

上海交通大学

硕士学位论文

视频图像稳定系统设计

姓名：石萍

申请学位级别：硕士

专业：生物医学工程

指导教师：朱贻盛;童善保

20070118

# 视频图像稳定系统设计<sup>\*</sup>

## 摘要

本文考虑盲人（或普通人）实际出行所带来的视频图像的不稳定特征，提出“基于特征块”的系统设计思想，将其应用到视觉假体获取的视频图像的稳像系统中。用此方法设计的系统较好地解决了人行进过程中图像序列的大幅度抖动问题。

本文首先对稳像系统的研究现状作了简要的介绍。第二章中，对课题中用到的数学模型——运动变换模型与摄像机模型给出了解释以及数学上的表示。对于稳像系统中的核心算法，论文在第三章中作了详细的研究比较。基于本课题的特点，确定采用新的设计思想——“基于特征块搜索”的思想进行稳像系统设计。论文的第五章对基于此思想的系统设计进行了详细的探讨。第六章里，用MATLAB进行系统设计和仿真，结果证明使用“基于特征块搜索”的稳像系统对于稳定人体行进过程中产生的抖动具有较好的效果。

**关键词：**图像稳定，运动估计，运动补偿，基于特征块搜索，

---

<sup>\*</sup> 973 一级子课题“视网膜视觉信号处理机制及仿生模型的研究”资助项目(2005CB724303)

# **STABILIZATION SYSTEM DESIGN OF VIDEO IMAGE**

## **ABSTRACT**

In this paper we introduce a stabilization method based on feature block in consideration of the trait in the men's walking. This method is used in the stabilization system design of video image and is resultful in stabilization.

At the beginning, the state-of-art review of the research on stabilization system is introduced briefly. Then, in chapter 2, the models of motion transformation and camera are explained and the mathematics formulae are given. Through analyzing and comparing existing algorithms each other in chapter 3, we adopt feature block searching (FBS) to design the system, which is discussed detailedly in chapter 5. Lastly, the software imitation did by MATLAB validate the feasibility of the method.

**Keywords :** Image stabilization , Motion estimation , Motion compensation , Feature block searching (FBS)

---

# 上海交通大学

## 学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。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作品成果。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结果由本人承担。

学位论文作者签名：石萍

日期：2007 年 1 月 16 日

---

## 上海交通大学 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。本人授权上海交通大学可以将本学位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、缩印或扫描等复制手段保存和汇编本学位论文。

**保密**，在\_\_\_\_年解密后适用本授权书。

本学位论文属于

**不保密**。

（请在以上方框内打“ ”）

学位论文作者签名：石萍

指导教师签名：朱贻盛

日期：2007 年 1 月 16 日

日期：2007 年 1 月 16 日

# 1. 概述

## 1.1 引言

人类视觉的目的是从眼睛得到的图像中获得关于世界的符号化描述，例如“这里有一张桌子，前面坐着一个人，这个人是李四”。在这个过程中，人的视觉系统必须要建立现实世界的模型，然后才能实现认知和推理。俗话说“眼见为实”，视觉对正常人来说是一件很简单的事情，然而，视觉系统完成的功能却是一个相当复杂的过程。人类对外界信息的获取，绝大多数是通过视觉得到的[1]。

对于视网膜疾病造成的永久性眼部病变，药物治疗方案已无能为力，目前有很多研究人员在研究采用视觉假体的方案来协助恢复部分视觉功能。如何部分或完全恢复视觉功能是目前神经工程和组织工程领域研究的热门课题。而目前通过组织工程方案修复永久性神经损伤离临床应用尚远，因此，神经工程方案中通过神经接口和神经假体方案来部分恢复神经系统功能是目前采用的主要手段。虽然人工视网膜假体的研究取得了引人瞩目的成果，但仍就有许多问题需要解决。

最近几年，各种不同的视觉假体方案发展很快[2][3][4][5]，包括视皮层刺激，视网膜上刺激，视网膜下刺激和视神经束刺激等电刺激方案，还有通过其它感觉如触觉、听觉替代视觉的方案。获取环境图像是各种假体产生功能性视觉的第一步骤。如何使获取的视频图像对应于刺激电极的有效刺激就是首先需要解决的问题。这个问题也是整个系统能够产生功能性视觉的关键问题。根据文献所述，图像处理子系统的主要功能是获取外界环境的图像，并将之转换成刺激序列。早期的电刺激视觉假体多用体外图像获取及处理系统，用摄像头获取图像转变成刺激序列。其中的部分功能可由使用者操控，如：图像缩放、亮度调节、对比度调节等。系统都具有相同或相似的图像处理算法，如：图像压缩、图像增强。另外，系统具有视网膜对图像处理的相似的基本功能，如：边缘增强、运动检测。很多科学家也都设计了许多仿生人眼的视觉机能的图像处理算法，力求恢复部分功能性视觉。如：Hallum[5]等人设计的图像处理方案主要在于减少数字化后的图像的冗余信息。他们使用互相关函数评价仿真试验的有效性，在研究中发现成六边形排列的像素获得的图像通过 Laplacian 或 Gaussian 窗函数滤波后可以得到最有

用的信息。但这些对视频图像的处理大多基于静态图像或稳定的动态图像。对于如何实时处理动态的视频图像提及甚少。而我们考虑，视觉假体的临床应用时，需要考虑到盲人在佩戴摄像系统获取的是不稳定的实时的视频图像这一前提条件非常关键。本文即是考虑盲人（或普通人）实际出行所带来的视频图像的不稳定特征，研究如何送给刺激器有效的稳定图像。

## 1.2 稳像研究现状

从最早的机械式稳像、光学稳像到电子稳像，稳像技术的研究已开展了多年。稳像技术向着更精确、更灵活、体积小以及价格低、能耗小、易于操作的方向发展。

1. 机械式稳像: 在一些大型的武器系统中，往往将整个仪器置于稳定平台上。而在一些小型导弹导引头或电视摄像头中，是将整个光学系统稳定起来。在这些系统中使用机电的方法来达到稳像目的，因此称为机械式稳像。这种系统像面是运动的，为后续图像处理系统增加了困难。
2. 光学稳像: 在一些光学仪器中用光学的方法实现图像稳定，称为光学稳像。在这里，一般是在光路中设置一个光学元件作为对不稳定图像的补偿。光学稳像方法不需将整个仪器或光学系统稳定，因此减轻了被稳定物的重量，同时物镜和像面均是固定的，便于对图像进一步处理，目前的稳像光学仪器广泛采用的是该方法。
3. 电子稳像：随着新的传感技术、计算机技术的发展，稳像系统已经逐步成为综合性系统。电子稳像实际上是应用图像处理的方法来直接确定图像偏移量并进行补偿的技术。

下面的两个例子分别是“机械式手动补偿法”和“电子式手抖动校正法”在现代电子技术中应用[7]。

### 1. 机械式手动补偿法

松下公司最早开发了用于补偿摄录机手抖动影响的机械——电子稳像器，应用于10倍变焦镜头的S-VHS摄录机。这种机械的原理是，变焦镜头和CCD组合的摄像单元，可以纵向和横向自由旋转，用特殊机构支承。为了使摄像单元可以纵向

和横向旋转,在各个旋转轴安装了线圈磁铁,装配了执行部件。为了检出各个旋转方向的抖动程度,在摄像单元设置了两个角度传感器。当开始摄像时,首先由两个角度传感器检出摄录机的纵向和横向抖动。接着由传感器输出信号供给伺服电路,分别控制旋转执行部件,使摄像单元朝与画面抖动相反的方向旋转,从而使拍摄到画面无抖动的稳定图像。但是,当防止画面抖动的功能过强时,手抖动和摄像机摇晃与摄像意图需要的摇镜头和摄像机倾斜等动作的判别,变得非常困难。因此,增加了用微机判别静止摄像、步行摄像、摇镜头和摄像倾斜等功能,即使在摇镜头和倾斜时,也能顺利摄像。

## 2. 电子式手抖动校正法

为了取代松下的机械式装置,日立公布了补偿摄录机手抖动影响非常有效的电子装置。过去,松下的电子式手抖动补偿方法,是将CCD输出的抖动视频信号,存储在视频场存储器,接着改变存储器地址,再重新读出视频信号,因而消除了图像的抖动。在这里,由于使用场存储器,因而整个装置变大了。日立公司开发的电子式手抖动影响补偿装置,是在补偿因手抖动使CCD产生摇晃电子图像后,直接取出扫描稳定的图像。它的工作原理如下:首先,从视频信号检出图像抖动,按照将其抖动消除来控制CCD的扫描范围,从而得到稳定的图像。其具体过程是,将来自CCD输出信号的像素部分(画面)称为扫描像素,首先,将该扫描像素的一半当作剩余像素,配置在这个扫描像素的周围,这个扫描像素的位置随控制信号上下左右自由移动。当这种CCD的扫描像素数约30万时,由于垂直方向有485像素,水平方向有632像素,在一个像素间距(相邻两个像素之间的距离)内的图像抖动方向能非常细致地检出来。因此,为了消除在一个像素间距内检出的像素抖动,扫描像素的位置在一个像素间距内上下左右微小移动,这种扫描像素位置的移动,由于是高速扫描剩余像素范围,在一场周期内,能够自由自在地控制。

电子式抖动校正装置与机械式装置对比,主要优点可以归纳为以下3条:(1)可以均匀补偿手抖动的影响;(2)由于是电子式的,故能补偿手轻微抖动的影响;(3)由于不使用场存储器,整个装置小型化了,可以安装在小型摄录机内,这是电子式手抖动影响校正器的最大优点。

下表就各种稳像系统性能进行了比较比较。



	稳像手段	稳像精度	控制方式	维护	功耗	体积	成本	适用场合
机械式稳像	机械以光电辅助	中	复杂	复杂	高	大	高	架设在大型武器装备或具有大平台空间的系统
光学稳像	光学	高	复杂	复杂	无	中	中	小型导引头电视摄像头
电子稳像	以电子为主，光、机为辅	高	简单	方便	低	小，且可集成程度高	低	任何场合，特别是用于对体积要求较高的小型导引头、摄像头

从上表中看出，机械式稳像受到控制方式、维护方式、成本等因素影响，发展受到极大的制约，且由于其体积庞大、功耗很高仅能架设在对体积要求不高的大型武器系统上，发展前景不甚乐观；光学系统的功耗几乎为零、体积较小且稳像精度很高，但是由于光学设备昂贵、使用寿命短、维护复杂等特点决定了光学系统使用范围较窄。与传统的光学稳像、机电结合的稳像方法相比，电子稳像具有易于操作、更精确、更灵活、体积小以及价格低、功耗小等特点，同时由于大规模集成电路技术的不断提高，又便于实现设备的小型化。可见电子稳像方法将是观瞄或摄像系统中的更新换代的方法。

电子稳像的前提条件是先获得图像的全局运动矢量，根据获取图像运动矢量方法的不同，电子稳像的基本方法有两种：

1. 利用传感器检测摄像机的运动矢量再转化为图像的运动量。即在摄像机上设置两个角速率传感器用于检测上下、左右方向上的角速率，通过放大、滤波取出对图像质量影响不大的高频成份，再经过 A/D 转换存入计算机中作为控制依据。然后采用像移补偿技术补偿图像运动，实现图像稳定。此方法易于实现，但其检测精度要依赖于具有高精度的速率陀螺来准确地敏感摄像机的振动。
2. 利用稳像算法实现电子稳像。图像获取后，直接从像面上来确定像偏移，并进行补偿。这时被摄对象和摄像系统之间的相互运动向量取决于图像偏移测量系统，图像偏移测量系统可以是这样或那样的算法，该算法就是对图像进行处理来分离出运动向量。与传统的光学稳像、机电结合的稳像方法相比，

电子稳像具有易于操作,更精确,更灵活,体积小以及价格低,能耗小等特点,同时由于大规模集成电路技术的不断提高,又便于实现设备的小型化。可见电子稳像方法将是观瞄或摄像系统中的更新换代的稳像方法。

有关稳像技术的相关信息可以参考综述性文献[8]。

### 1.3 目前视觉假体视频图像处理研究现状

视网膜电刺激系统的基本框架是通过植入电极(阵列),将CCD摄像头获取的图像转变成对应于这些电极阵列上的电刺激脉冲序列,通过电流刺激的方式来部分恢复视觉功能。系统一般包括三个部分,即植入眼内的电极和刺激器,眼外的图像采集系统以及图像处理识别,序列产生的软件部分[9][10]。

视觉假体系统中,图像处理是系统的关键部分。图像处理系统将由摄像机拍摄的图像转换为刺激图样。图像处理系统需要具有较好的实时性、强大的处理功能和便携性,以便刺激能随外界的变化及时更新并方便用户使用。由于商用摄像机至少是 $320 \times 240$ 像素,电极的多少对应于图样分辨率的高低,而刺激电极无法和每个像素一一对应,故图像处理系统算法主要包括压缩和增强。Liu[11]的系统中,使用的是小型CMOS摄像机和大规模集成电路芯片进行图像采集和处理。

但目前国内外的有关视觉假体设计的文献中,大都未提及有关视频图像采集后的稳像问题。

我们考虑到盲人(或普通人)在行走,头部转动,无意识的身体晃动等都可能造成采集的图像整体画面不稳定、颠簸的情况,使得进一步的图像处理难度增大,为产生合适的刺激组型带来困难;我们希望通过数字图像处理技术改善图像序列中的稳定状况,消除因摄像机的无意抖动而造成的视频变化,以保证摄像机在一定的变化范围内,输出的视频是稳定的。

对于图像视频稳定的技术研究来说,大多数科学工作者都是基于数码产品或航空以及军事等领域的研究,在这些领域的研究中,力求最优的图像稳定算法,得到最好的效果,并且具有强大运算功能的芯片也为实现这些要求提供了可能性。

视觉假体系统的特殊性使得我们在进行视频图像稳定的实现中具有不同的出发点:

- 1) 视觉假体实时性要求较高，所以要求简单而快速的算法
- 2) 视觉假体芯片的开发中较高的成本不利于假体的临床推广，需要较为简单的硬件实现；
- 3) 稳定后的视频图像还会经过诸如图像压缩，图像增强，轮廓提取等一系列的图像处理过程，所有的图像预处理过程都是符合视觉感知机制进行的，所以视频图像的稳定结果需要达到视觉感知稳定的效果。

图像序列的帧间运动有两种：一是由于摄像机参数或者位置变化而引起的整个图像的变化称为全局运动；二是由于拍摄对象的运动而引起的局部图像变化称为局部运动。而全局运动又分为有意和无意，比如扫描景物获得的图像就是有意的全局运动，而外界干扰造成的图像抖动就是无意的全局运动。我们所要做的是对无意的全局运动做出估计和补偿。

## 1.4 论文的主要研究内容及创新点

### 1.4.1 主要研究内容

论文的具体安排如下：

第一章是概述。简要介绍了稳像研究现状以及目前视觉假体视频图像处理研究现状和论文的主要研究内容及创新点

第二章里，讨论了运动变换模型和摄像机模型。

在第三章里，介绍了现有的主流的视频稳像算法。包括光流法、基于空间像素点灰度的方法、位平面匹配法、基于空间特征的方法、基于图像块的方法。在讨论这些文算算法的同时比较了各算法之间的优缺点。

第四章是文章的研究重点：基于特征块的稳像算法系统的设计。此部分时在盲人视觉假体视觉感知机制的基础上，力求设计出快速处理，实时有效的、适合视觉假体上视频图像系统的设计。

第五章是系统设计的Matlab仿真。

第六章对论文进行了总结，并对视觉假体视频图像系统设计临床应用进行了展望。

### 1.4.2 创新点

以往的研究多基于由于非人为因素（风，震动）所产生的摄像系统摇摆或晃动，其对图像的处理方法主要考虑如何恢复图像的清晰度。本研究具有相似之处，但最大的不同在于消除由于走路时身体的摆动和上下平移而造成的目标图像平移。由于视觉假体相对于摄像系统地分辨率来说小得多，且我们要假定盲人需要“足够”的时间来辨认刺激，所以进入视觉假体刺激器的图像，需要当前目标物体及其相对背景图像具有坐标系位置不变特性，才能产生有效的刺激序列。

刺激所需的快速稳像算法中的运动矢量检测采取绝对误差和准则（SAD），能够快速准确实现本研究中要求的稳像目标。系统设计中加入匹配模板，用来定位图像中的基准目标，通过模板的稳定进行整个图像的稳定实现，系统设计中采用基于特征模块匹配的图像稳定方法，达到了对“特征目标”的定位。

## 2 运动变换模型与摄像机模型

### 2.1 成像理论与数字图像的数学表示方法

#### 2.1.1 阿贝成像理论

1873年,阿贝(E.Abbe,18 40--1905)首先提出,“二次成像”假说。今天,我们计算机图像处理的许多思想,都可以从那里借鉴,它是近代光信息处理的基础。

阿贝二次成像理论认为,物体经过光学系统到像经历了两个过程:物体 $L$ 经过光学系统后,在它的后焦面上形成衍射图样。以衍射图样为次波源,在像平面上产生振幅叠加而构成图像 $L'$ 。由此推出两点重要结论:

由于透镜的孔径的有限性,有些光总是被挡住,因而丢失了物的部分信息,那部分光为高频分量,也就是说像总是丢掉物体的高频信息。物体的细节部分必然造成大的 $\theta$ 角衍射,这部分信息因光学系统孔径受限而丢失,说明分辨率有极限,再则亦说明物和像总是有差异的;

若在系统的后焦面上换上不同形式的光阑,即部分地挡去第一过程的衍射图样,像也将随着变化。

这样根据阿贝理论,将图像通过傅立叶变换到频域,在频域上采取不同形式的约束,可以改善和提高图像的质量。如“图像边缘增强”,“目标识别”等技术,应用与军事,工农业生产等领域。从而发展了光信息处理科学,也成为计算机图像处理学科的物理基础。

#### 2.1.2 数字图像的表达

图像记录的是物体辐射能量的空间分布,它是关于空间点坐标 $(x, y)$ 及光波长 $\lambda$ 的函数。对于活动的图像,它是一个四维的函数,可表示成 $F(x, y, t, \lambda)$ ,对于多光谱图像(如:彩色图像和遥感图像)观察到的像场可模拟为图像函数对光谱响应 $S(\lambda)$ 的加权积分,因而对第 $i$ 个光谱像来说图像函数仅表示成空间坐标 $(x, y)$ 与时间 $t$ 的函数:

$$F(x, y, t) = \int_0^\infty F(x, y, t, \lambda) S_i(\lambda) d\lambda \quad (2.1)$$

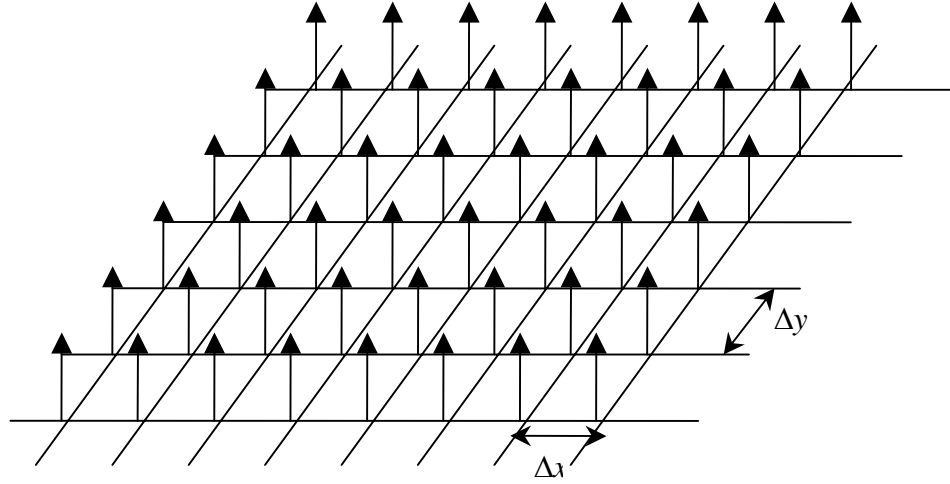


图 2.1：连续函数  $f(x, y)$  的抽样

对单色图像来说， $F(x, y, t)$  就表示与空间坐标及时间有关的亮度。如果认为一幅图像是不随时间变化而变化的，时间变量  $t$  可以从图像函数中去除，这样图像函数从四维变成二维，即仅表示成空间坐标点与亮度之间的关系  $F(x, y)$ 。经过采集和量化，连续像场离散化。因此可以用矩阵表示一幅图像，如果  $M \times N$  矩阵  $F$  是由抽样的矩形区域所限制，则图像函数可表示成：

$$F = \begin{bmatrix} F(1,1) & F(1,2) & \Lambda & F(1,N) \\ F(2,1) & F(2,2) & \Lambda & F(2,N) \\ \Lambda & \Lambda & O & \Lambda \\ F(M,1) & F(M,2) & \Lambda & F(M,N) \end{bmatrix} \quad (2.2)$$

## 2.2 图像坐标系、摄像机坐标系、世界坐标系

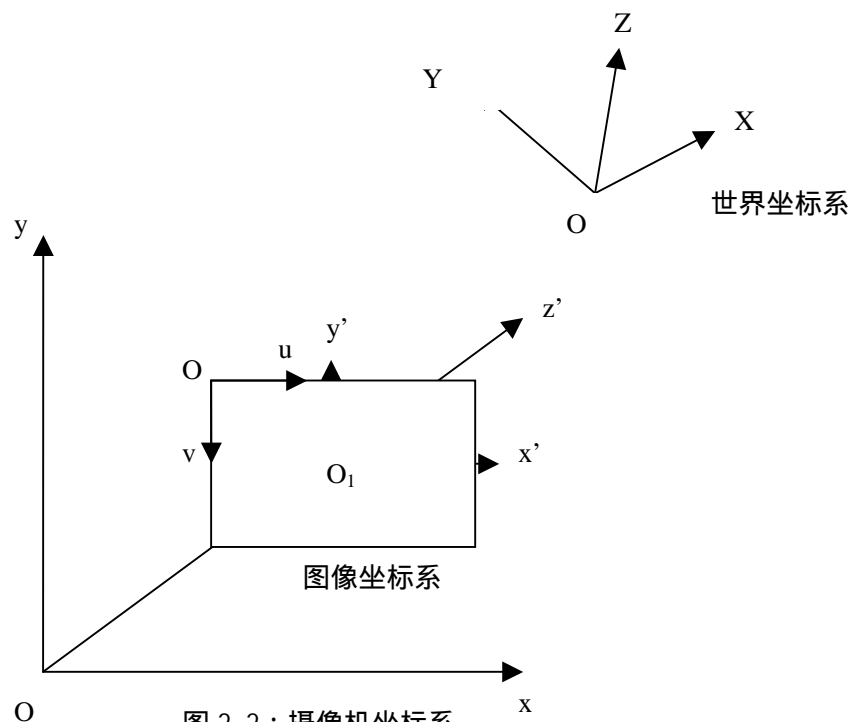


图 2.2：摄像机坐标系

3-D 空间景物成像时涉及到的坐标系统有：

**世界坐标系:** 也称真实或现实世界坐标系统，它是客观世界的绝对坐标 (所以也称客观坐标系统)。一般的3-D场景用这个坐标系统来表示。如图2.2中坐标系  $O-XYZ$  所示。

**摄像机坐标系:** 以摄像机为中心制定的坐标系统  $xyz$ 。一般常取摄像机的光学轴为  $z$  轴。摄像机坐标系如图2.2中的坐标系  $O-xyz$  所示，其中坐标原点  $O$  又称为摄像机光心， $x$  轴和  $y$  轴与成像平面坐标系的  $x'$  轴和  $y'$  轴平行， $z$  轴与成像平面垂直，称为摄像机的光轴 (或称为主轴)。光轴与成像平面的交点为图像平面的主点  $O_1$ ， $OO_1$  为摄像机焦距。摄像机坐标系和世界坐标系之间的关系可用旋转矩阵  $R$  与平移向量  $t$  来描述。

**像平面坐标系:** 由于图像坐标系只表示像素位于数字图像中的列数和行数，并没有用物理单位表示出该像素在图像中的物理位置，因而需要再建立以物理单位 (例如毫米) 表示的坐标系，如图2.2中所示的以点  $O_1$  为坐标原点的坐标系  $O_1-x'y'$  称为成像平面坐标系。

计算机图像坐标系统: 在计算机内部数字图像所用的坐标系统MN。数字图像最终由计算机内的存储器存放, 所以要将像平面的投影坐标转换到计算机图像坐标系统中。

摄像机采集的数字图像在计算机内是以数组的形式存储的, 数组中的每一个元素(称为像素, pixel)的值是图像点的亮度(或称灰度, 若为彩色图像, 则图像的像素亮度将由红、绿、蓝三种颜色的强度表示)。如图2.2所示, 在图像平面上以左上角o为坐标原点建立的直角坐标系  $O-uv$  称为图像坐标系, 每一像素的坐标  $(u, v)$  分别是该像素在数组中的列数和行数, 即  $(u, v)$  是以像素为单位的图像坐标系坐标。

## 2.3 图像平移

图像平移就是将图像中所有的点都按照指定的平移量水平、垂直移动。设  $(x_0, y_0)$  为原图像上的一点, 图像水平平移量为  $t_x$ , 垂直平移量为  $t_y$ , 则平移后点  $(x_0, y_0)$  的坐标将变为  $(x_1, y_1)$ 。 $(x_0, y_0)$  和  $(x_1, y_1)$  的关系为:

$$\begin{cases} x_1 = x_0 + t_x \\ y_1 = y_0 + t_y \end{cases} \quad (2.3)$$

用矩阵表示为:

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix}, \quad (2.4)$$

对矩阵求逆, 可以得到逆变换:

$$\begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -t_x \\ 0 & 1 & -t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \text{ 即 } \begin{cases} x_0 = x_1 - t_x \\ y_0 = y_1 - t_y \end{cases}. \quad (2.5)$$

这样, 平移后的图像上的每一点都可以在原图像上找到对应的点。

## 2.4 图像旋转

一般图像的旋转是以图像的中心为原点, 旋转一定的角度。旋转后, 图像的



大小一般会改变。和图像平移一样，我们既可以把转出区域的图像裁去，也可以扩大图像以显示所有的图像。

下面推导一下旋转运算的变换公式。点  $(x_0, y_0)$  经过旋转  $\theta^0$  后坐标变成  $(x_1, y_1)$ 。在旋转前：

$$\begin{cases} x_0 = \gamma \cos(\alpha) \\ y_0 = \gamma \sin(\alpha) \end{cases} \quad (2.6)$$

旋转后：

$$\begin{cases} x_1 = \gamma \cos(\alpha - \theta) = \gamma \cos(\alpha) \cos(\theta) + \gamma \sin(\alpha) \sin(\theta) = x_0 \cos(\theta) + y_0 \sin(\theta) \\ y_1 = \gamma \sin(\alpha - \theta) = \gamma \sin(\alpha) \cos(\theta) - \gamma \cos(\alpha) \sin(\theta) = -x_0 \sin(\theta) + y_0 \cos(\theta) \end{cases} \quad (2.7)$$

写成矩阵表达式为：

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} \quad (2.8)$$

其逆运算如下：

$$\begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \quad (2.9)$$

上述旋转是绕坐标原点  $(0, 0)$  进行的，如果是绕一个指定点  $(a, b)$  旋转，则先将坐标系平移到该点，再进行旋转，然后平移回新的坐标原点。将坐标系平移到坐标系 处，其中坐标系 的原点在坐标系 中坐标为  $(a, b)$ 。两种坐标系坐标变换矩阵表达式为：

$$\begin{bmatrix} x_{II} \\ y_{II} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -a \\ 0 & -1 & b \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_I \\ y_I \\ 1 \end{bmatrix} \quad (2.10)$$

其逆变换矩阵表达式为：

$$\begin{bmatrix} x_I \\ y_I \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & a \\ 0 & -1 & b \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{II} \\ y_{II} \\ 1 \end{bmatrix} \quad (2.11)$$

假设图像未旋转时中心坐标为  $(a, b)$ ，旋转后中心坐标为  $(c, d)$ （在新的坐标系下，

以旋转后新图像左上角围原点), 则旋转变换矩阵表达式为:

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & c \\ 0 & -1 & d \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -a \\ 0 & -1 & b \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} \quad (2.12)$$

其逆变换矩阵表达式为:

$$\begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & a \\ 0 & -1 & b \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -c \\ 0 & -1 & d \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \quad (2.13)$$

即:

$$\begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta & -c \cos \theta - d \sin \theta + a \\ -\sin \theta & \cos \theta & c \sin \theta - d \cos \theta + b \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \quad (2.14)$$

因此,

$$\begin{cases} x_0 = x_1 \cos \theta + y_1 \sin \theta - c \cos \theta - d \sin \theta + a \\ y_0 = -x_1 \sin \theta + y_1 \cos \theta + c \sin \theta - d \cos \theta + b \end{cases} \quad (2.15)$$

## 2.5 摄像机模型

一般说来,所有成像系统都是将三维场景变换成二维灰度或彩色图这种变换可以用一个从三维空间到二维空间的映射来表示:

$$\begin{aligned} f: R^3 &\rightarrow R^2 \\ (x, y, z) &\rightarrow (x', y') \end{aligned} \quad (2.16)$$

如果考虑时变三维场景,侧上述变换是四维空间到三维空间的变换,如果再考虑某一波段或某几个波段的光谱,则上式的维数将增加到五维或更高维。这里我们只考虑三维空间到二维空间的两种常用映射:透视投影变换和正交投影变换。

### 2.5.1 透视投影

透视投影(Perspective Projection)是最常用的成像模型,可以用针孔(Pinhole)成像模型来近似表示。透视投影成像模型的特点是所有来自场景的光线均通过一个投影中心,它对应于透镜的中心,经过投影中心且垂直于图像平面(成像平面)

的直线称为投影轴或光轴，如图 2.3 所示。其中  $oxyz$  是固定在摄像机上的直角坐标系，遵循右手法则，其原点  $o$  位于投影中心， $z$  轴与投影轴重合并指向场景， $x$  轴、 $y$  轴与图像平面的坐标轴  $x'$  和  $y'$  平行， $xy$  平面与图像平面的距离为摄像机的焦距  $f$ 。在实际摄像机中，图像平面位于投影中心后面距离为  $f$  的位置(图 2.3)，其投影图像是倒立的。为了避免这种倒立图像，假定图像平面位于投影中心的前面，如图 2.3 所示。场景中一点  $(x, y, z)$  在图像平面上的投影位置  $(x', y')$  可以通过计算点  $(x, y, z)$  的视线(空间点与投影中心的连线)与图像平面的相交位置得到。

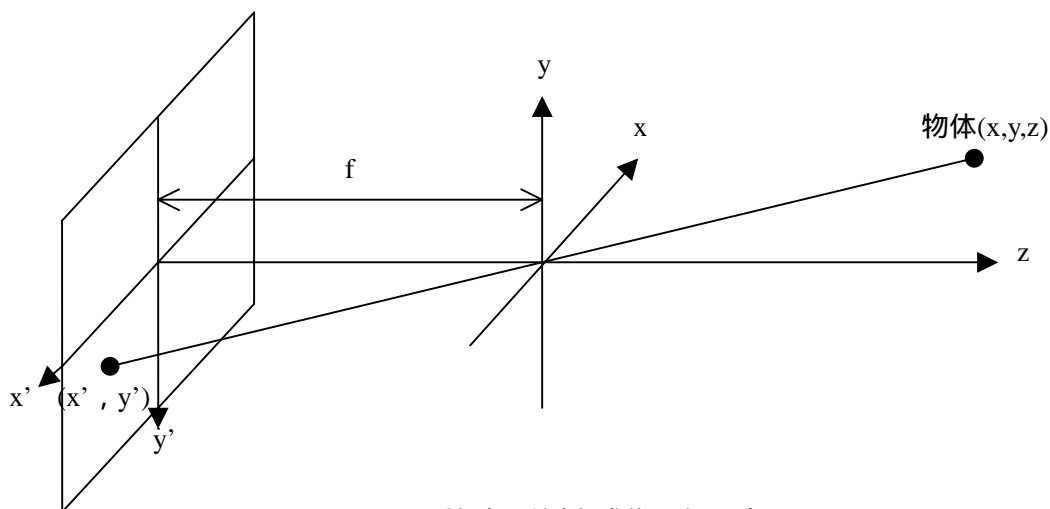


图 2.3：透视投影倒立成像几何示意图

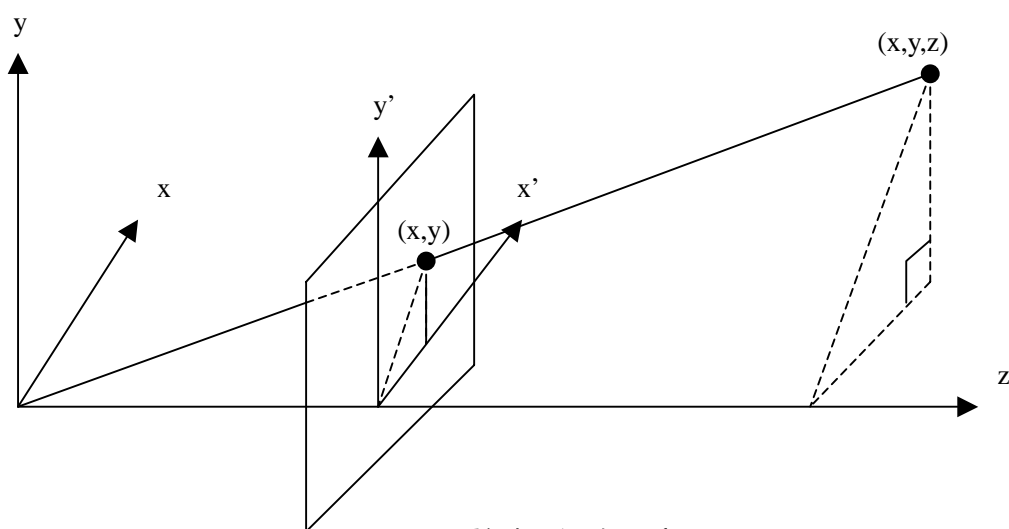


图 2.4：透视投影几何示意图

点 $(x, y, z)$ 的视线、点 $(x, y, z)$ 到 $z$ 轴的垂直线段以及 $z$ 轴形成了一个三角形。图像平面上点 $(x', y')$ 的视线、点 $(x', y')$ 到轴垂直线段以及 $z$ 轴形成了另一个三角形。这两个三角形是相似三角形，因此得到透视投影方程：

$$\frac{x'}{x} = \frac{y'}{y} = \frac{f}{z} \quad (2.17)$$

点 $(x, y, z)$ 在图像平面中的位置由下式给出：

$$x' = \frac{f}{z} x \quad y' = \frac{f}{z} y \quad (2.18)$$

## 2.5.2 正交投影

正交投影 (Orthogonal Projection)指用平行于光轴的光将场景投射到图像平面上，因此也称为平行投影(Parallel Projection)，如图 2.5 所示。正交投影是透视投影的一个特例，当透视投影模型的焦距 $f$ 很大且物体距投影中心很远时，透视投影就可以用正交投影来近似。此时，投影方程为：

$$x' = x \quad y' = y \quad (2.19)$$

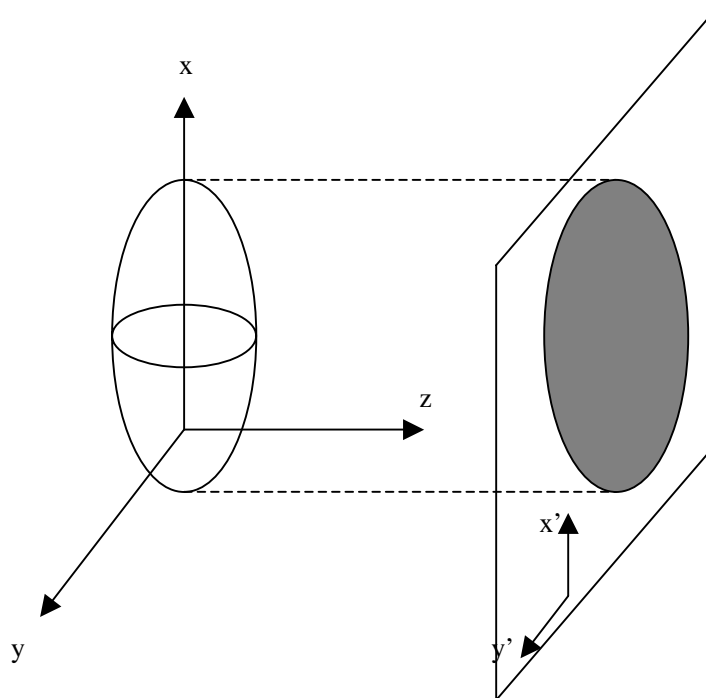


图 2.5：正交投影示意图

### 2.5.3 摄像机模型

摄像机模型即是成像几何模型。它是由前述各个坐标系统不同的相互关联而得到的，实质上就是几何透视变换。

通用摄像机模型不但要考虑各个坐标系统的不重合，还要有两个因素要考虑：一是摄像机镜头会有失真，所以在像平面上的成像位置会与用前述公式算出的透视投影结果有偏移；二是计算机中使用的图像坐标单位是存储器中离散像素的个数，所以对像平面上的坐标需取整变换(假设像平面仍用连续坐标)。

我们只考虑世界坐标与摄像机坐标重合时的摄像机模型。它所涉及的变换即为上面所述的透视投影变换。

摄像机坐标系统的 $xy$ 平面与像平面重合而光学轴沿 $z$ 轴，这样像平面的中心处于原点，镜头的中心坐标是 $(0,0,f)$ 其中 $f$ 为焦距。推出：

$$x = \frac{fX}{f - Z} \quad y = \frac{fY}{f - Z} \quad (2.20)$$

两式关于 $x$ 和 $y$ 的等式都是非线性的，可以借助齐次坐标的方法将它们表示成线性矩阵的形式。

一个点对应笛卡尔坐标系 $XYZ$ 的齐次坐标定义为 $(kX, kY, kZ, k)$ ，其中 $k$ 是一个任意的非零常数。一个笛卡尔世界坐标系统中的点可用矢量形式表示为：

$$W = [Z, Y, X]^T \quad (2.21)$$

它对应的齐次坐标是：

$$W_h = [kX, kY, kZ, k]^T \quad (2.22)$$

定义透视变换矩阵：

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{f} & 1 \end{bmatrix} \quad (2.23)$$

那么 $PW_h$ 给出的元素是齐次形式的摄像机坐标矢量：

$$C_h = PW_h = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{f} & 1 \end{bmatrix} \begin{bmatrix} kX \\ kY \\ kZ \\ k \end{bmatrix} = \begin{bmatrix} kX \\ kY \\ kZ \\ -\frac{kZ}{f} + k \end{bmatrix} \quad (2.24)$$

故摄像机坐标系统中任一点的笛卡尔坐标可表示为:

$$C = [X, Y, Z]^T = \left[ \frac{fX}{X - Z}, \frac{fY}{f - Z}, \frac{fZ}{f - Z} \right]^T \quad (2.25)$$

因为 $xy$ 和 $x'y'$ 两平面重合,所以 $C$ 的前两项也表示3-D点 $W(X, Y, Z)$ 在像平面上的投影坐标。

## 2.6 本章小结

我们所要处理的图像是一幅通过光学系统后所记录的像(而不是计算机图形学中采用函数关系构成的图形)。对这样的一幅图像要做计算机处理的第一步必须要建立一个数学模型。数学模型的建立都是对物理过程的一种抽象。本章从成像理论与数字图像的数学表示方法开始,建立图像坐标系、摄像机坐标系、世界坐标系,在图像旋转和图像平移的基础上,建立了摄像机模型的数学表示。本章所介绍的运动变换模型与摄像机模型是进行稳像系统设计的基础。

## 3 视频稳像算法研究

### 3.1 假设模型

稳像算法是有关运动估算的算法。由于该问题具有“不适定”的性质，稳像算法需要有关二维运动场结构的附加假设模型。

**参数模型：**参数模型时为了描述曲面的三维运动（位移和速度）在图像平面上的正交或透视投影。一般来说，三位曲面的表达式决定了带参数的二维运动场模型。例如，一个由平面的三维刚体运动产生的二维运动场，在正交投影下，可以用6个参数仿射模型描述；在透视投影下，可用8个参数非线性模型描述。对于二次曲面还存在更复杂的模型。

**非参数模型：**参数模型的主要缺点是它只适用于三维刚体运动。换言之，在不适用三维刚体运动模型下，可将非参数均匀性（平滑度）约束条件强加于二维运动场上。非参数约束条件可归纳为确定性和随机性的平滑度模型。下面介绍结构主要的非参数方法。

### 3.2 稳像算法

#### 3.2.1 光流法

光流(optical flow)法的概念是由Gibson于1950年首先提出的，光流(optical flow, image flow)是空间运动物体的被观测表面上的像素点运动的瞬时速度场，包含了物体与成像传感器系统之间相对运动的关系。当物体在运动时，它在图像上对应点的亮度模式也在运动。这种图像亮度模式的表观运动(apparent motion)就是光流。光流研究是计算机运动视觉的一个重要部分，是利用运动图像序列中的强度数据的时域变化和相关性，确定图像像素位置的运动情况。光流表达了图像的变化，由于它包含了目标运动的信息，因此被观察者可用来确定目标的运动情况[12][13]。由光流的定义可以引申出光流场，它是指图像中所有像素点构成的一种二维(2D)瞬时速度场，其中的二维速度矢量是景物中可见点的三维速度

矢量在成像表面的投影。所以光流不仅包含了被观察物体的运动信息，而且还包含有关景物三维结构的丰富信息。光流运动分析涉及3D刚体运动在2D图象平面的透视投影。

目前提出许多不同的光流计算方法，新的方法还在不断地出现。归纳起来，光流计算方法可以分为以下几类：微分法（differential method）、基于频率的方法（frequency-based method）、基于相关的方法（correlation-based method）、多重运动方法（multiple motion method）和时间细化方法（temporal refinement method）等。光流计算方法的另外一种划分为：微分方法、区域匹配法（region-based matching method）、基于能量的方法（energy-based method）和基于相位的方法（phase-based method）等。

在图像序列中，记时刻 $t$ 时图像上的点 $(x, y)$ 处的灰度为 $I(x, y, t)$ ，它是场景中时刻 $t$ 时物体上某点 $(X, Y, Z)$ 在图像上的像。在时刻 $t + \Delta t$ 时，这一点运动到 $(X + \Delta X, Y + \Delta Y, Z + \Delta Z)$ ，它在图像上的像变为 $(x + \Delta x, y + \Delta y)$ ，于是时刻 $t + \Delta t$ 时图像上点 $(x + \Delta x, y + \Delta y)$ 处的灰度可记为 $I(x + \Delta x, y + \Delta y, t + \Delta t)$ ，设它与 $I(x, y, t)$ 相等，即

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) \quad (3.1)$$

显然，这一约束还不能唯一地求解 $\Delta x$ 和 $\Delta y$ ，因此还需要其它约束，比如运动场处处连续等。设亮度随 $x, y, t$ 光滑变化，且 $\Delta x$ 和 $\Delta y$ 足够小，则可将上式左边用泰勒级数展开，得到新的等式

$$I(x, y, t) + \frac{\partial I}{\partial x} \cdot \Delta x + \frac{\partial I}{\partial y} \cdot \Delta y + \frac{\partial I}{\partial t} \cdot \Delta t + e = I(x, y, t) \quad (3.2)$$

其中 $e$ 是关于 $\Delta x, \Delta y, \Delta t$ 的二阶和二阶以上的多项式。整理(3.2)式，两边的 $I(x, y, t)$ 相互抵消，在两边除以 $\Delta t$ ，并取极限 $\Delta t \rightarrow 0$ ，略去等式左边二阶及二阶以上的恶多项式，从而得到

$$\frac{\partial I}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial I}{\partial y} \cdot \frac{dy}{dt} + \frac{\partial I}{\partial t} = 0 \quad (3.3)$$



记

$$u(x, y, t) = \frac{dx}{dt} \quad v(x, y, t) = \frac{dy}{dt} \quad (3.4)$$

并称 $(u, v)$ 为光流场，则有：

$$\frac{\partial I}{\partial x} \cdot u + \frac{\partial I}{\partial y} \cdot v + \frac{\partial I}{\partial t} = 0 \quad (3.5)$$

这就是光流场计算的基本等式，称为光流约束方程。它也可以写成：

$$(\nabla I)^T \cdot U = -\frac{\partial I}{\partial t} \quad (3.6)$$

其中 $\nabla I = \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{bmatrix}$ ， $U = \begin{bmatrix} u \\ v \end{bmatrix}$ ， $T$ 表示转置； $\nabla I$ 是图像灰度值的梯度， $U$ 为光

流场。

由于光流场 $U$ 有两个变量。而基本的光流约束方程只有一个等式，因此，只能求出光流场沿图像灰度值梯度方向的值，而不能同时求出 $u$ 和 $v$ 。因此，必须引入新的约束方程。

Nagel方法[14][15]使用二阶导数来估计光流。提出了一种面向平滑的约束(oriented-smoothness constraint)，并不是强加在亮度梯度变化最剧烈的方向(即边缘方向)上，这样做的目的是为了处理遮挡(occlusion)问题。Lucas-Kanade方法假设在一个小的空间邻域上，运动矢量保持恒定，然后使用加权最小二乘法(weighted-least-squares)来估计光流。Horn-Schunck方法[16]使用光流在整个图像上平滑变化的假设来求解光流，该法搜寻一个运动场，它的流量矢量逐像素变化的最小值满足0FE，即运动场不但满足光流约束方程而且满足全局平滑性。至于自适应法引入的是有向平滑的概念：即沿着边界加平滑约束条件，而不在垂直遮挡边界上加此条件。Fogel[17]在精巧的分级公式中使用有向平滑约束条件和自适应加权概念。

光流与运动场虽然有密切关系，但又不完全对应[18]。场景中的目标运动导致图像中的亮度模式运动，而亮度模式的可见运动产生光流。在理想情况下光流与运动场相对应，但实际中也有不对应的时候。换句话说：运动产生光流。因而有光流一定存在运动，然而并不是有运动就必定有光流，光流和运动是有区别的。

考虑下列两种情况:

图像缺乏足够的空间梯度: 在实际运动能被观察到的运动范围内要有足够的灰度等级(颜色)变化。下图(图3.1)给出了一个不可被观测的运动的例子。图中一个均匀亮度的圆绕它的中心旋转, 这个运动没有产生任何光流, 因此不可观测。

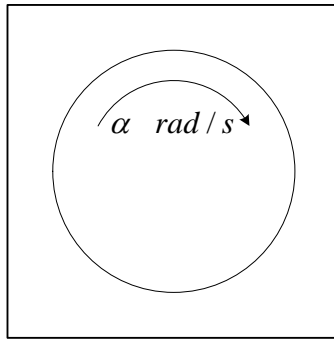


图 3.1 各投影运动不产生光流

外部光照的变化: 一个可观测到的光流(图3.2), 并不总是对应于实际的运动, 例如: 外部照明的色调按照一帧接一帧的速度变化, 即使没有运动, 光流也可被观测到。因此, 外部照明变化妨碍了真实的二维运动场估算。

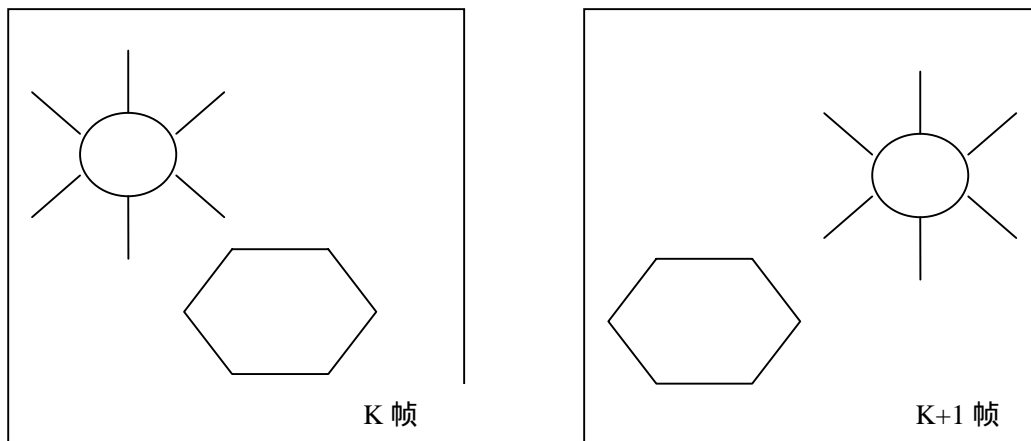


图 3.2 : 一个可观测到的光流

在一些情况下, 即使外部照明没有变化, 色调亦会一帧接一帧地变化。例如, 如果一个物体绕它的表面轮廓线变化, 则会引起色调的变化。这个色调变化可能引起沿运动轨迹的像素的亮度变化, 这些都是二维运动估算中需要考虑的。此外, 光流法的计算量非常大, 而且对噪声比较敏感, 对硬件要求比较高。

### 3.2.2 基于空间像素点灰度的方法

又称代表点匹配法(Representative Points Matching, RPM)[19]。代表点匹配法应用在小型摄像系统中,是在图像中选择一些离散的像素点构成统一的坐标系。代表点是图像上最简单的特征量,通常是将图像分成四个区域,在每个区域选取一定数量的代表点进行匹配。每个点有一个确定的搜索区域,建立先前帧与当前帧图像代表点之间的关系式,最后以搜索区域内选定的代表点为相对参考点做一个相同位移值,相应的有一个灰度绝对差值,对所有代表点的绝对差值求和,获得一个相关匹配函数关系式如下:

$$P(\delta_x, \delta_y) = \sum \left[ I_t(x_k + \delta_x, y_k + \delta_y) - I_{t-1}(x_k, y_k) \right] \quad (3.7)$$

其中  $I_{t-1}(x_k, y_k)$  表示先前帧图像上第  $k$  个代表点像素的灰度值;

$I_t(x_k + \delta_x, y_k + \delta_y)$  表示当前帧图像第  $k$  个代表点像素在偏移  $(\delta_x, \delta_y)$  个单位后像素的灰度值。 $P(\delta_x, \delta_y)$  表示当前帧图像在偏移  $(\delta_x, \delta_y)$  个单位后与参考帧图像的相关匹配值。

$$v_e = \{\delta_x, \delta_y\} = \min(P(\delta_x, \delta_y)) \quad (3.8)$$

$v_e$  表示在当前帧图像所有偏移位置中与参考帧代表点相关匹配值最小时的偏移矢量,就是该区域的图像偏移矢量,将四个区域综合考虑,就可得到我们所需的图像间的偏移矢量。此方法的优点是算法简单、计算量小。但是由于代表点是固定的,并非是图像上有明显特征的点,因此对图像的变化不敏感,且图像的旋转晃动不能得到补偿。

### 3.2.3 位平面匹配法

此方法利用图像的位平面和灰度码实现快速二值匹配来确定图像局部运动矢量,显著的降低了图像处理的运算量,提高了系统的运行速度[20][21]。

对一幅用多比特表示其灰度值的图像来说,其中的每个比特都可以看作一个二值的平面,也称为位平面。一个8位图像有8个位平面,对应每一个位平面有一个级别,第  $k$  个位平面的级别是  $2^k$ , 这样一个8位图像在  $(x, y)$  点处的像素灰度值

表示为：

$$f(x, y) = b_7 * 2^7 + b_6 * 2^6 + \Lambda + b_k * 2^k + \Lambda + b_1 * 2^1 + b_0 * 2^0 \quad (3.9)$$

其中  $b_k$  取值是0或1。图像第  $k$  个位平面用  $b_k(x, y)$  表示，这样一个8位图像就可以看作是由8个二值位平面图像  $b_0(x, y)$ ,  $b_1(x, y)$ , ...,  $b_7(x, y)$  构成，每个位平面描述图像部分信息。但是这种分解方法一个明显的缺陷是像素点灰度值的微小变化有可能对位平面的复杂程度产生明显的影响。例如灰度值为127，比特值为(01111111)，灰度值为128，比特值为(10000000)，灰度值只差1，而8个位平面的比特数值却各不相同。因此，为了减小灰度值变化对位平面的影响采用如下的格雷码位平面分解：

$$g_k(x, y) = \begin{cases} b_k(x, y) \oplus b_{k+1}(x, y) & 0 \leq k \leq 6 \\ b_k(x, y) & k = 7 \end{cases} \quad (3.10)$$

式中  $\oplus$  表示异或符号。经过比特值转换后，相邻的灰度码仅有一位比特值不同。这样，灰度级的变化对所有  $k$  个位平面的影响很小，例如灰度级分别为127和128时，根据公式(3.10)计算得到灰度码分别为(11000000)和(01000000)，即相邻的灰度级仅有一位灰度码不同。式(3.11)表示了当前帧和参考帧位平面的相关匹配公式：

$$c_i(m, n) = \sum_{(x, y) \in S_i} g_k^t(x, y) \oplus g_k^{t-1}(x+m, y+n) \quad (3.11)$$

其中  $g_k^t(x, y)$  表示当前帧第  $k$  层位平面上  $(x, y)$  点的比特值， $g_k^t(x+m, y+n)$  表示参考帧从  $(x, y)$  点沿X轴和Y轴分别运动  $m$  和  $n$  个单位后第  $k$  层位平面像素的比特值， $S_i$  表示参与相关运算的匹配块的面积。公式(3-7)对于搜索范围内的每一个  $(m, n)$  计算  $c_i(m, n)$ ， $c_i(m, n)$  是当前图像和参考图像间的不匹配的比特数。 $c_i(m, n)$  最小时可获得最佳匹配效果，这时候的  $(m, n)$  就是该区域的偏移矢量。这种算法不同于以往的匹配算法，只需利用简单的布尔函数与异或操作，大大减小了计算量。

由于每一个位平面包含的信息不同，算法中又只用到了一个位平面，所以选取一个合适的位平面是非常重要的。或者使用多个位平面的加权值进行计算。同

样的，该算法也只能补偿平移引起的图像抖动。

### 3.2.4 基于空间特征的方法

特征跟踪法[22]是稳像算法中获取图像运动矢量的重要算法，由于在获取图像运动矢量时采用像素点匹配计算量大，而基于频域的分析方法对图像所含信息的利用程度低，所以特征跟踪法是较常用的方法。主要被利用的特征量有角点、直边缘、曲边缘等局部特征和型心、表面积、惯量矩的长短轴等全局特征。特征量跟踪时，会面临一个严重问题，在真实场景中，各特征量所处的环境往往很复杂，其中经常有各种干扰，如目标的遮挡、景物的阴影、背景的变化等，这些变化使得特征量的跟踪变得十分困难。利用特征跟踪法获取各特征量的运动矢量是图像的局部运动矢量，稳定图像还要求出图像由摄像机引起的全局运动矢量。全局运动矢量是在局部运动矢量的基础上，利用准确设定的图像运动数学模型来求取的。

### 3.2.5 基于图像块的方法

基于块(Block-based)的运动分析(运动估算和补偿)[23][24]可以说是最通用的算法，在国际标准中已经采用基于块的运动补偿进行数字视频的压缩(如国际标准H. 261和MPEG1, MPEG2)。块运动估计与光流算法不同，无需计算每一个像素的运动而只是计算由若干像素组成的像素块的运动，是一种近似的表达。

物体的运动由多种元素构成，包括平移、旋转及扭曲等，完整地描述物体运动的模型是复杂的。在目前通用的图像压缩标准及大多数实用的图像压缩算法中，均使用了简化的运动模型，即假设运动是由平移构成，这样只用  $X$  和  $Y$  方向（或水平与垂直方向）的两个平移参数  $dx$ ， $dy$  表征运动参数。由  $dx$ ， $dy$  构成一个运动矢量  $\vec{D}$

$$\vec{D} = (dx, dy)^T \quad (3.12)$$

当前帧  $(i, j)$  坐标点的像素值由前一帧预测为

$$MCP(i, j, k) = \hat{S}(i, j, k) = S(i + dx, j + dy, k - 1) \quad (3.13)$$

这里  $k$  表示当前帧， $k-1$  表示前一帧， $S(i, j, k)$  表示原始像素值， $\hat{S}(i, j, k)$  表示预测的像素值。

当前帧原始图像与预测图像之间可能会有误差，由于这个误差是由平移运动补偿预测引起的，故称为帧间位移误差 DFD (displacement frame difference)，表示如下：

$$DFD(i, j, k) = S(i, j, k) - \hat{S}(i, j, k) = S(i, j, k) - \hat{S}(i + dx, j + dy, k - 1) \quad (3.14)$$

有了这些结果，对图像帧内像素的编码变成对 DFD 场和运动矢量  $\vec{D}$  的编码。由于相邻帧的高相关性，运动补偿是有效的，DFD 场的能量是非常低的，可以用很少的码字表示；运动矢量是稀疏的，同样用较少码字表示，这样就可以获得有效的码率压缩。

这里，平移运动假设的有效性是有条件的，如果一个运动物体较大时，将它分成一些小块，对于每个小块的平移与旋转运动可以用一个平移运动矢量来逼近。如果进行分块运动估计，则每个块的运动矢量可以逼近平移和旋转运动，所以块匹配运动估计是目前应用最广泛的。

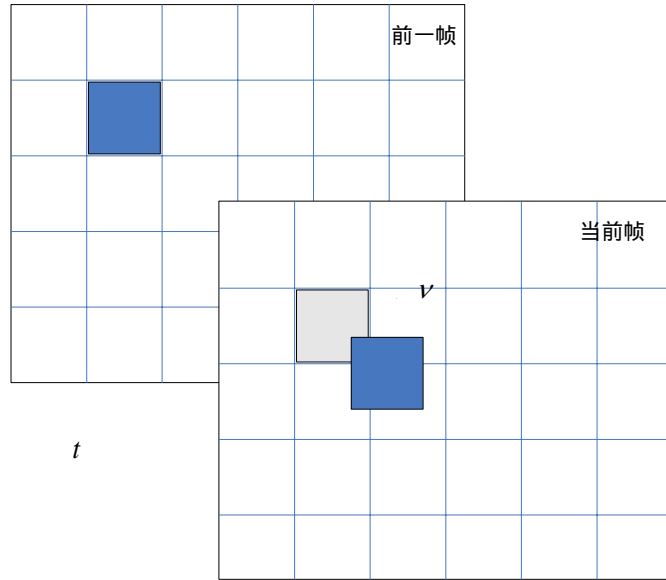


图 3.3：基于图像块的方法示意图

### 3.3 本章小结

本章详细叙述和比较了目前在稳像系统设计中所用到的关键算法。它们包括光流法、基于空间像素点灰度的方法、位平面匹配法、基于空间特征的方法、基于图像块的方法。光流表达了图像的变化,由于它包含了目标运动的信息,因此利用光流法可以用来确定目标的运动情况。但应用光流法有诸多的限制和约束条件,且光流法的计算量非常大,而且对噪声比较敏感,对硬件要求比较高。基于空间像素点灰度的方法中,代表点是固定的,并非是图像上有明显特征的点,因此对图像的变化不敏感,位平面匹配法利用图像的位平面和灰度码实现快速二值匹配来确定图像局部运动矢量,显著的降低了图像处理的运算量,提高了系统的运行速度。运用此方法选取一个合适的位平面是非常重要的基于空间特征的方法在获取图像运动矢量时采用像素点匹配计算量大,而基于频域的分析方法对图像所含信息的利用程度低,所以特征跟踪法是较常用的方法基于图像块的方法无需计算每一个像素的运动而只是计算由若干像素组成的像素块的运动,是一种近似的表达。

## 4 基于特征块的稳像算法的设计

### 4.1 图像稳定系统的一般结构

通常大多数的图像稳定系统按功能都可分为两大部分[8][25]: 运动估计和运动补偿。各类系统之间的差异在于运动估计模型的选择、运动补偿的算法、实现的软硬件载体及应用场合的不同等。很明显, 第一步是第二步的基础。

在序列图像中, 相邻帧间的主要变化是由于构成景物的各物体的运动引起的, 检测物体的运动参数, 并通过这些运动参数由前一帧预测当前帧, 这就是运动补偿预测 MCP (motion compensated prediction), 或简称运动补偿 (MC)。其中主要的任务是检测物体的运动参数, 称为运动估计 ME (motion estimation)。正如

运动估计 (Motion Estimation) 的主要问题是运动模型的选择和图像运动的检测。因此它首先需要的是建立一个运动的模型或者说是摄像机成像的模型, 然后在此模型的基础上计算和估计摄像机的运动参数。

运动补偿 (Motion Compensation) 的主要目的是恢复图像序列各帧中的“本来”面目, 或者说是根据运动估计的结果去除不想要的 (unwanted) 运动, 使观察者看起来有摄像机的晃动已被消除或减弱的感觉, 从而达到稳定图像序列的目的。运动补偿的算法往往决定了整个软件的结构。

正如前章所述, 物体的运动由多种元素构成, 包括平移、旋转及扭曲等, 完整地描述物体运动的模型是复杂的。在目前通用的算法中, 均使用了简化的运动模型, 即假设运动是由平移构成,

下图所示的是普通数字图像稳定系统的模型图 (图4.1)

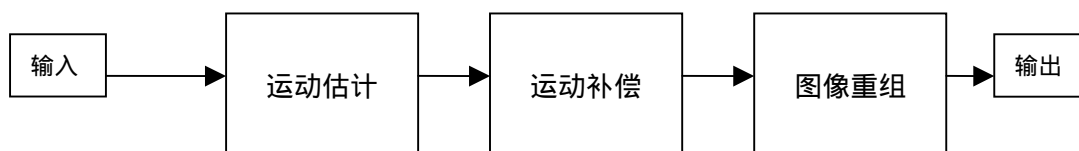


图 4.1 普通数字图像稳定系统的模型图



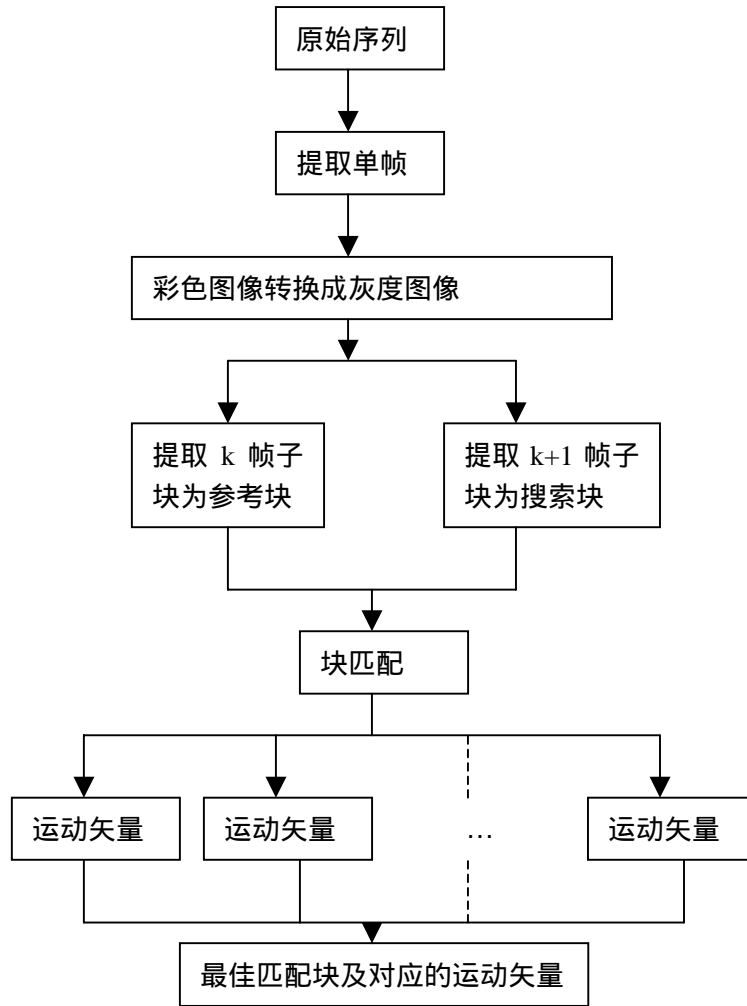
与消除由于摄像系统颤动而引起视频图像模糊的稳像思想不同的是,我们需要消除的是清晰视频图像中存在的由于行走或类似原因(如:开车)而造成的图像大幅度平移和晃动,即对“视觉中心”的稳定。

“基于特征块”的稳像算法即是通过“特征块”在搜索区域内实施快速有效的稳像得出当前帧的运动矢量(运动矢量)。图像帧根据此位移参数进行上下或左右的平移(运动补偿)。“特征块”的位置可以通过参数进行调整。

## 4.2 运动估计算法

在运动估计中求得平移、旋转参数后,利用这些参数对原始图像进行矫正以获得稳定的图像序列。运动估计就是求两帧间的运动,稳像算法就是关于运动估计的算法(在前一章已经作了详细的介绍)。

运动估计流程图



### 4.3 二维运动模型

不同的 2D 运动模型描述了不同类型的运动，如仿射模型能够描述平面对象在正交投影下的任意 3D 运动，并能描述大部分的摄像机运动。我们将摄像机提取的动态图像通过二维运动模型来获得运动矢量。根据运动复杂性的不同，描述二维物体的运动模型有很多种。仿射运动模型是一种线性变换，他可以用来描述物体的平移、旋转和缩放运动，仿射变换可以表示为：

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \quad (4.1)$$

这里  $(x_0, y_0), (x_1, y_1)$  分别是参考图像和当前图像中的像素点的坐标，其中  $c_1, c_2$  用来描述平移运动， $(a_1, a_2, b_1, b_2)$  用来描述缩放和图像的旋转变换，如果我们不考

虑图像的旋转与缩放变换，则仿射变换可表示为：

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \quad (4.2)$$

这时一个相对简单的平移模型，运动矢量为  $V = (c_1, c_2)^T$ 。 (4.3)

## 4.4 平移算法原理及准则

假设图像只在  $x$  和  $y$  方向上作平移运动。平移的实现过程如下：

- (1) 定义判别函数；
- (2) 确定参考块的大小和搜索范围，在比较帧的搜索块内，利用一定的准则，计算参考块与每个搜索块的运动估计；
- (3) 按照判别准则确定与参考块最优的搜索块，得到相对位移矢量。

下面对主要的平移准则作揖简要的介绍。

三种平移准则：绝对误差和(SAD, sum of absolute difference)、均方误差(MSE, mean square error)和归一化互相关函数(NCCF normalized cross correlation function)，分别定义如下：

- (1) 绝对误差和

$$SAD(\delta_x, \delta_y) = \sum_{i=1}^n \sum_{j=1}^m |I_t(i, j) - I_{t-1}(i + \delta_x, j + \delta_y)| \quad (4.4)$$

式中  $(\delta_x, \delta_y)$  是位移矢量。若在某一位置  $SAD(\delta_x, \delta_y)$  达到最小，则该点就是要找的最优匹配点， $I_t$  和  $I_{t-1}$  分别是当前帧和上一帧的灰度值。

- (2) 均方误差

$$MSE(\delta_x, \delta_y) = \frac{1}{mn} \sum_{i=1}^n \sum_{j=1}^m [I_t(i, j) - I_{t-1}(i + \delta_x, y + \delta_y)]^2 \quad (4.5)$$

同样，MSE值最小点位最优匹配点。

- (3) 归一化互相关函数

$$NCCF(\delta_x, \delta_y) = \frac{\sum_{i=1}^n \sum_{j=1}^m I_t(i, j) I_{t-1}(i + \delta_x, j + \delta_y)}{\left[ \sum_{i=1}^n \sum_{j=1}^m I_t^2(i, j) \right]^{\frac{1}{2}} \left[ \sum_{i=1}^n \sum_{j=1}^m I_{t-1}^2(i + \delta_x, j + \delta_y) \right]} \quad (4.6)$$

这里，NCCF的最大点为最优匹配点。

具体选用那种准则对匹配的结果影响不大[26]。但上述三种准则中，由于绝对误差和准则不需要做乘法，实现简单、方便，所以用的最多。我们的系统需要实时快速有效的算法，故采用SAD准则。

## 4.5 搜索过程

确定了匹配准则后，就要通过运动搜索获得最佳匹配块，运动搜索时按照搜索算法指定的搜索路径寻找最佳匹配块以及最匹配的运动矢量。各种搜索算法的主要区别在于如何选择指定窗口内的搜索路径和搜索次数，也就是尽量提高搜索精度和减少搜索次数。

寻找最佳匹配块需要在每个象素 $(n_1, n_2)$ 所有可能的候选位移矢量上优化平移算法准则。这可以通过所谓的“全面搜索”来实现，即对每一个像素的所有 $(d_1, d_2)$ 值评测平移算法准则，这样做的话，相当费时。

为了减少计算负担，我们通常限制搜索区域在

$$-M_1 \leq d_1 \leq M_1 \text{ 且 } -M_2 \leq d_2 \leq M_2 \quad (4.7)$$

“搜索窗口”中心定位在运动矢量将被估算的每一个像素上，其中 $M_1$ 和 $M_2$ 是预定的整数。另一个常用来降低计算负担的做法是：在像素的稀疏网格上估算运动矢量，例如，使用一个 $16 \times 16$ 块的每8个像素和8条线估算一次，然后插入运动场来估算剩余的运动矢量。

在多数情况下，常常使用比全范围搜索快得多的快速搜索方案，虽然它们只产生次优解。为了加快运算速度，保证精度，人们提出了很多快速搜索算法[27][28][29]：三步法（TSS）及基于三步法的改进算法、二维对数法（LOGS）、交叉搜索法（CS）、四步法（4SS）、预测搜索法（PSA）、钻石搜索法（DS）等。

这些快速搜索算法仅在候选的运动矢量位置的预定子集上评测准则函数。我们必须注意到,这里运动估算所期望的精确度是依据应用变化的。在运动补偿的压缩中,依据某个度量标准,我们所搜索的是匹配块。

### 三步搜索 (TSS) Three Step Search

与二维对数搜索法类似,由于简单、健壮、性能良好的特点,为人们所重视。

第一步:以搜索区的中心点(0,0)为中心点,首先搜索一个含有9个点的 $9 \times 9$ 的窗口,找到一个最佳匹配点。

第二步:以上一步找到最佳匹配点为中心,搜索一个含有9个点的 $5 \times 5$ 窗口,找到一个最佳匹配点。

第三步:以上一步找到的最佳匹配点为中心,搜索一个含有9个点的 $3 \times 3$ 的窗口,找到一个最佳匹配点就是最终的匹配点。

### 二维对数法 Two Dimensional Logarithmic Search

二维对数法(TDL)的基本思想是通过快速搜索法来跟踪最小 SAD 点。从运算矢量(0,0)开始,以十字形分布的五个点构成每次搜索的点群,求出 SAD 最小点。若 SAD 最小点出现在十字形点群的边缘点,则下次搜索以该点为新十字点搜索点群的中心点,而步长不变。若 SAD 最小点出现在十字点群之中心点,则下次搜索仍以该点为中心点构成十字形点群,但步长减半。以此类推,直至补偿为 1,此时所找到的 SAD 最小点即为所求最优匹配点。需要注意的是,在上述搜索过程中,若发现新的十字形点群的中心点位于搜索区的边缘,则步长也减半。

该方法的前提是假设搜索区内只有一个谷点,但如果搜索区内存在多个谷点时,该方法找到的可能是局部最优点,而不是全局最优点,这也是大部分快速搜索算法的通病。

### 钻石搜索法 (DS)

又叫菱形搜索法,以其搜索模式的形状而得名,使一种优秀的搜索方法,其思想是减少进行块匹配的搜索点。

钻石搜索法如图 4.2 所示,可以看到搜索点大都位于图像的水平 and 垂直方向

上,这是考虑到现实中的物体在这两个方向的运动概率比较大,图像的频谱多呈菱形分布,也就是说,是由于图像在水平和垂直方向的相关性大于斜线方向。

在搜索过程中,一旦某一搜索点的 SAD 值小于预先设置的门限制,则停止搜索。如果搜索完毕仍然不能找到小于门限值 的点,也不再继续搜索,以其中最小 SAD 点计算,然而这种情况发生的概率很小。

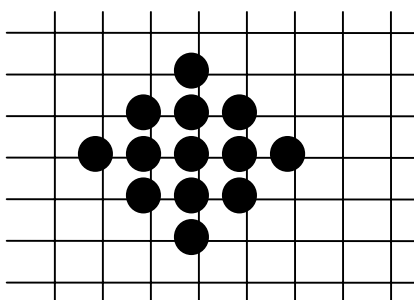


图 4.2 : 钻石搜索法

采用两种搜索模式: 共计 9 个点的大钻石搜索模式 (LDSP) 和 5 个点的小钻石搜索模式 (SDSP)。先用 LDSP 搜索, 由于步长大, 搜索范围广, 可以进行粗定位, 使搜索过程不会陷于局部最小。当粗定位结束后, 可以认为最优搜索点就在 LDSP 周围 8 个点所围的菱形区域中, 这是再用 SDSP 来准确定位, 使搜索不至于有大的起伏, 所以它的性能优于其它快速算法。

钻石搜索法的 LDSP 模式下是一个循环结构, 在每次循环中均需计算围绕中心点的 8 个点。但是如果每次均计算 8 个点, 则无论该中心点是前次搜索过程中钻石形模式的边点或是角点, 都必然有 5 个或 3 个点在相邻两次搜索中重复计算, 这样增加了时间复杂度。

### 金字塔搜索法

这是一种减少匹配搜索时间的有效方法, 它采用金字塔式的数据结构, 通过从粗糙图像(即低分辨率图像)开始模板匹配, 找粗匹配点, 逐步找到原始图像(即最高分辨率图像)的精确匹配点。这是一种由粗到细的匹配方法, 其步骤如下:

- A. 对待匹配的两幅原图像中 $2 \times 2$ 领域内的像点灰度值取平均, 得到分辨率低一级的图像。
- B. 对低一级图像中 $2 \times 2$ 领域内的像点灰度值再取平均, 得到分辨率更低一级的图像。依次处理, 可以得到一组呈金字塔式的图像, 假设有 $L+1$ 级, 其中 $L=0$ 即

原图像。

C. 从待匹配的两幅图的某一低分辨率级开始进行匹配搜索，由于这两幅图不但像素数目少，高频信息也平滑掉一部分，因此粗匹配结果可能出现不止一个匹配位置，但由于图中点数少，匹配搜索的过程很快。

D. 对高一级分辨率的图像进行匹配搜索，但搜索空间只限于一个或几个粗匹配点附近，所以计算量并不大。依次类推，直到 $K=0$ 的那一级(即原图像)找到二图的匹配点为止。上述算法的加快程度从总搜索位置数的减少可以看出：第一次(最低分辨率一级)是全部搜索，搜索位置数为 $(N/2^L - M/2 + 1)^2$ ，第二次到最高分辨率一级都只在粗匹配点附近搜索，所以本方法总的搜索位置数约为普通不分级搜索位置数的 $1/4^L$ ，若 $L=3$ ，就只有普通搜索位置的 $1/64$ ，扣除预处理多花的时间，计算总量仍减少了许多。

通过对以上各个搜索算法的比较分析我们了解了每种搜索算法的适用性及优缺点。我的系统设计中考虑到假体稳像系统的目标选择性要求，采用“基于特征块”的全搜索方法。此方法的主要思想是通过特定区域的全搜索法达到视频图像稳定的目的。这里的“特定区域”指的是：在系统设计中大小一定的搜索框，搜索框的初始位置根据初始参数来确定（此参数的设定在 matlab 仿真一章中具体说明），我们的全局搜索算法就在此搜索框中进行，这样搜索法在避免了全局搜索的较大的计算量的同时保证了搜索后匹配的精确性。关于系统得设计和仿真和参数设置我们将在下一章给出详细的介绍。

“基于特征块”的稳像方法如图 4.3 所示，选定视频图像中一定大小的区域作为“特征块”，就是我们定位的“视觉中心”。我们力求达到消除这个“视觉中心”在视频图像中的大幅度晃动。将“特征块”作为我们的搜索模板，采用基于模板的全搜索法则实现特征目标稳像的目的。

以视频图像左上角为坐标原点建立坐标系，特征块是一个左上角坐标为 $(x_0, y_0)$ ，长宽分别为  $n, m$  的矩形区域。搜索区域(search window)为左上角坐标为 $(x_1, y_1)$ ，边缘与特征块距离分别为  $a, b$ ，长宽分别为  $n+2a, m+2b$  的矩形区域。“基于特征块”的稳像算法即是通过“特征块”在搜索区域内实施快速有效的

稳像 得出当前帧的运动矢量（运动矢量）。图像帧根据此位移参数进行上下或左右的平移（运动补偿）。接着，再将当前帧作为参考帧（previous frame）重复此算法。

特征块是我们设定的“视觉中心”，将此特征块在搜索块中进行全搜索算法的比对，计算出“视觉中心”的位移矢量，达到获知整个视频图像位移矢量的目的。搜索区域是视频图像的一部分，特征块是搜索区域的一部分。虽然全搜索算法是最为精确的算法，但由于其计算量大，在以往的研究中，全搜索算法用于整幅图像的搜索不尽如人意[7]。而我们用在“基于特征快”的搜索中，其算法优势得到了充分的体现。并且，基于特征块的稳像算法是面向对象的：特征块即是我们的“视觉中心”，是我们感兴趣的内容，其初始位置可以一通过参数进行调整，这样，同时到达了对目标跟踪的目的。

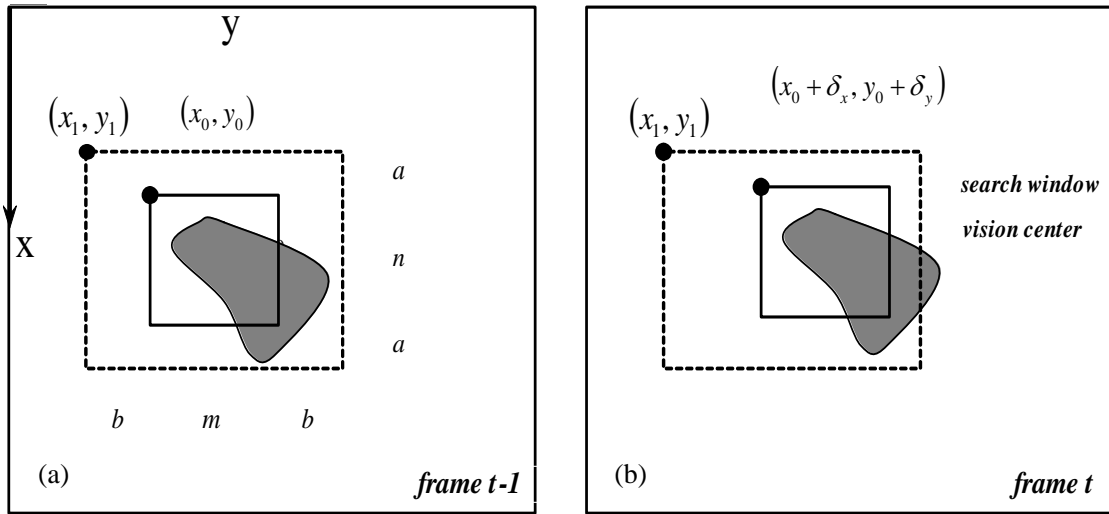


图4.3：基于特征块稳像的核心思想

Fig.4.2 : Illustration of FBS method. (a) is previous frame. Initial positions of vision center and search window are  $(x_0, y_0)$  and  $(x_1, y_1)$  respectively. The size of vision center, which is centered on the search window, is  $n \times m$  at a given pixel. (b) is current frame.  $\delta_x$  and  $\delta_y$  are respectively the horizontal and vertical motion vectors which are acquired by tracking the vision center within search window.



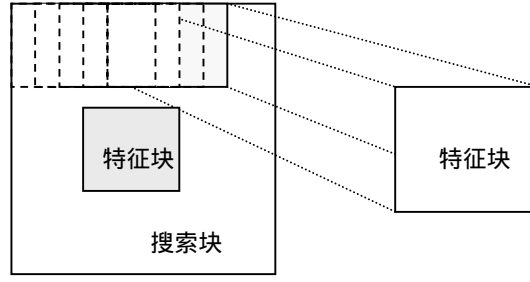


图4.4：搜索方法原理图

Figure 4.4 : Illustration of Feature-block-based method

## 4.6 运动补偿算法

运动补偿的目的就是保持某种运动估计记录以产生稳定图像序列。前面所述运动估计过程的结果,可计算得到两帧间的运动,为了产生一个稳定的图像序列,必须从当前帧到参考帧进行校准变换,运动补偿模块,结合帧间估计,得出这个整体变换。下面的仿射变换用来校正  $f_{t-1}$  已产生相应的稳定帧  $\hat{f}_{t-1}$

$$\hat{f}_{t-1} = R_{t-1}f_{t-1} + T_{t-1} \quad (4.8)$$

帧间换由帧间旋转  $R_t^i$  和平移  $T_t^i$  定义

$$f_{t-1} = R_t^i f_t + T_t^i \quad (4.9)$$

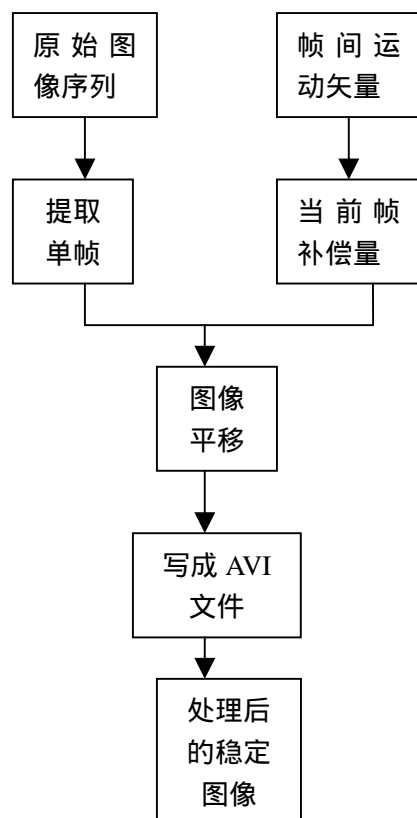
将(4.9)代入(4.8)得:

$$\hat{f}_t = R_{t-1}R_t^i f_t + R_{t-1}T_t^i + T_{t-1} \quad (4.10)$$

所以,新的旋转和平移量由下式给出:

$$R_t = R_{t-1}R_t^i \quad ; \quad T_t = R_{t-1}T_t^i + T_{t-1} \quad (4.11)$$

运动补偿流程图



## 4.7 本章小结

针对基于前一章在稳像核心算法的分析和比较的基础上,本章在进行稳像系统设计时采用“基于特征块”的稳像系统设计方法。本章从稳像系统的一般结构入手,将稳像系统分为运动估计、运动补偿、图像重组三个重要的组成部分。分别设计了运动估计算法流程框架和运动补偿算法流程框架。在运动补偿的平移算法中,采用绝对误差和(SAD, sum of absolute difference)算法,和其它常用的平移准则相比 SAD 准则不需要做乘法,实现简单、方便。在图像块的搜索算法一节中,比较了目前流行的搜索算法,基于本稳像系统得特点,采用基于模块的全匹配搜索法。这种搜索法在避免了全局搜索的较大的计算量的同时保证了搜索后匹配的精确性。

## 5 Matlab 仿真

### 5.1 影像格式

Matlab 支持多种视频图像格式，下面就采集视频图像中最常见的 Matlab 支持的视频格式做一介绍。

#### Microsoft Windows Media 文件--.AVI

AVI ( Audio Video Interleave ) 是 1992 年初由 Microsoft 公司推出的数字视频格式。AVI ( Audio Video Interleaved 的缩写 ) 是一种 RIFF ( Resource Interchange File Format 的缩写 ) 文件格式，多用于音视频捕捉、编辑、回放等应用程序中。通常情况下，一个 AVI 文件可以包含多个不同类型的媒体流 ( 典型的情况下有一个音频流和一个视频流 )，不过含有单一音频流或单一视频流的 AVI 文件也是合法的。AVI 可以算是 Windows 操作系统上最基本的、也是最常用的一种媒体文件格式[30]。

在 AVI 文件中，运动图像和伴音数据是以交织的方式存储，并独立于硬件设备。构成一个 AVI 文件的主要参数包括视像、伴音和压缩参数等：

##### 视像参数：

1. 视窗尺寸 ( Video size )：AVI 的视窗大小可按 4:3 的比例或随意调整，视窗越大，数据量越大。
2. 帧率 ( Frames per second )：帧率也可以调整，而且与数据量成正比。不同的帧率会产生不同的画面连续效果。
3. 伴音参数：在 AVI 文件中，视像和伴音是分别存储的，因此可以把一段视频中的视像与另一段视频中的伴音组合在一起。WAV 文件是 AVI 文件中伴音信号的来源，伴音的基本参数也即 WAV 格式的参数。

除此以外，AVI 文件还包括与音频有关的其他参数：

4. 视像与伴音的交织参数 ( Interlace Audio Every X Frames )：AVI 格式中每 X 帧交织存储的音频信号，也即伴音和视像交替的频率 X 是可调参数，X 的最小值是一帧，即每个视频帧与音频数据交织组织，这是 CD-ROM 上使用的

默认值。交织参数越小，回放 AVI 文件时读到内存中的数据流越少，回放越连续。因此，如果 AVI 文件的存储平台的数据传输率较大，则交错参数可设置得高一些，如几帧，甚至 1 秒。

5. 同步控制 ( Synchronization ) : 在 AVI 文件中，视像和伴音是同步得很好的。但实际上由于处理能力的不够，回放 AVI 时有可能出现视像和伴音不同步的现象。

#### **压缩参数：**

在采集原始模拟视频时可以用不压缩的方式，这样可以获得最优秀的图像质量。生成 AVI 文件时需要根据应用环境的不同选择合适的压缩参数，而其中 AVI 压缩算法是首先要确定的参数之一。

#### **AVI 的特点**

AVI 及其播放器 VFW 已成为了 PC 机上最常用的视频数据格式，是由于其具有如下的一些显著特点：

1. 无硬件视频回放：根据 AVI 格式的参数，其视窗的大小和帧率可以根据播放环境的硬件能力和处理速度进行调整。在低档 MPC 机上或在网络上播放时，VFW 的视窗可以很小，色彩数和帧率可以很低；而在高档系统上，可实现质量较好的回放效果。因此，VFW 就可以适用于不同的硬件平台。
2. 同步控制和实时播放：通过同步控制参数，AVI 可以通过自调整来适应重放环境，如果 MPC 的处理能力不够高，而 AVI 文件的数据率又较大，在 WINDOWS 环境下播放该 AVI 文件时，播放器可以通过丢掉某些帧，调整 AVI 的实际播放数据率来达到视频、音频同步的效果。
3. 高效地读取数据：由于 AVI 数据的交叉存储，VFW 播放 AVI 数据时只需占用有限的内存空间，播放程序可以一边读取视频数据一边播放，而无需预先将容量很大的视频数据加载到内存中。这种方式不仅可以提高系统的工作效率，同时也可以实现迅速地加载和快速地启动播放程序，减少播放 AVI 时用户的等待时间。
4. 开放的文件结构：AVI 文件结构不仅解决了音频和视频的同步问题，而且具有通用和开放的特点。它可以在任何 Windows 环境下工作，而且还具有扩展环境的功能。用户可以开发自己的 AVI 视频文件，在 Windows 环境下可随

时调用。

5. 文件易于再编辑：AVI 一般采用帧内有损压缩，可以用一般的视频编辑软件如 Adobe Premiere 或 MediaStudio 进行再编辑和处理。

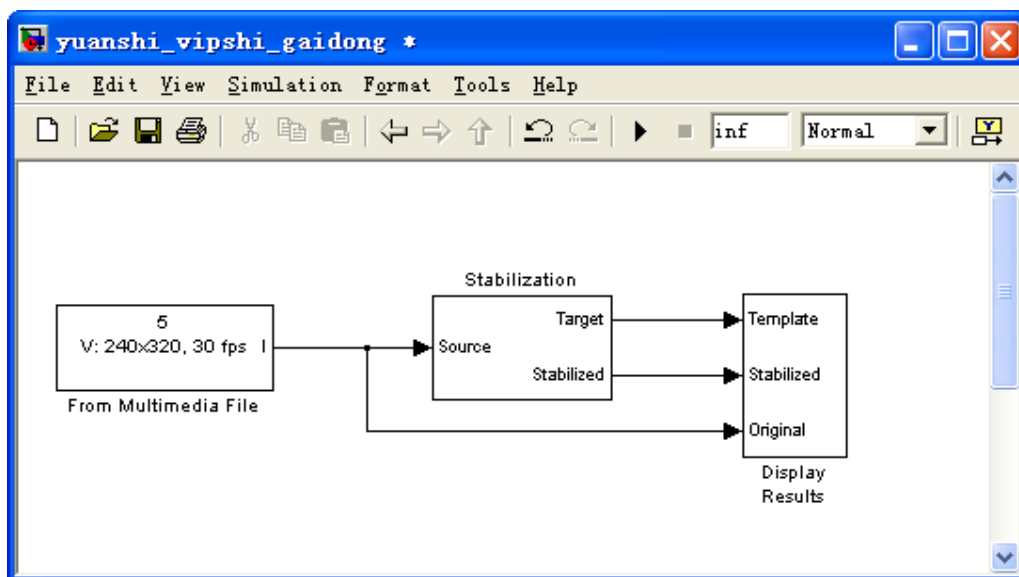
### QuickTime 文件--.MOV/.QT

QuickTime 是 Apple 计算机公司开发的一种音频、视频文件格式，用于保存音频和视频信息，具有先进的视频和音频功能。QuickTime 文件格式支持 25 位彩色，支持 RLE、JPEG 等领先的集成压缩技术，提供 150 多种视频效果，并配有提供了 200 多种 MIDI 兼容音响和设备的声音装置。新版的 QuickTime 进一步扩展了原有功能，包含了基于 Internet 应用的关键特性，能够通过 Internet 提供实时的数字化信息流、工作流与文件回放功能，此外，QuickTime 还采用了一种称为 QuickTime VR (简作 QTVR)技术的虚拟现实(Virtual Reality, VR)技术，用户通过鼠标或键盘的交互式控制，可以观察某一地点周围 360 度的景像，或者从空间任何角度观察某一物体。QuickTime 以其领先的多媒体技术和跨平台特性、较小的存储空间要求、技术细节的独立性以及系统的高度开放性，得到业界的广泛认可，目前已成为数字媒体软件技术领域的事实上的工业标准。国际标准化组织(ISO)最近选择 QuickTime 文件格式作为开发 MPEG 4 规范的统一数字媒体存储格式。

## 5.2 系统仿真设计

本部分的系统仿真设计使用 Matlab 软件仿真[31]。包括对系统总体框架的描述，系统得运动估计，运动补偿，以及视频输出子系统设计的说明。同时，还对系统中的重要参数设置作出了解释和说明。

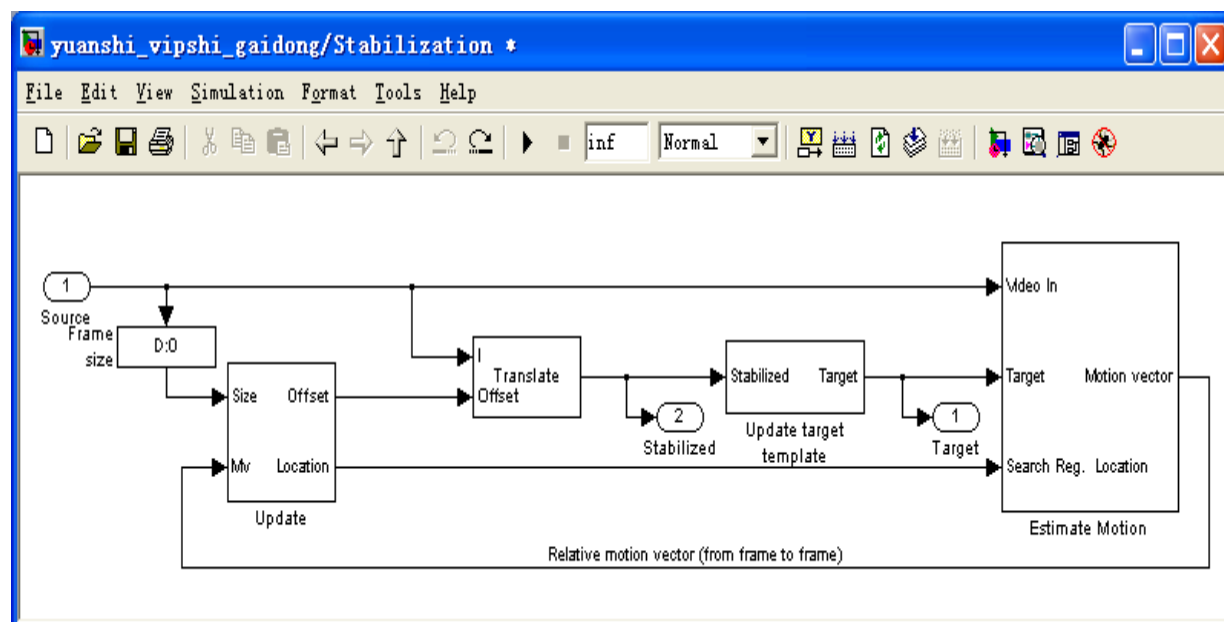
## 5.2.1 总体框架



系统的总体框架如图所示，视频图像通过“From Multimedia File”模块读入系统，之后进入系统的主体模块进行稳像处理。处理后的图像通过视频图像“输出显示”子系统显示结果。

1. From multimedia file 模块：读入视频图像。
2. Stabilization 模块：稳像系统得核心部分。
3. Display results 模块：视频输出显示。

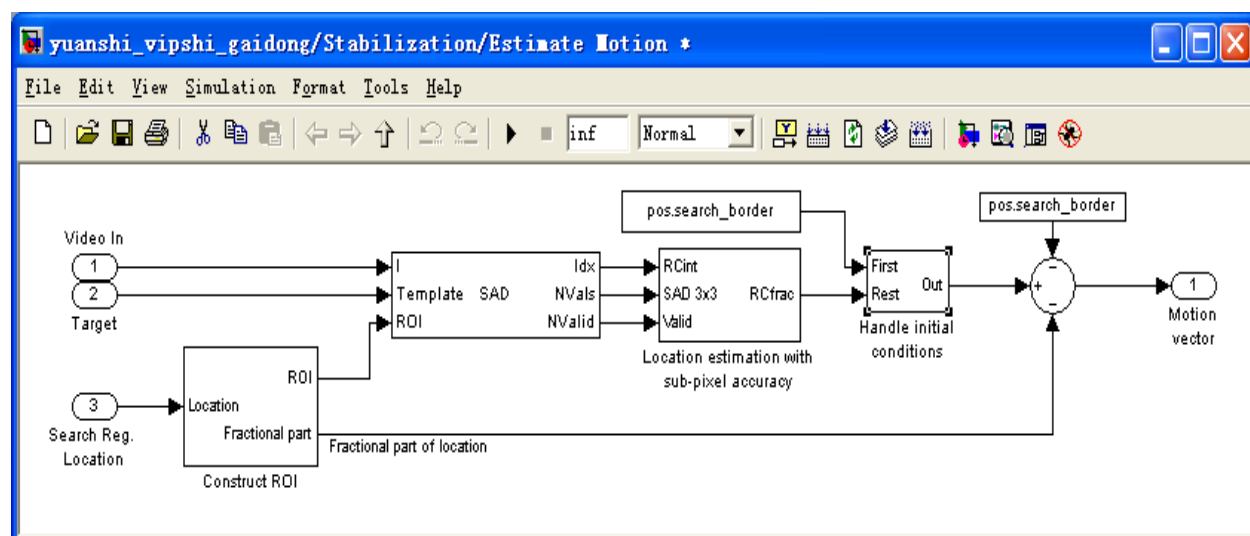
## 5.2.2 视频图像稳定子系统结构



“图像稳定子系统”通过系统的参数设置确定需要跟踪的特征块的位置，搜索区域的大小，定位好系统需要计算和处理的区域后，通过运动估计子系统估计运动参数（运动矢量），系统通过反馈进行更新。

1. Update 模块：接收 Estimate motion 模块的运动矢量，将相邻两帧的运动矢量进行比较，得出相对运动矢量，送给下一模块。
2. Translate 模块：对图像帧进行移动的模块。此模块‘I’口输入原始视频图像，‘offset’输入位移参数，图像帧根据此位移参数进行上下或左右的平移。如位移参数[2 -3]表示图像向下平移 2 个像素，向左平移 3 个像素。
3. Update target template 模块：特征模块更新。
4. Estimate motion 模块：运动估计子系统。对选定区域采用全搜索法来估计相邻两帧的运动矢量。

### 5.2.3 运动估计子系统

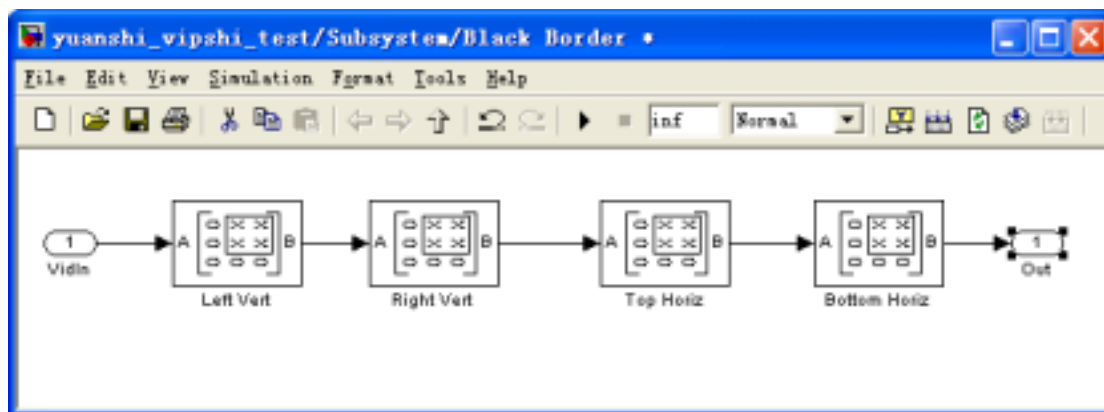


运动估计子系统采用 SAD 搜索准则，提取运动矢量。此运动矢量送给上一级子系统进行帧间的稳像处理。

1. SAD 模块：运用 SAD 搜索准则对指定区域进行搜索。在其参数设置上，采用全搜索法。
2. Location estimation with sub-pixel accuracy 模块：视频图像稳定精确度调整。
3. Handle initial condition 模块：将符合搜索准则的区域和原始特征模块进行比较，计算位移矢量。

4. Conduct ROI 模块：接受反馈，根据为一矢量对视频图像前后帧进行调整。

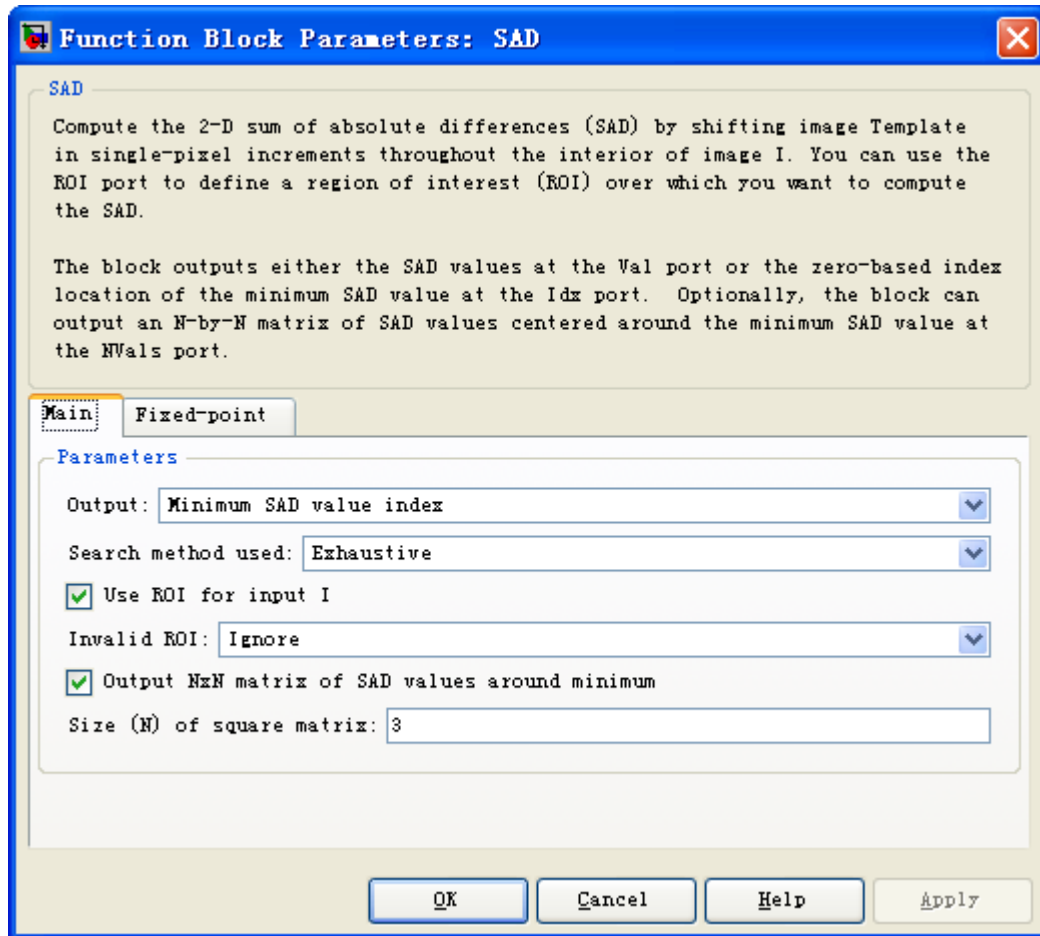
## 5.2.4 图像显示模块设计



上图模块的作用是：给稳像后输出的视频图像四周添加黑色覆盖区域。其中的四个“over write value”模块分别对图像的左右上下进行矩阵“0”填充处理。稳像输出的处理设计：当把设定好的“视觉中心”稳定后，原始视频图像其边缘部分会因为晃动而在视频图像中“时隐时现”，处理后的视频图像应该设定一个输出图像的大小范围，根据测试，一般的行进过程中将图像画面周边按整个图像5%的比例“剪裁”掉，就不会出现因为“边框”晃动而产生的“虚晃动”。

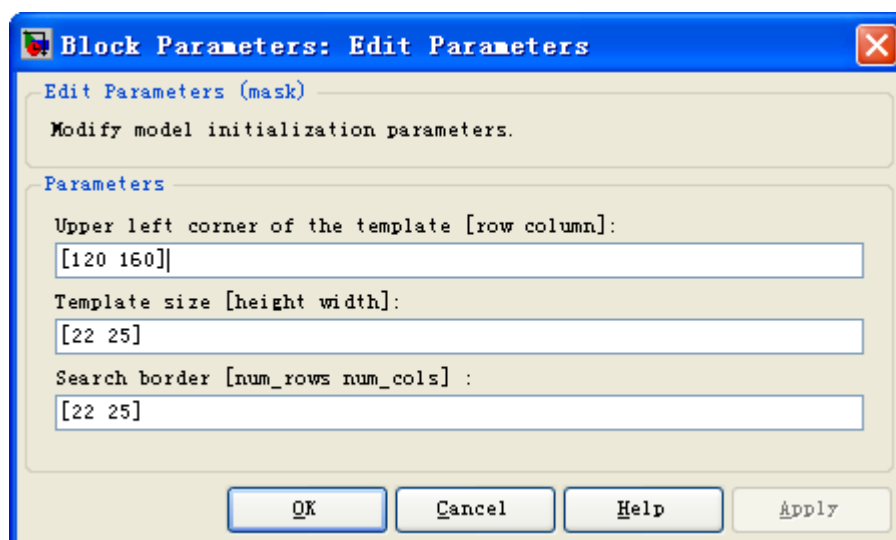


## 5.2.5 系统的参数设计



在 SAD 模块的参数设置中,根据系统的要求(基于特征块的算法),我们需要此模块具有 ROI 端口来确定我们的‘特征块’。

在 SAD 模块参数中的搜索方法我们选择“Exhaustive”,即全搜索法。根据系统的实际情况,全搜索在系统指定的特征模版内进行搜索。此种方法,避免了在整幅图像中进行搜索造成的搜索时间过长,又减少了搜索次数,提高了准确度。



如上图所示是 simulink 中对稳定系统中匹配模版参数的设置。第一列参数是匹配模版中心距离视频图象左上角的距离。第二行参数是搜索模版的大小，第三行参数设定的是搜索区域的大小。

## 5.3 仿真结果

### 3.1 仿真结果

实验采用 blackcar 图像序列的第一帧到第 50 帧。视频图像是  $320 \times 240$  的图像序列，红色特征块是  $50 \times 20$  的图象块，白色搜索区域为  $70 \times 40$  的图像块。



图5.1：原始视频图像与稳定后的视频输

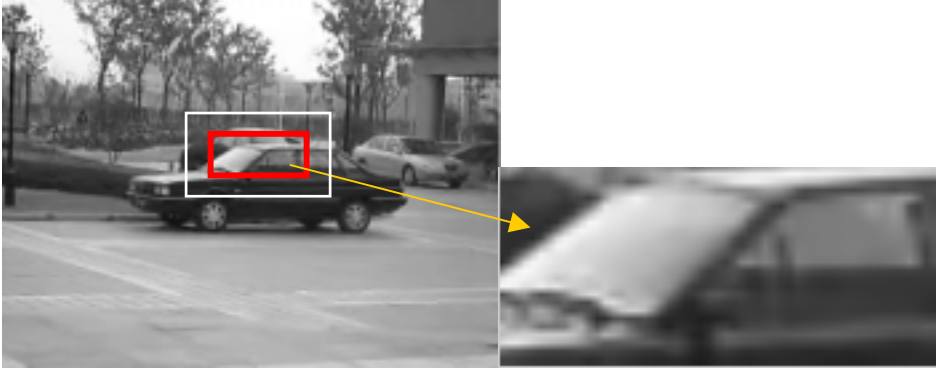


图5.2 : 特征块作为视觉中心

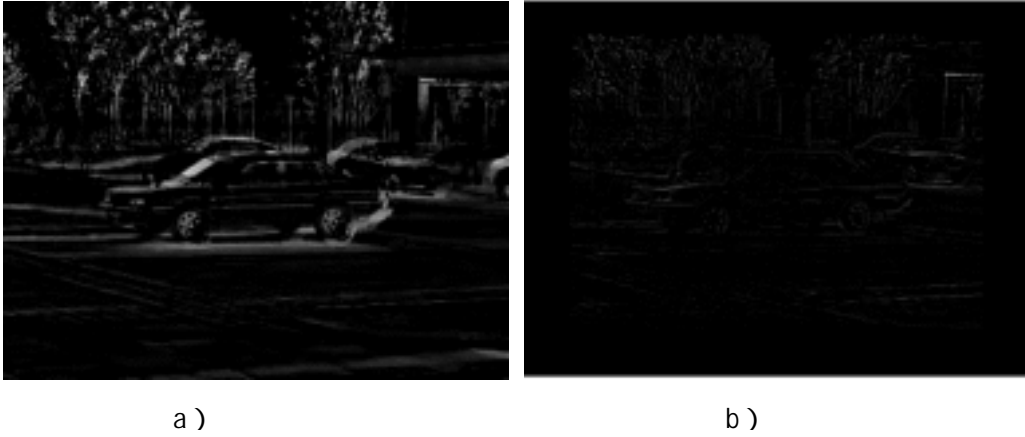


图5.3 : 性能评估. a) 原始图像44与45帧之差; b) 稳定后图像序列44与45帧之差。

## 5.4 性能评估

我们通过反映稳像系统对视频图像经过稳像处理后前后两帧之间的差别来反映稳像系统得性能<sup>4</sup>。在统计中，我们评价一组数据差别的最简单的指标是方差和标准差，也就是说方差和标准差是表示一组数据离散程度的最好指标，是最常用的差异量数。我们利用仿照标准差的定义来定义“连续帧间差异标准差”。

首先计算差异矩阵：后一帧减前一帧  $D_{i,j} = D_t - D_{t-1}$  (5.1)

矩阵元素平均值：矩阵元素平均  $\bar{D} = \frac{\sum_{j=1}^m \sum_{i=1}^n D_{i,j}}{M \times N}$  (5.2)

帧间标准差  $SD = \left( \frac{1}{M \times N - 1} \sum_{j=1}^N \sum_{i=1}^M (D_{i,j} - \bar{D})^2 \right)^{\frac{1}{2}}$  (5.3)

其中  $D_{i,j}$  为差异矩阵中的每一个元素值。

如图 5.4，通过“帧间差异标准差”的度量，我们看到经过稳像系统处理的视

频图像的的稳定程度明显提高。

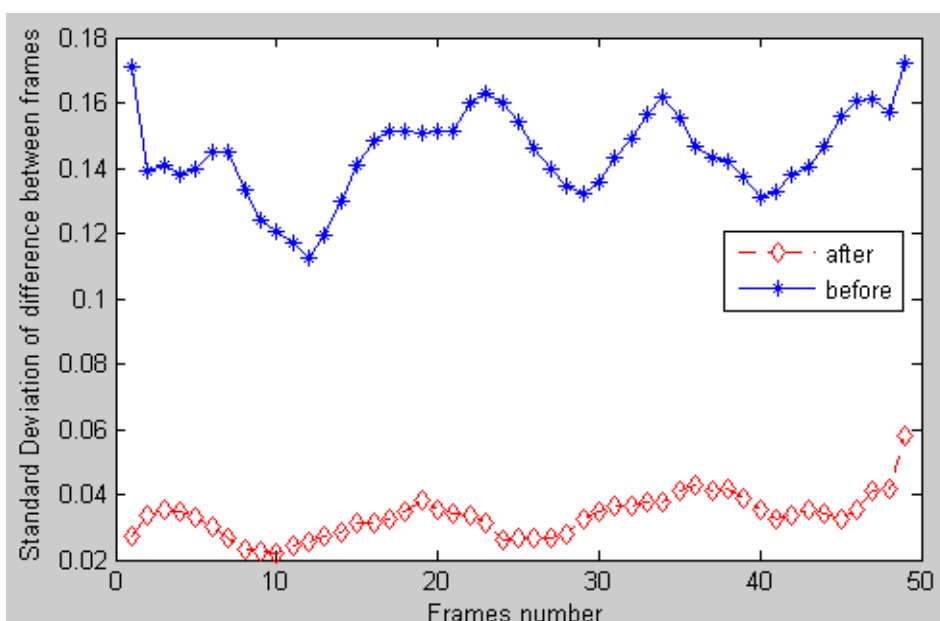


图5.4：性能评估.

## 5.5 本章小结

本章是稳像系统的 Matlab 仿真。Matlab 支持多种视频图像格式，章节重点介绍了试验中用到的 AVI 格式。之后对系统的 Matlab 仿真设计框架给予详细的图示和说明，其中包括整体框架的结构，运用“基于特征块”稳定系统的结构以及运动估计子系统的设计，另外还给出了其中关键参数设置的说明。

在运用此仿真系统时，输入了一段自己录制的带有较大晃动的视频图像，此图像是在正常走动过程中进行录制。通过系统获得的结构显示相邻两帧间的大幅度晃动得到了很好的消除。

在评价此系统的有效性时，采用“帧间差异标准差”进行度量，度量结果显示稳像系统具有很好的稳像效果。

## 6 全文总结

### 6.1 主要结论

视觉假体系统中，图像处理是系统的关键部分。盲人（或普通人）在行走，头部转动，无意识的身体晃动等都可能造成采集的图像整体画面不稳定、颠簸的情况，使得进一步的图像处理难度增大，为产生合适的刺激组型带来困难；我们希望通过数字图像处理技术改善图像序列中的稳定状况，消除因摄像机的无意抖动而造成的视频变化，以保证摄像机在一定的变化范围内，输出的视频是稳定的。

对于图像视频稳定的技术研究来说，大多数科学工作者都是基于数码产品或航空以及军事等领域的研究，在这些领域的研究中，力求最优的图像稳定算法，得到最好的效果，并且具有强大运算功能的芯片也为实现这些要求提供了可能性。视觉假体系统的特殊性使得我们在进行视频图像稳定的实现中具有不同的出发点：（1）视觉假体实时性要求较高，所以要求简单而快速的算法；（2）视觉假体芯片的开发中较高的成本不利于假体的临床推广，需要较为简单的硬件实现；

但目前国内外的有关视觉假体设计的文献中，大都未提及有关视频图像采集后的稳像问题。

本系统设计考察现有的各类稳像方法，电子稳像算法优缺点。改进已有的电子稳像算法，运用快速有效的运动估计和运动补偿准则，设计实用的稳像系统：适合盲人的，即行进过程中产生的图像平移及晃动等等。用 MATLAB 实现视频稳定仿真。评估系统的性能表明，系统具有较好的稳像效果。

### 6.2 研究展望

目前此系统仅仅是离线的系统设计，在系统进一步的设计中，需要增加实时读取视频图像的功能。并增加系统面板的设计部分，为系统的实际运用提供基础。目前系统的人工调节功能包括：目标块位置锁定的设置，根据目标块设定的搜索区域的大小，以及根据晃动程度设定的输出图像大小。我们希望进一步的研究将以上对于系统得的人工数值设定设计成可视化界面中可选或可设定的操作。

我们的目标是将此视频稳定系统用在视觉假体中,对于如何确定“视觉中心”仍然是一个难点,需要结合计算机视觉的研究成果进一步结合计算机视觉的理论,增加系统的特征目标识别功能,以人类知识库来锁定已知内容的目标,并且按照目标需求的不同,根据面板操作来转化“视觉中心”。

## 参 考 文 献

1. 马颂德,张正友.计算机视觉——计算理论与算法基础.北京:科学出版社,1998.
2. Hayes JS, Yin JT, Piyathaisere DV, Weiland J, Humayun MS, Dagnelie G.  
Visually guided performance of simple tasks using simulated prosthetic vision.  
Artif. Organs 27(11):1016–28,2003.
3. Veraart, C., et al., Visual sensations produced by optic nerve stimulation using an implanted  
self-sizing spiral cuff electrode. Brain Res, 1998. 813(1): p. 181-6.
4. Thompson R, Barnett D, Humayun M, Dagnelie G., Reading speed and facial  
recognition using simulated prosthetic vision. Invest. Ophthalmol. Vis. Sci.  
41:4571 (Abstr.), 2000.
5. Zrenner, E., Will retinal implants restore vision? Science, 295(5557):  
1022-5,2002.
6. Hallum, L.E., et al. Towards photosensor movement-adaptive image analysis for  
a retinal prosthesis [C]. in Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 26th., San  
Francisco, CA, 2004.
7. 刘远航,刘文开,刘畅,丁启芬 编著,视频捕捉设备——数字摄录一体机,  
人民邮电出版社出版发行,北京,2001.6
8. DUFAUX F, MOSCHENI F. Motion estimation techniques for digital TV[A]:a  
review and a new contribution,Proc. IEEE: ,83(6):858 – 876, 1995
9. 石萍,邱意弘,朱贻盛,童善保.人工视觉假体研究综述(1)——视网膜假  
体研究现状,生物医学工程学杂志,已录用;
10. 石萍,邱意弘,朱贻盛,童善保.人工视觉假体研究综述(2)——视皮层、  
视神经束、感觉替代假体的研究现状,生物医学工程学杂志,已录用。
11. Liu WT, Sivaprakasam M, Singh PR, et al. Electronic visual prosthesis. Artificial  
Organs, 27(11): 986,2003;
12. Pauwels, K. and Van Hulle, M.M., Optic Flow from Unstable Sequences  
containing Unconstrained Scenes through Local Velocity Constancy  
Maximization. Proceedings of the British Machine Vision Conference (BMVC),

- Edinburgh, September 4-7 2006, volume 1, pp. 397-406.
13. J. Chang, W. Hu, M. Cheng, and B. Chang. Digital image translational and rotational motion stabilization using optical flow technique. *IEEE Transactions on Consumer Electronics*, 48(1):108-115, 2002.
  14. [Nagel 1987] H H Nagel. On the estimation of optical flow : Relations between different approaches and some new results. *Artificial Intelligence* , 33 : 299~324, 1987.
  15. [Uras 1988] S Uras , F Girosi , A Verri , V Torre. A computational approach to motion perception. *Biol. Cybern.* , 66 : 79~97 , 1988.
  16. [Horn 1981] B K P Horn , B G Schunck. Determining optical flow. *Artif. Intel.* , 17 : 185~ 203 , 1981.
  17. [Fogel 1991] S V Fogel. Estimation of velocity vector fields from time-varying image sequences. *CVGIP : Image Understanding* , 53 : 253~287 , 1991.
  18. [Verri 1989] A Verri , T. Poggio. Motion field and optical flow : Qualitative properties. *IEEE Trans. Patt. Anal. Mach. Intel.* , PAMI-11 , 490~498 , 1989.
  19. 韩亚楠,冯驰 基于灰度编码位平面匹配的视频图像稳定技术, *应用科技* , 32(5) , 13-15 , 2005
  20. Sung-Jea Ko; Sung-Hee Lee; Seung-Won Jeon; Eui-Sung Kang; Fast digital image stabilizer based on Gray-coded bit-plane matching , *Consumer Electronics, IEEE Transactions on* , 45(3):598 – 603,1999.
  21. LEE, K.-H., LEE, S.-H., KO, S.-J., “Digital image stabilizing algorithms based on bit-plane matching”. In *IEEE Transactions on Consumer Electronics*, 44(3):617-622,1998.
  22. A. Censi, A. Fusiello, and V. Roberto. Image Stabilization by Features Tracking. Technical report, University of Udine, Machine Vision Laboratory, Dept. of Mathematics and Informatics, 1998.
  23. K. Uomori, A. Morimura, H. Ishii, T. Sakaguchi, and Y. Kitamura. Automatic image stabilizing system by full-digital signal processing. *IEEE Transactions on Consumer Electronics*, 36(3): 510- 519, 1990.
  24. KELLER, Y., AVERBUCH, A. 2003. “Fast Global Motion Estimation or



- MPEG-4 Video Compression”. Packet Video 2003.
25. MORIMOTO, C.H., CHELLAPPA, R., Automatic digital image stabilization. Proceedings of IEEE International Conference on Pattern Recognition, 1996
  26. Takeo Kanade, Hiroshi Kano, Shigeru Kimura, et al, Development of a Video-Rate Stereo Machine", IROS'95, Aug.7-9, 1995, Pittsburgh, PA
  27. Lai-Man Po, Wing-Chung Ma "A Novel Four Step Search Algorithm For Fast Block Motion Estimation" IEEE Transactions on Circuits and Systems for Video Technology,6(3):313-7, June 1996.
  28. M.Ghanbari "The cross search algorithm for Motion Estimation" IEEE Transactions on Communications, 38(7): 950-3, July 1990.
  29. [http://www.ece.cmu.edu/~ee899/project/deepak\\_mid.htm](http://www.ece.cmu.edu/~ee899/project/deepak_mid.htm)
  30. 林福宗, 图像文件格式, 清华大学出版社, 1998, p120-130.
  31. 徐飞,施晓红. MATLAB应用图像处理[M]. 西安:西安电子科技大学出版社, 2002.

## 致 谢

本课题是在朱贻盛和童善保两位导师的悉心指导和殷切关怀下完成的,在此衷心感谢导师对我的谆谆教导,以及对我在生活、工作、学习、研究等各方面的无私帮助。朱老师作为我的科研启蒙老师,引导我步入科学殿堂的大门,并在科研、学习和生活中给予我无微不至的关怀和帮助。朱老师孜孜以求的敬业精神和虚怀若谷、平易近人、宽容大度的人格魅力是我一生的学习楷模。童老师严谨细致、实事求是的治学态度,认真勤奋、不知疲倦的工作作风,以及对事业的执着追求使我终生难忘,并时时鞭策我在科学研究的道路上奋发向上。如果说我从导师那里学会了怎样做好学问,那么首先应该说我从导师那里领略了真正的学术精神,导师严谨的治学态度和坚韧的探索精神将使我终生受益。

衷心地感谢邱意弘老师在学习和生活各方面给予我的帮助。她的热心、严谨的科研思路、实事求是的治学态度给我留下了深刻的印象。

感谢生命科学学院和生物医学工程系的各位老师,在我工作和学习中,时常能感受到他们对我的关心和帮助。

感谢师实验室的全体师兄弟、师姐妹,他们与我探讨了许多问题,提出了许多有益的建议和帮助,同时感谢他们在我学习与生活上的关心与帮助。

衷心地感谢我的家人给予我的支持和关怀。

最后,感谢国家重点基础研究发展规划项目的资金支持。

## 硕士期间发表的论文

1. 石萍, 邱意弘, 朱贻盛, 童善保, 人工视觉假体研究综述(1)——视网膜假体研究现状, 生物医学工程学杂志, 已录用;
2. 石萍, 邱意弘, 朱贻盛, 童善保, 人工视觉假体研究综述(2)——视皮层、视神经束、感觉替代假体的研究现状, 生物医学工程学杂志, 已录用;
3. 石萍, 邱意弘, 朱贻盛, 童善保, 视觉假体系统组成及关键技术, 国际生物医学工程杂志, 已录用。