摘 要

研究人、计算机以及它们之间相互影响的技术称之为人机交互，是人与计算机通过人机界面进行某种形式上的信息的交流以完成一定的交互任务的过程。人机交互技术已经从以计算机为中心逐渐的转变为一人为中心，手势的交互是一种很容易学习的、自然、直观的人机交互手段，手势交互界面有着广阔的应用前景。近年来，以手机、PDA和掌上电脑为代表的手持移动设备得到了日益广泛的应用，手持移动计算机已经逐渐成为当今世界的主流计算模式之一。随着移动设备自身的软硬件性能的提高和带宽、网络覆盖等条件的改善，人机交互的效率过低和自然性不好的问题暴露得越来越明显，计算机应用的主要障碍也已由硬件技术转变为人机交互和用户界面，基于视觉的手势识别技术已经成为一个研究的重点。

本文总结和介绍了现有的手势识别技术，手势识别研究的关键内容以及手势识别技术的发展历史。接着，本文主要对手势识别涉及的主要技术进行了研究。

关键词：手势；识别；跟踪；人机；交互

摘　要

1 绪 论

1.1 课题背景

计算机系统是由人、计算机软件、计算机硬件来共同构成的人机系统。人机交互研究的是人和计算机之间相互影响的技术，是人与计算机通过人机界面进行某种形式上的交流用来完成一定交互任务的一个过程。人与硬件、软件的交叉部分就构成了人机界面。人机界面是介于用户和计算机系统之间，是计算机与人之间传递、交换信息的媒介，是人们使用计算机系统的综合操作环境。它作为计算机系统的一个重要的组成部分，是计算机科学、认知科学、心理学的交叉研究领域，也使计算机行业竞争的焦点从硬件转移到软件之后，又一个新的、重要的研究领域。随着计算机系统的发展，用户界面的发展经历了批处理、联机终端、菜单等阶段，现在正处于以图形用户界面为主流的阶段。交互式系统的发展趋势正逐渐以“以机器为中心”转移到“以人为中心”、“人际和谐交互”的方向上。而“以人为中心”的人机交互的一个重要研究方向，就是通过模拟与人类类似的感知类型进行信息传递，这些研究包括人脸识别、面部表情识别、头部运动跟踪、手势识别、以及体势识别等等[1][2]。

1.2 研究意义

手势语言是一种靠视觉和动作进行交流的一种特殊的语言，它还是一种包含信息量最多的人体语言，它与语音及书面语等自然语言的表达能力相同，因而在人机交互方面，手势完全可以作为一种有效的、自然、直接的交互手段，具有很强的表意能力，可以在很多特殊的场合表达一些特定的信息。而基于计算机视觉的手势识别技术应用于人机交互接口具有用户友好、直接而有效等等优点。这使得手势交互可以成为人机交互过程中的一个非常自然的、直观的交互通道，符合“人际和谐交互”及“以人为中心”的人机交互发展方向。

作为一种自然，直接的交互方式，基于视觉的手势交互方式有着广泛的应用前景，现有的主要领域包括：

1. 在控制机器人和机器人远程操作中的应用。例如：在伊拉克战争中使用的

智能机器人去拆炸弹，远程医疗的远程控制领域。

2. 辅助聋哑人生活。辅助聋哑人的生活，通过手语界面可以减少聋哑人在生活中的障碍。
3. 在电子游戏领域。现在更多的游戏实现了手势识别来代替以前的按键控制，实现了人机交互的综合应用。例如：游戏厅里面的跳舞系统，就是人机结合的良好平台。

1.3 论文主要的研究内容

本文主要研究了基于视觉的手势识别技术，重点研究的是手势的分割、手势的跟踪、手势特征提取和手势识别算法。

首先，本文选择 HSV 空间的 H 分量等信息对手势的原始图像进行分割，并对图像进行平滑滤波和形态学处理，得到完整的手势二值图像，通过八领域搜索法计算手势的轮廓。之后，采用 Camshift 算法对手势进行跟踪。最后，根据手势图像和轮廓的结构信息和统计信息，对手是进行识别。

2 基于视觉分析的手势图像预处理

2.1 手势识别概述

2.1.1 手势识别的定义和分类

通常我们把手势的定义分为:手势是手或者手和臂结合产生的各种姿势和动作，以助于表达情绪、想法或强调所说的话。根据不同的标准，手势还有着不同的分类：

根据手势的空间特性，手势可分为动态手势和静态手势。动态手势强调的是手在做一个动作的一个过程，表现为手在一个时间段上的手部动作的姿势的一个序列；静态手势的意思是在某一个时刻点上手在一定空间的姿势，包括朝向、手形、与身体的相对位置。

首先，对于静态手势，是通过八连通搜索法来计算手势的轮廓，并根据手势轮廓的特点，研究了手势特征提取的方法，提出了统计特征、结构特征结合的特征提取方法。对于动态手势，本文讨论了手势跟踪技术，采取了 Camshift 算法对手势来进行跟踪。在手势识别方面利用不同特征的差异，简化了识别计算，来提高系统的实时性，所以本文提出了一种分层识别的方法，这是一种根据手势特征值特性来设计的识别方法。在动态识别方面，按照自然交互的要求，设计了简单易用的分区域识别法。

所以将手势按功能做如图 2.1 所示的划分：

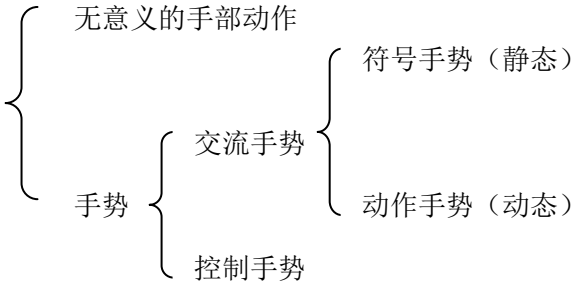


图 2.1　手势的分类

在人机交互领域中，根据手势的所表达含义可以将手势分为两类：一种是无意义的动作，一种是传递这用户意图的手势。根据功能的特点，手势又可分为控制手势和交流手势。交流手势的本质就是为了更好的传递信息，比如交谈中的伴随着不同的手势，包括具有语言描述作用的一些符号手势以及表示指示方向的手势等等。控制手势是用来控制环境中的物体，如平移，旋转，托起等等。

对于手势识别系统的分类，根据手势的采集设备可以把手势识别系统分为两类：基于视觉的手势识别系统和基于数据手套的手势识别系统。

基于数据手套的手势识别系统是最早的手势识别系统，由于当时机器视觉技术以及视觉采集设备的限制，需要用户佩戴数据手套，通过数据手套来测量出手指或者手臂的关节角度和位置等信息，进而来识别用户的手势。采用数据手套手势识别系统的优点是采集装备的应用使手势建模的难度大大降低，同时采用数据手套对手势信息的采集的有效性较高，使得基于数据手套的手势识别系统有实时性和准确性的特点。然而这种方法对于用户而言，在人机交互的过程中必须佩戴昂贵而且比较笨重的手势采集装备，是一种间接的交互方式，限制了手势交互的自由性，这就导致了基于数据手套的手势交互方法无法成为一种自然的交互方式。

随着目前计算机视觉技术的发展，基于视觉的手势识别技术也越来越成熟，它主要通过摄像机来采集手势的视觉信息，从视频图像中提取手势，并进行识别，用户不需要佩戴任何的设备，可以直接与计算机之间进行交互，与基于数据手套的手势识别技术相比，从视觉信息中要完整地恢复出原始的手势信息难度相对较大，可识别的识别率和手势数量、实时性方面还不能达到基于数据手套手势识别的效果。因为基于视觉的手势识别技术对输入设备的成本低，对用户的限制少，人手是处于一种自然状态，使人能够以自然的方式与计算机之间进行交互的优点，所以基于视觉的手势识别技术符合人机交互技术发展的方向，也是未来手势识别技术发展的趋势和目标。因此本文主要关注的是基于视觉的手势识别技术。

2.1.2 基于视觉的手势识别技术的主要研究内容

手势建模、手势分析、手势识别是基于视觉的手势识别技术的三个核心部分[3]。

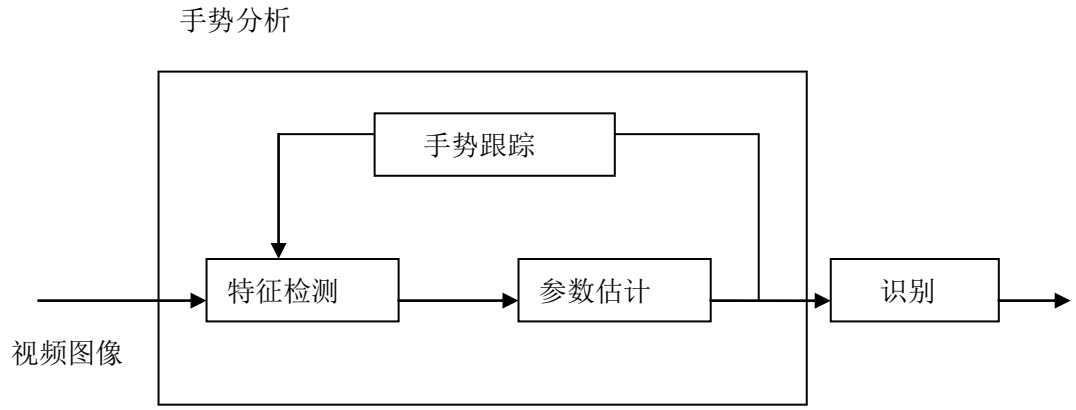


图 2.2 手势识别系统流程图

在建立了手势建模的情况下，手势识别系统的流程图如 2.2 图中所示。手势建模的意思是选取什么样的模型来描述所要表达手势，这个数学模型包含了手势的时间特征和空间属性。一旦模型确定好了之后，接下来的任务就是从单幅或序列图像中计算出模型的参数，这些参数描述了手的姿势和运算的轨迹。手的定位、跟踪以及选择合适的图像特征是手势分析阶段的关键。手势识别阶段就是将手势特征参数与已知的模型进行一一匹配，从而来判断手势的类别和属性。

1.手势建模

手势建模对于手势识别系统至关重要的，特别是对确定识别范围起关键性作用，一般来说手势建模方法被分为两大类[4]：基于表观的手势建模和基于 3D 模型的手势建模。前者是直接从观察到的视频图像去推断手势；而后者考虑了手势产生的中间媒介（手和臂）。图 2.3 是对两种建模方法的进一步分类。

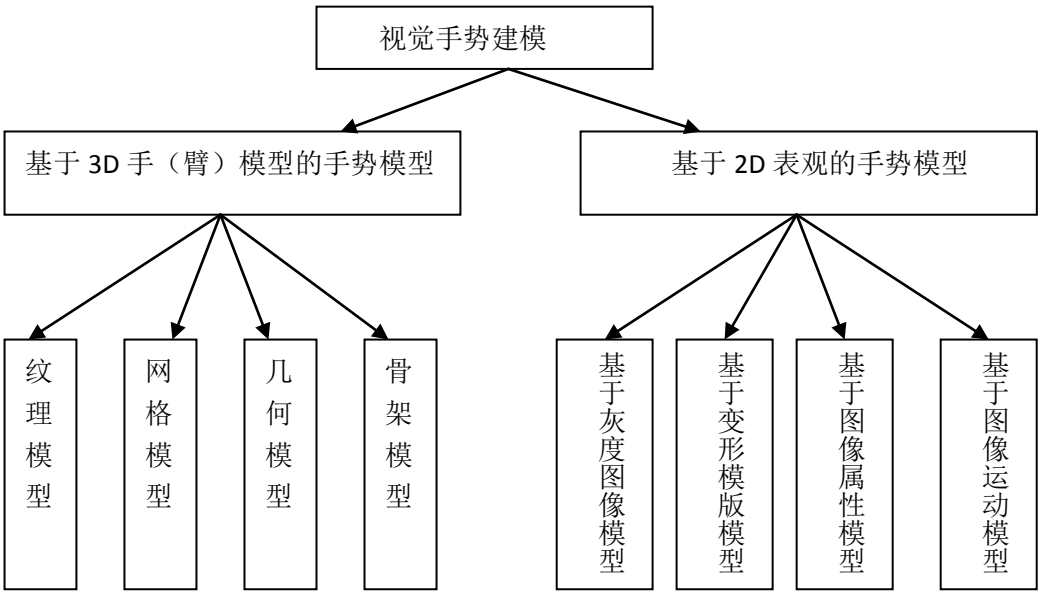


图 2.3　手势模型的分类

2.手势分析

在模型确定之后，手势分析的任务是估算所选手势模型的参数，由手势跟踪、收拾图像分割、参数和特征检测估计几个部分组成。

在进行特征检测的过程中，首先要从场景中分割出手势。手势图像分割技术有基于运动信息、基于颜色信息和基于综合信息这三种方式。基于运动信息的分割方法是利用图像差分法把用户的手部区域从背景中分离出来，常用的差分法有帧间差分法和背景差分法。基于颜色信息的手势图像分割技术利用肤色在颜色空间中的分布特性，利用阀值分割的方法将手势分割出来。第三种方法则综合利用肤色信息和运动信息来定位用户的手势。

在手势图像分割结束之后，为了提取手势的动态信息，同时避免在视频序列中每一帧的定位手势以节省计算的资源，需要在场景中跟踪用户的手势。在动态手势识别中，手部动作速度比较快，一般在 5m/s 以上，为了达到人机交互界面所需要的准确性和实时性的要求，需要一种在手部变形和复杂背景下跟踪实时性好而且鲁棒性高的图像跟踪算法。目标检测算法和运动估计算法是视频目标跟踪算法主要的两个部分。在目标检测算法方面，常用的方法有光流法、背景分差法等。运动估计算法主要包括粒子滤波算法、卡尔曼滤波算法以及在此基础上的各种改进的算法。

在结束特征检测之后，根据所选用的手势模型来估计模型参数。不同的手势

模型需要估计不同的模型参数，但是用于计算模型参数的图像特征通常都是基本相似的。常用的图像特征包括二值图像[6]、灰度图像[7]、边界[8]、区域[9]及轮廓[10]或者指尖[11]等。对于人机交互的手势识别系统，提取的手势特征首先要能够区分不同手势，另外还需要对手势的旋转、平移、缩放等变化能够很快的适应。

2.2 基于肤色的手势图像分割

手势图像分割的第一步是基于视觉的手势识别过程，是最为重要的一步，手势图像分割的好坏直接影响到后面的手势跟踪、手势特征提取以及手势识别的结果。手势图像分割就是将有意义的区域—手势从摄取的手势图像中划分出来，可以通过以下几种方法来实现：

1. 增加限制

通过简化背景，比如使用白色或者黑色的墙壁、深颜色的服装作为手势图像采集的背景。或者戴特殊的手套，通过强调前景来简化手和背景域之间的划分，加深两者之间的对比。但是这些人为附加的限制影响了手势交互的失去了自由性。

2. 模型的比较

首先建立手势在各个时刻、不同比例、不同位置下的手形图像，然后通过匹配的方法来实现手势的分割。这种方法的缺点是计算量非常大，而且无法实现实时识别。

3. 轮廓的跟踪

典型的有基于 Snake 模型的手势分割，利用 Snake 模型对噪声和对比度的敏感性来有效跟踪目标的形变和非刚体的复杂运动，达到将目标从复杂背景中分割出来的目的，这种方法的效果比较好，但同样无法用于实时系统。

4. 目标背景相减法及其改进算法

就是将目标图像和背景图像相减，此方法对消除背景图像具有很明显的效果，但要求已知背景并且背景不变，这一点限制了算法的适用范围。

5. 基于肤色的分割

主要根据肤色在颜色空间中分布的特点，通过快速的找到手可能运动的候选区域，缩小后续检测的范围。从背景图像中分割出肤色区域，用肤色特征信息来实现手势和背景的分离。基于肤色的分割方法有着直观、高效并且准确的有点，

也是本文采中手势图像分割所采用的方法。

由于颜色是人手表面最为显著的特征之一，所以在计算机视觉技术中利用颜色检测人手是一个很自然的想法，然而在不同环境下，不同的用户的肤色表现有着很大的差别，所以如果想要实现良好的肤色分割效果，那么首先需要选择一个肤色可以良好聚类并且可以适应光线环境变化的颜色空间。

手势图像预处理模块主要包括了对图像的平滑去噪、基于 H 分量的手势分割及对手势二值图像的形态滤波以及手势轮廓的提取。预处理的目的是提取手势的二值图像和轮廓，为手势特征提取提供条件。预处理的流程为图 2.4 所示：

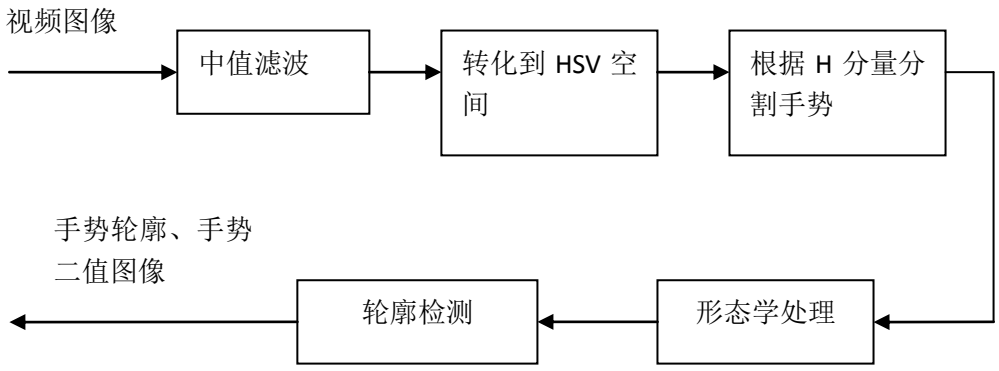


图 2.4 手势预处理模块流程图

2.2.1 肤色检测所采用的颜色空间

颜色空间指的是颜色的数学表示方法，用来指定和产生颜色，让颜色更加的形象化。颜色通常是用三维模型来表示，用三维坐标来指定空间中的颜色，在一些特定的颜色空间中，一个指定的三维坐标对应着颜色空间中的一个特定的颜色，常用的颜色空间有：RGB 颜色空间、YXZ 颜色空间、YUV 颜色空间、HSV 系列颜色空间。

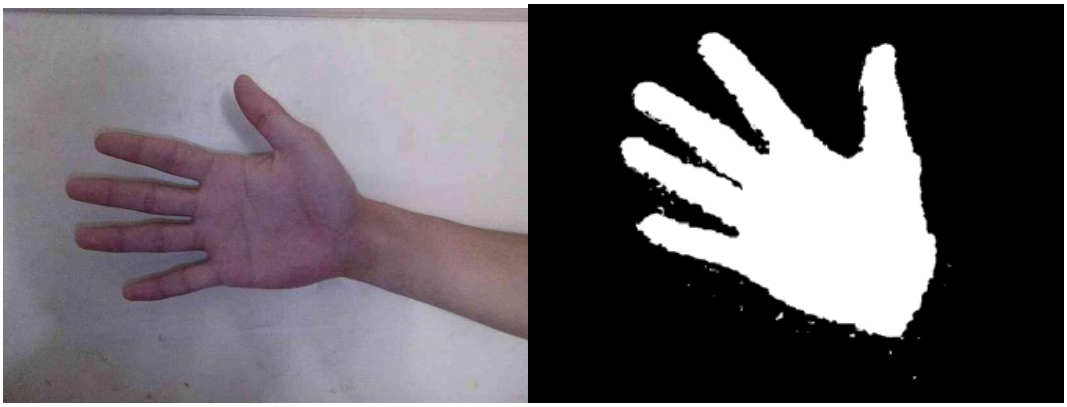
由于 HSV 空间中的 H 和 S 分量是独立于 V 分量的，也就是说颜色和饱和度信息与亮度信息是互相独立的，因此可以有效的减小受到光线的影响，所以在本文中采取 HSV 系列颜色空间来对手势进行分割，使得对肤色的分割算法可以通过更简单的方法来进行，这样就有利于增强系统的实时性。

2.2.2 HSV 空间下的手势图像分割

本文采用的是 HSV 空间中的 H 分量来进行肤色的统计。具体的步骤是对 100 幅不同光照的情况下不同的人手部的肤色图像进行统计，第一步：将 RGB 格式的图像转换成为 HSV 的格式。第二部：统计每个像素的 H 值在颜色空间分量中的取值范围内的分布范围直方图。根据对手部的肤色信息，可以对采集到的包含手部的原始的图像进行二值化处理，从而得到人手的区域。根据原图想（x，y）处的 H 值和 R、G、B 值判断二值图（x,y）处的像素值为：

$$f(x,y)=\begin{cases}255 & \mathrm{H}\in\{2,20\}\ \text{and R>G>B}\\ 0 & \text{else}\end{cases} \qquad (2,1)$$

式 2.1 中 $f(x,y)$ 是二值化图像中坐标为 (x,y) 的像素值，H、R、G、B 分别为原图像中 (x,y) 处的像素的 RGB、HSV 颜色空间所对应的分量的值。经过处理之后，就可以得到手势区域的二值图，其结果如 2.1 图所示。



(1) 简单背景下的手势图像分割

(2) 复杂背景下的手势图像分割

图 2.1 手势原图及二值化后的图像

图 2.1 中(1)为简单背景为单一颜色下的手势图像的分割，(2)为复杂背景下基于肤色的手势图像的分割。从实验的结果可以看出，基于 H 分量的肤色分割在简单背景和复杂背景下都是可以实现稳定的手势图像的分割，从而得到完整的手势二值化图像，但是图像的分割会受到视频采集设备的性能和光线等的影响，当有噪声存在的情况下，分割出来的手势图像边缘比较粗而图像内部存在空洞，解决这个问题需要后续再对图像进行滤波处理。

2.3 图像平滑

在图像的采集、传输过程中，无法避免的都会受到一些外界环境的影响，例如：噪声。那么得到的图像画质会因噪声而在不同程度上出现一些变异，因此我们必须首先对图像进行平滑操作，过滤掉部分噪声的影响。

图像的平滑是一种最常用的数字图像处理技术，主要是为了减少噪声对图像的影响。一般情况下，在空间域内可以使用领域平均来减少噪声：在频率域，由于噪声频谱通常都分布在高频段，因此可以采用各种形式的低通滤波的办法来减少噪声。常用的图像平滑方法有：中值滤波、领域平均、频域平滑技术。

根据系统视频采集设备的特点，为了需要解决采集到的原始图像中的噪声对手势图像二值化的影响，所以本文采用中值滤波技术进行平滑的处理，在消除噪声影响的同时还保留了原图想的细节。

(1)原始图像

(2) 未经平滑处理的二值化图像

(3)经过中值滤波处理后的二值化图像

图 2.2 图像中值滤波处理后的实验结果

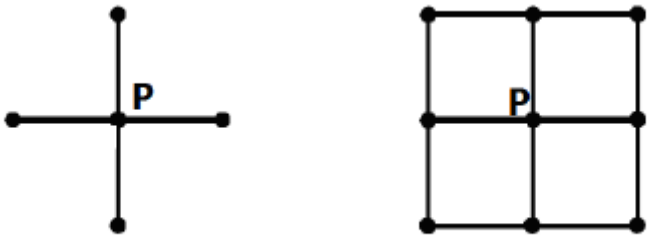
图 2.2 是本文中所采用的中值滤波对直接分割出来的二值化的图像进行处理后的结果，图中(1)为原始的图像，（2）是根据上文分割方法所得到的手势图像，(3)是用 5x5 模版对二值图像进行中值滤波处理后的结果。从上面的图像我们可以看出，在直接分割得到的二值图像中，是存在严重的噪声干扰的；经过中值滤波处理之后，就去掉了大部分的噪声的影响。不过在处理之后的图像中，虽然去除了绝大部分的噪声，手势图像的边缘还是比较粗糙的，并且还有是一些空洞的存在，为了解决这个问题，可以通过形态学滤波对图像进行进一步的处理。

2.4 形态学滤波

在处理图像二值化的过程中，本文采用形态学方法对得到的二值化图像进行滤波。因为判断肤色的标准是按照经验得到的统计信息，并不能完全准确的判断出每一个像素的性质：手部的边缘部位易受光照的影响，同时也容易受到噪声的影响，H 值不稳定，所以二值化得到的图像无法避免噪声、孔洞和粗糙的边缘的存在。

在数字图像处理的过程中，由于数学形态学算法有填充洞孔、平滑轮廓、连接断裂区域等特性，常常被用在处理各种图像操作中。数学形态学是一种以形态为基础从而对图像进行分析的一种数学工具。其基本运算分为 4 种，即膨胀、腐蚀、开运算和闭运算。通常都是用集合论名词来定义的，并用专门定义的结构元素对图像进行一定的操作。图 2.3 中为应用最为广泛的由 4 连通的 3x3 领域（5 点）和 8 连通的 3x3 领域（9 点）组成的结构元素。基于这些基本运算还可以推导和组

合各种数学形态学实用算法。



2.3 常用的结构元素

对于本实验中的二值化手势图像，存在边缘毛刺和内部的空洞，我们可以结合开运算和闭运算去除这些空洞，如图 2.4 所示的是对一幅包含噪声、内部小空洞和边缘毛刺的二值化之后的手势图像进行开闭运算的结果。其中(1)是对肤色分割得到的手势二值图像在经过中值滤波处理之后得到的图像，（2）是对(1)中图像用 5x5 结构元素进行形态学处理后的结果。

(1) 经过中值滤波得到的手势的二值图像

(2) 经过开、闭运算得到的手势二值图像

图 2.4 二值手势图像开闭运算后结果

从上图中可以看出，中值滤波处理之后的图像中还是包含了少量的孔洞和噪声，而孔洞和噪声的尺寸和手势相比是比较小的，经过开闭运算处理之后，孔洞和噪声得到了有效的减少。

2.5 边缘检测与轮廓提取

在得到了手势的二值图像之后，为了简化图像的信息，突出手势的结构特征，要对图像进行轮廓提取和边缘检测。这两种图像处理方法都是从图像中提取目标物体的形状，用提取的形状来描述特定的手势。其目的是把图像中人们感兴趣的部分分离出来，突出想要的最终目标，减少处理的信息量。两者区别在于边缘检测是一种并行的检测方法，提取图像中所有目标物体的边缘，同时检测出来的边缘有可能不是封闭的，而对于轮廓提取是一种串行的方法，提取图像中目标物体最外层的轮廓，提取出的轮廓是封闭的曲线图形。

2.5.1 边缘检测

图像的基本特征就是图像的边缘，边缘指的就是它周围像素灰度有阶变化或者屋顶变化的像素的一个集合。

Canny 边缘检测器是最为精确的一种边缘算子，在现在已经得到了广泛的应用，它按照优良检测、精确定位和对边缘单一响应这三个标准，Canny 边缘检测器

是通用性最优的一种方法。Canny 检测主要检测的是阶跃性边缘。它的基本思想就是在图像中找出具有局部最大梯度幅值的那个像素点，检测边缘的主要工作是寻找能够用于实际图像的梯度数字逼近。由于实际的图像是经过了摄像机的光学系统和电路系统固有的低通滤波器的平滑，所以，图像中的阶跃性边缘不是十分的独立。图像也容易受到噪声和场景中不希望出现的一些其他因素的干扰。

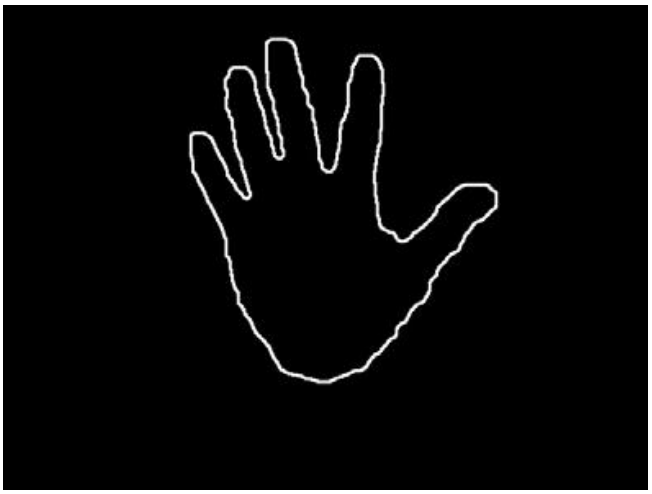
2.5.2 轮廓提取

边缘检测是一种并行的处理技术，检测出来的边缘往往不是封闭的，而轮廓提取它是一种串行的检测技术。其基本的方法是：先根据一些严格的“探测标准”找出其目标物体轮廓上的像素点，再凭这些像素点的某一些特征用一定的“跟踪标准”找出目标物体其他的像素点。

在试验中，为了方便对手势特征的提取，就采用了八领域搜索法对手势二值图像的轮廓进行提取，得到以链码形式表示的手势图像的轮廓。

手势轮廓提取效果如图 2.5 所示

(1) 手势二值图像



(2)八领域搜索法提取的手势轮廓

图 2.5 手势轮廓提取效果图

图 2.5 中，(1)是通过上面的方法所提取得到的手势二值图像，(2)图是用八领域搜索法根据手势二值图像提取的手势轮廓，并且根据此轮廓信息绘制出的手势轮廓图像。从结果中我们可以看出，通过八领域搜索法从手势二值图像中准确的提取出手势的轮廓。

2.6 本章小结

本章详细介绍了手势图像的预处理的问题，包括了对肤色空间的选择，对分割得到的二值图像的形态学滤波、手势图像的分割、对原始图像的平滑以及手势图像轮廓的提取。手势图像的预处理的目的是为了对下一步将要进行的手势分析提供一个良好的手势模型，对于手势预处理质量的好坏直接关系到手势分析的准确度，对整个识别系统的效果也是至关重要，是所有后续工作的基础。