示一个姿势的时候，手与 Kinect 之间的距离应该是最近的。Kinect 的深度图像提供的深度信息，即显示了手距离 Kinect 之间的距离。因此，使用深度图像可以得到手部区域。在得到手部区域之后，手部的中心则再次用距离的转化得到。然后，手掌区域被移除，剩余的部位可以根据深度值得到手指的个数。然而 Jagdish 的方法在此实验数据中并不适用。当小孩的手在动的时候，光靠深度信息并没有办法将手部分离出来。小孩的手部并不总是距离摄像机最近的。

Tao 则同时使用肤色模型和深度信息来识别手部区域[9]。首先，手部区域可以通过设置深度阈值来得到。其次，使用 YCrCb 颜色空间可以检测到与肤色相近的区域。由深度信息得到的图像和由肤色信息筛选得到的图像的重合部分将被认为是真正的手部区域。而手指的个数则被认为是被最终手部区域有用最小深度值得区域数。遗憾的是，像之前说的一样，深度信息在手会前后移动的情况下是不可信的。

另外，在手部被提取出来后，为了识别姿势。Choi 根据手掌的几何特征来识别手部的姿势[10]。他依靠已有的特征向量数据集来决定姿势和手的形状，包括提取的特征。

现在有很多基于视频的活动和行为识别[11,12]。然而，大部分的工作专注于只有一个人的场景，或者是两个人之间非常简短的交流行为，比如拥抱了一下[11]的行为。在只有一个人的场景的情况下，比如自主地准备早餐[13]，个人的活动会是复杂的并且持续时间会特别短。然而，当一个人是在执行一个特定的任务时，比如遵照食谱的步骤在做早餐，这样就会是一些可预料的行为，时间也是可以估计的。而比起这些研究，成人与小孩子之间的社交领域给研究者带来了很大的挑战，因为当谈话或交流进行时，小孩和成人之间的交流不再是单一对象的，而是二元的。并且小孩子的行为是不能预估的，多模态的。

近来，一些研究者解决了识别人群中的社交行为的问题[14,15]。有些研究者甚至给 Youtubue 上视频的社交游戏进行分类[14]，而这些视频中都存在很多成人与小孩的二元交流的。但是这些工作只是给群体的行为进行粗略的分类，比如从对话中提取一个人的自言自语等。而在这篇文章中，我们希望能够明确找到一个社交的动作的描述，比如小孩的参与度，就表示在小孩是否有主动去指兔子上。
在这篇文章中，在成人，即诊疗师和小孩在进行交流时，我们注重于小孩子的反应。特别的，作为一个二元分类问题，我们在这里需要判断小孩有没有指向摄像头。而有没有指向摄像头，则会成为诊疗师进行自闭症诊断的一个重要指标。而相关的指标如时间等，也能起到帮助诊疗师进行诊断的作用。

第三章 手部区域检测方法

图 3-1 显示了手部区域检测方法的大致流程，首先从视频中一帧一帧的读取图片，然后进行肤色分离和形态学操作，再讲彩色图片二值化成黑白图片，用轮廓检测检测手部区域，用脸部检测避免脸部区域的干扰，再用椭圆拟合检测手指是否指向 Kinect。

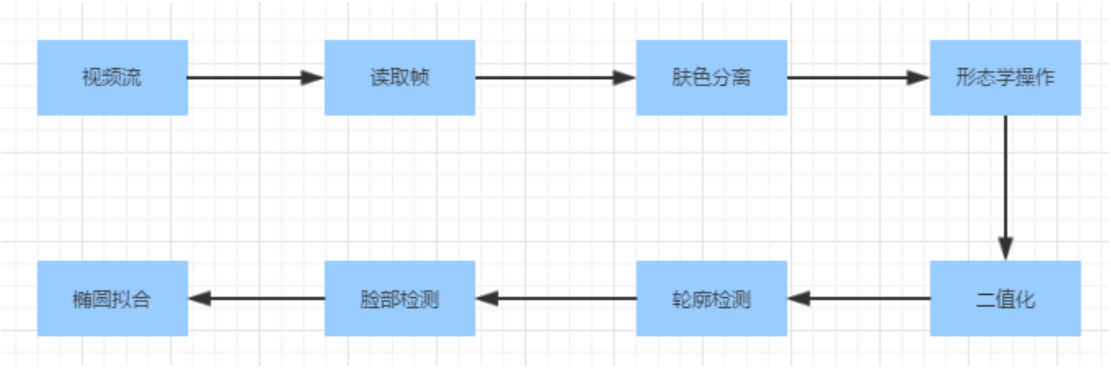

图 3-1：基本流程图

3.1 肤色模型

在手势识别系统中，最主要也是最基础的就是给定图像中的手部的检测和分离。手部的检测和分离很重要因为，它与任务相关的像素数据和与任务无关的图片背景分割开来，用这些提取出来的数据来进行下一步的操作。在很多文章中，都提到了相应的方法来进行手部的检测与分割。通过分析图像的各种视觉特征，或者是它们的组合来分离。例如说，皮肤的颜色，手部的形状，手部的移动以及手部的解剖模型等。有文章专门比对了这些方法。[16]在这篇文章中，我们着重用肤色模型来检测手部。

首先，我们需要转换颜色空间。颜色空间，用于在某些标准下用通常可接受的方式对彩色加以说明。颜色空间有许多种，常用有 RGB，CMY，HSV，HIS 等。绝大部分视频中，使用的都是 RGB 颜色空间。颜色空间从提出到现在已经有上百种，大部分只是局部的改变或专用于某一领域。在这篇文章中，我们使用的是 HSV 颜色空间。相对于 RGB 的立方体模型，HSV 的圆锥体模型在用于指定颜色分割时，有比较大的作用。Androutsos[17]等人通过实验对 HSV 颜色空间进行了大致划分，

亮度大于 75%并且饱和度大于 20%为亮彩色区域，亮度小于 25%为黑色区域，亮度大于 75%并且饱和度小于 20%为白色区域，其他为彩色区域。对于不同的彩色区域，混合 H 与 S 变量，划定阈值，即可进行简单的分割。H 参数表示色彩信息，即所处的光谱颜色的位置。该参数用一角度量来表示，红、绿、蓝分别相隔 120 度。互补色分别相差 180 度。纯度 S 为一比例值，范围从 0 到 1，它表示成所选颜色的纯度和该颜色最大的纯度之间的比率。S=0 时，只有灰度。V 表示色彩的明亮程度，范围从 0 到 1。RGB 转化为 HSV 的算法如下所示：

```
max=max(R,G,B);
min=min(R,G,B);
V=max(R,G,B);
S=(max-min)/max;
if (R = max) H =(G-B)/(max-min)* 60;
if (G = max) H = 120+(B-R)/(max-min)* 60;
if (B = max) H = 240 +(R-G)/(max-min)* 60;
if (H < 0) H = H+ 360
```

通过统计可得，人皮肤的色度是在 9<h<15，50<s<255，50<v<255 之间，因而我们就设置相应的阈值，将不在这个阈值之间的像素点都变成黑色。

3.2 形态学操作

形态学操作是基于形状的一系列图像处理操作。通过将结构元素作用于输入图像来产生输出图像。最基本的形态学操作有两种，腐蚀与膨胀。它们一般用于消除噪声，分割独立的图像元素以及连接相邻的元素。寻找图像中的明显的极大值或极小值区域。[18]例如有一个图形如图 3-2 的左图所示。我们将通过图像的示意来展现膨胀和腐蚀对图像产生的影响。

图 3-2：原图及进行形态学操作之后的二值图

膨胀：

此操作将图像 A 与任意形状的内核(B)，通常为正方形或圆形,进行卷积。内核有三种形状，矩形，交叉形，椭圆形。内核 B 有一个可定义的锚点，通常定义为内核中心点。进行膨胀操作时，将内核 B 划过图像,将内核 B 覆盖区域的最大相素值提取，并代替锚点位置的相素。显然，这一最大化操作将会导致图像中的亮区开始”扩展”。对上图采用膨胀操作我们得到图 3-2 的中间那张。从图 3-1 中间的图可以看得出来，背景(白色)膨胀，而黑色字母缩小了。

腐蚀：

腐蚀在形态学操作中与膨胀通常是成对存在的。它提取的是内核覆盖下的相素最小值。进行腐蚀操作时，将内核 B 划过图像,将内核 B 覆盖区域的最小相素值提取，并代替锚点位置的像素。以与膨胀相同的图像作为样本,我们使用腐蚀操作。从下面的结果图我们看到亮区(背景)变细，而黑色区域(字母)则变大了。如图 3-2 右边的图看出来。

用于彩色图像时，得到的结果如下所示。原图如图 3-3 所示。

图 3-3：彩色原图

腐蚀和膨胀操作之后，结果如图 3-4 所示。

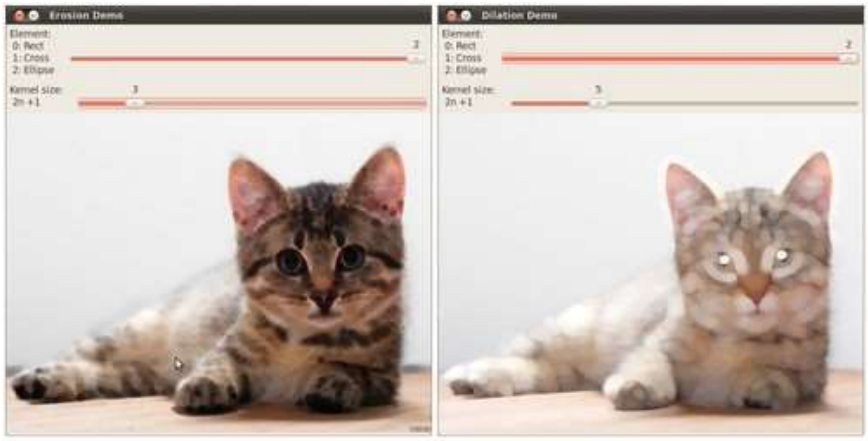

图 3-4：进行膨胀和腐蚀操作之后的彩色图像

在这篇文章中，我们对应用了肤色模型检测的图像进行形态学操作。我们对图像进行两次形态学操作，我们给第一次形态学操作定义了两个核，一个是正方形核，另一个是椭圆形核。分别是 11×11 和 5×5 大小。第一次对图像依次进行膨胀，腐蚀，膨胀操作。得到第一次形态学操作的图像后，将图像模糊化处理，进行第二次操作。在第二次操作中，定义两个新的椭圆核，分别是 8×8 和 5×5 大小。用两个椭圆形核依次对第一次操作后得到的图像进行两次膨胀操作。

3.3 去除脸部影响因素

对第二次形态学操作得到的结果进行模糊化处理，然后进行二值化。
二值化的好处在于，我们可以通过提取二值图像的轮廓而分离出手部区域。
轮廓可以简单的解释为顺着物体边缘的连续点的曲线，它对于形状分析和物体检测与识别非常有用。在这里我们默认，手部区域比人体的脸部区域小。在皮肤区域只有手部和脸部的情况下，我们只需要比较两个轮廓区域的大小，然后选择更小的那个轮廓区域即可。

这里有个问题是，在人手移动的过程中，人手的轮廓区域并不总是比人脸的区域小的。我们需要一个方法解决当人手的轮廓区域比人脸的轮廓区域大的时候的出现的问题。这里采用人脸识别算法，将视频中的人脸给识别出来。然后将轮廓区域与脸部区域进行重叠率比对，如果两个区域的重叠率大于 60%，则我们可以姑且认为是同一个区域，那么这个轮廓区域即是脸部区域。那么另一个区域就是手部区域了。

3.4 手势识别

为了识别人的手部是否有指向摄像头，我们需要用到椭圆拟合。如图 3-5 所示。在上一小节提到过，手部的轮廓实质上是一系列的物体边缘的连续的点。拟合的意思就是用椭圆将二维点包含起来。这里我们使用的是直接最小二乘拟合算法，它是一种非迭代的椭圆拟合算法。[19]通过拟合，我们可以得到椭圆的旋转角，即近似的可以看做手指与摄像头所成的角度。基于误差的考虑，手指不可能与摄像头刚好成九十度，所以我们认为手指与摄像头的角度为 80 到 100 度之间手指就是正指着摄像头。

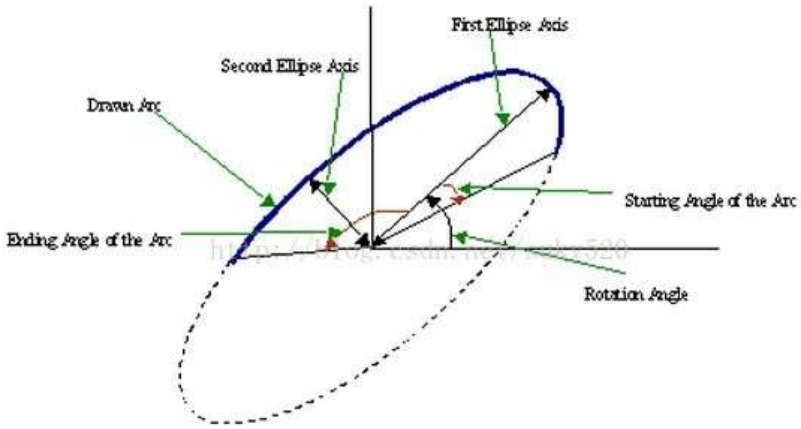

图 3-5：椭圆拟合示意图

第四章 实验结果

4.1 多模态数据集

我们使用 Kinect 采集自己的多模态数据集，数据包含了视频，音频，记录了小孩和诊疗师之间的社会交流行为。为了控制变量，数据都在 SMIIP 实验室里采集，尽量保持唯一的变量只有测试的小孩和他们的父母，环境以及诊疗师都是一样的。诊疗师和小孩之间的交互严格按照 ADOS 规定的环节。在每个环节中，诊疗师，即医生都会与小孩进行来回的对话，或者直接用肢体语言来吸引小孩的注意，从而让小孩能展现出相应的行为。这些行为反映了小孩的社交发展水平，其中肢体动作和面部表情的变化被认为是能够展现小孩早期自闭症的重要诊断因素。整个视频包括三个部分的内容，一个是叫名反应，诊疗师在小孩的视野范围内，看小孩是否有转头面向声音来源，以及反应的时间。第二个是手指指物，诊疗师出乎小孩意料地令玩具突然发出声音，并用言语以及肢体语言引导小孩看向物体并用手指指向物体。第三个是陌生人情景测试，在测试中途时，小孩的父亲或母亲突然起身离开，但是离开的时候尽量保持在小孩的余光范围内，看小孩是否有跟随父母，哭闹或者仍然只是在玩自己玩具的反应。在这篇文章中，我们专注于第二部分的内容，即判断小孩是否有看向物体并用手指指物。从视频数据中可以看到，诊疗师在引导小孩看向物体的时候，会通过发出惊讶的语气，以及同样用手指指向物体来引导小孩进行指物，这些都是外部给小孩的刺激。为了给小孩更多的刺激，我们不只布置了一个发声的玩具，而是三个，依次发声，从而令小孩更有可能做出指物的动作。

在诊疗师的角度来看，小孩有没有看向发声的物体是判断小孩患有自闭症几率的重要指标，当然同时也包括小孩有没有主动指物等。另外一部分数据为，诊疗师通过观看视频，而做出的判断。即小孩患有自闭症的几率大不大。

具体的布置如图 4-1 所示，四周由泡沫墙挡着，保证实验环境的单一性。实验中只有三个人，诊疗师，小孩的父亲或母亲。物品只有吸引小孩注意力

的玩具和限制他们行为的椅子以及桌子。

图 4-1：多模态数据集实验室布置

4.2 样本数据

在将最终的方法正式应用于多模态数据集之前，我们先在样本数据上进行应用。样本数据的实验环境与多模态数据集的实验环境完全一致。区别在于实验对象由小孩换成了实验员，这样做的目的是为了减少小孩乱跑而出了 Kinect 视野范围的因素。实验员的动作都是尽可能模仿小孩的行为，以减少实验误差。

4.3 结果

因为视频的长度太长，在验证中间步骤的时候我们使用的是 Kinect 的实时帧。最后整体的方法使用的样本数据进行验证。

4.3.1肤色模型结果

如图 4-1 所示，应用在帧上的肤色模型成功将其他的物体给屏蔽了，只剩下了手部区域和脸部区域。

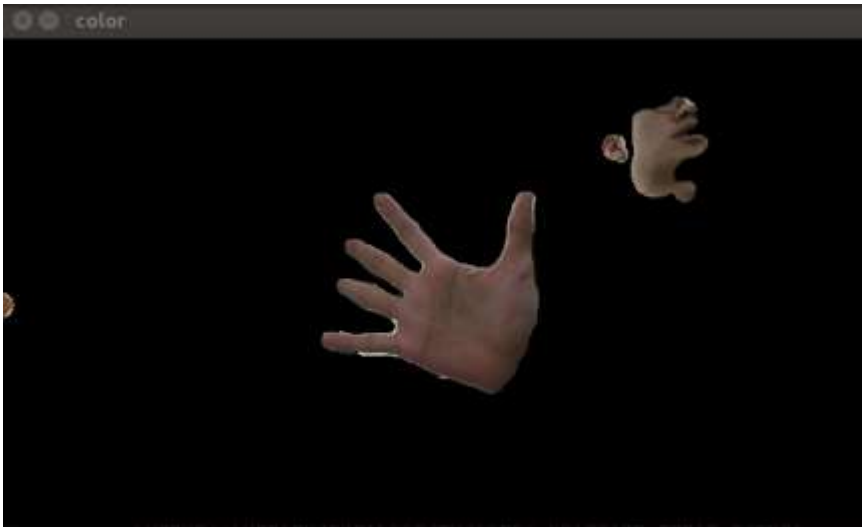

图 4-1：肤色模型结果

4.3.2形态学操作和二值化的结果

如图 4-2 所示，从左至右从上至下依次是第一次膨胀，腐蚀，模糊化，第二次膨胀，第三次膨胀，再次模糊化最后加上二值化的结果。从图中可以看到，原本帧中存在的一点一点的噪声已经没有了，物体的边缘也已经没有那么锐利，物体和物体之间的过渡变得更加圆滑。这种特点对于人眼来说可能更加不好识别，但是对于计算机来说却更容易掌握物体的特征。

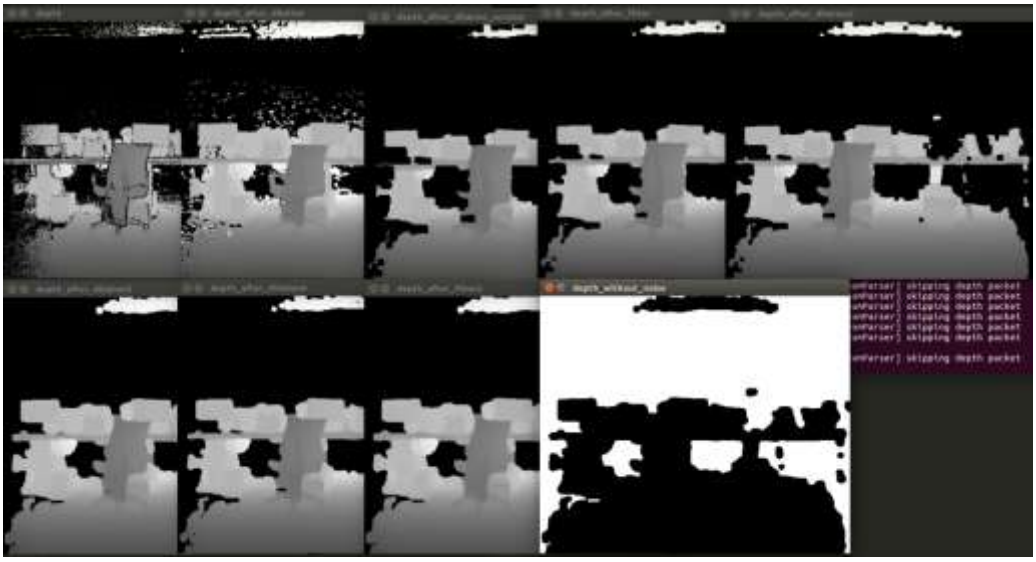

图 4-2：形态学操作结果

4.3.3轮廓检测

如图 4-3 所显示的结果，为了使结果更加明显，我们将二值图所检测到的轮廓画在了原始的色彩帧上。轮廓由白色的线表示，蓝色的点表示轮廓上的点，绿色的框表示最贴近轮廓的矩形，而红色的圆表示通过图像我们可以看出，无论人手的手势是怎样的，我们都可以正确的检测出手部的轮廓。

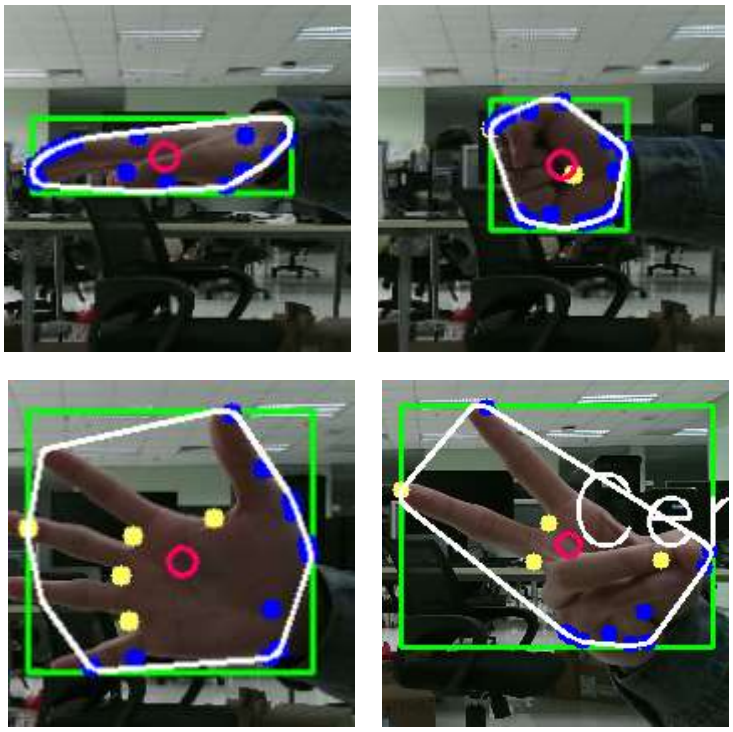

图 4-3：轮廓检测结果

4.3.4脸部检测

如图 4-4 和图 4-5 所示，在引入脸部检测前，当手部区域大到足够能被检测之前，脸部会被误当做手部，因为此时整个帧中存在的轮廓只有脸部的轮廓，手部还没能检测到。左上角的 Pointing！是当检测到的轮廓被认为是正对着 Kinect 才会显示的，因为现在人的脸部是对着 Kinect，因而左上角会显示出 Pointing！的字样。这种情况会给不理解情况的诊疗师造成困扰，因而我们需要引入脸部检测，从而引入先验条件，帮助我们做判断。在引入脸部检测后，在有了脸部区域所在位置的先验条件下，将脸部区域误认为是手部区域的情况可以被自动识别并筛除。

图 4-4：脸部检测前

图 4-5：脸部检测后

4.3.5椭圆拟合

如图 4-6 所示，在引入了脸部检测之后，当人的手指向 Kinect 时，被检测到的手部会被组成轮廓的点包围，并且会被最贴合这些点的绿色矩形所包围。左上角显示红色的 Pointing!，代表检测到的区域基本上是正对着 Kinect 的。

图 4-6：最终椭圆拟合结果

第五章 总结与展望

这篇文章主要是提出了手势识别的方法，即利用肤色模型得到人体的裸露区域。再利用形态学操作，降低帧的噪声，并将其模糊化和二值化，将其转化为更容易让计算机处理的模式。之后使用轮廓识别，将脸部和手部的所在区域识别出来并将两者分离开来。另外，为了减少当手部未到足够让计算机识别时误将脸部当做手部进行识别的情况，我们引入了脸部识别，给计算机一个脸部区域所在位置的先验知识，以便其自动筛选脸部区域。

本篇文章的创新点在于，手势识别与自闭症检测的相结合。到目前为止，还没有文章将手势识别应用于自闭症检测当中。我们的目的在于，希望在自闭症检测这种相对主观的模式中，引入客观变量，以减轻诊疗师的负担。如果这些客观的因素能全部被计算机自动检测出来，那么留给诊疗师的就只有根据电脑的分析进行综合的评估，这会大大减少诊疗师的工作。

我们未来希望，将提出的方法完整的应用在多模态数据集上，并且结合视线检测，重复行为检测，和音频数据来提取更多的客观数据，将整个自闭症诊断的模式的客观性加重，以加速诊疗过程的智能化和自动化。

参考文献

[1]: 五彩鹿自闭症研究院. 中国自闭症教育康复行业发展状况报告 II[R].北京. 2016.

[2]: Am. Psychiatr. Assoc.: Diagnostic and statistical manual of mental disorders.5th ed. Washington, DC; 2013

[3]: Chen JA, Peñagarikano O, Belgard TG, Swarup V, Geschwind DH. The emerging picture of autism spectrum disorder: genetics and pathology.Annu Rev Pathol: Mech Dis. 2015;10:111–44

[4]: Rehg J M, Abowd G D, Rozga A, et al. Decoding Children's Social Behavior[C], IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2013:3414-3421.

[5]: General Chair-Wang J Z, General Chair-Boujemaa N, Program Chair-Natsev A. Proceedings of the international conference on Multimedia information retrieval[C], International Conference on Multimedia Information Retrieval. ACM, 2010:405-408.

[6]: Jain H P, Subramanian A, Das S, et al. Real-time upper-body human pose estimation using a depth camera[C], International Conference on Computer Vision/computer Graphics Collaboration Techniques. Springer-Verlag,2011.

[7]: Han S, Choi J, Park J-I Two-hand based interaction method using a hybrid camera[C]. In: Proceedings of the of IPIU' 13.2013.

[8]: Raheja J L, Chaudhary A, Singal K. Tracking of Fingertips and Centers of Palm Using KINECT[C], Third International Conference on Computational Intelligence, Modelling & Simulation. IEEE, 2013:248-252.

[9]: Hongyong T, Youling Y. Finger Tracking and Gesture Recognition with Kinect[C], IEEE, International Conference on Computer and Information Technology. IEEE, 2012:214-218.

[10]: Choi J, Park H, Park J I. Hand shape recognition using distance transform and shape decomposition.[J]. 2011, 263(4):3605-3608.

[11]: Laptev I, Marszalek M, Schmid C, et al. Learning realistic human actions from movies[C], IEEE Computer Society Conference on Computer Vision and Pattern Recognition. DBLP, 2008:1-8.

[12]: Blank M, Gorelick L, Shechtman E, et al. Actions as Space-Time Shapes[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005, 29(12):2247-2253.

[13]: Fathi A, Farhadi A, Rehg J M. Understanding egocentric activities[C], IEEE International Conference on Computer Vision. IEEE, 2011:407-414.

[14]: Prabhakar K, Rehg J M. Categorizing Turn-Taking Interactions[M], Computer Vision – ECCV 2012. 2012:383-396.

[15]: A. Fathi, J. K. Hodgins, and J. M. Rehg. Social interactions: a fi rst-person perspective[C], IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2012.

[16]: Côté M, Payeur P, Comeau G. Comparative study of adaptive segmentation techniques for gesture analysis in unconstrained environments[J]. 2006, 46(2):28-33.

[17]: Terrillon J C, Fukamachi H, Akamatsu S, et al. Comparative Performance of Different Skin Chrominance Models and Chrominance Spaces for the Automatic Detection of Human Faces in Color Images[C], IEEE International Conference on Automatic Face and Gesture Recognition. DBLP, 2000:54-63.

[18]: Kaehler, Adrian. Learning OpenCV, [M]. O'Reilly, 2008.

[19]: ELISEO STEFANO MAINI. ENHANCED DIRECT LEAST SQUARE FITTING OF ELLIPSES[J]. International Journal of Pattern Recognition & Artificial Intelligence, 2012, 20(06):939-953.

致　谢

这次的毕业论文设计总结是在我的指导老师李明老师亲切关怀和悉心指导下完成的。从毕业设计选题到设计完成，李老师给予了我耐心指导与细心关怀，有了李老师耐心指导与细心关怀我才不会在设计的过程中迷失方向，失去前进动力。李老师有严肃的科学态度，严谨的治学精神和精益求精的工作作风，这些都是我所需要学习的，感谢李老师给予了我这样一个学习机会，谢谢!感谢与我并肩作战的舍友与同学们，感谢关心我支持我的朋友们，感谢学校领导、老师们，感谢你们给予我的帮助与关怀;感谢中山大学，特别感谢数据与计算机科学学院四年来为我提供的良好学习环境，谢谢!

毕业论文成绩评定记录

指导教师评语：   成绩评定： 指导教师签名：　　年　　月　　日
答辩小组或专业负责人意见：   成绩评定： 签名（章）：　　年　　月　　日
院系负责人意见：   成绩评定： 签名（章）：　　年　　月　　日