

学校代码: 10475  
学 号: 104753040717

## 河南大学硕士学位论文

# 一种基于特征匹配的鲁棒性稳像算法

## A Robust Video Stabilization Algorithm based on Feature Matching

专 业 名 称: 应用数学

专 业 代 码: 070104

研 究 方 向: 数字图像处理、Web 服务

年 级: 二〇〇四级

研 究 生 姓 名: 黄亚博

导师姓名、职称: 陈志国 教授、焦建彬 教授

完 成 日 期: 二〇〇七年五月

论 文 主 题 词: 视频稳像/抖动检测/鲁棒性

## 关于学位论文独立完成和内容创新的声明

本人向河南大学提出硕士学位申请。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完成的，对所研究的课题有新的见解。据我所知，除文中特别加以说明、标注和致谢的地方外，论文中不包括其他人已经发表或撰写过的研究成果，也不包括其他人为获得任何教育、科研机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位申请人（学位论文作者）签名：\_\_\_\_\_

20 年 月 日

## 关于学位论文著作权使用授权书

本人经河南大学审核批准授予硕士学位。作为学位论文的作者，本人完全了解并同意河南大学有关保留、使用学位论文的要求，即河南大学有权向国家图书馆、科研信息机构、数据收集机构和本校图书馆等提供学位论文（纸质文本和电子文本）以供公众检索、查阅。本人授权河南大学出于宣扬、展览学校学术发展和进行学术交流等目的，可以采取影印、缩印、扫描和拷贝等复制手段保存、汇编学位论文（纸质文本和电子文本）。

（涉及保密内容的学位论文在解密后适用本授权书）

学位获得者（学位论文作者）签名：\_\_\_\_\_

20 年 月 日

学位论文指导教师签名：\_\_\_\_\_

20 年 月 日

## 摘 要

随着摄像设备应用的日益广泛,摄像平台变得越来越多样。通过车载摄像系统、手持摄影设备、飞机或者舰船摄影平台等得到的视频信号往往会因为摄像机的运动而发生抖动。视频抖动不仅会影响视觉效果,容易使观看者产生晕眩进而导致误判和漏判,并且给进一步的图像分析带来困难。可见,如何将这些视频信号转化为高质量的稳定视频具有重要的应用价值。

通常电子稳像算法主要包括三部分:全局运动估计模块,运动平滑模块和运动补偿模块。其中运动估计模块是通过各种不同的算法对图像进行分析,估算摄像机的全局运动参数;运动平滑模块将全局运动分为主观运动和抖动,符合摄像者主观意愿的为主观运动,需保留;运动补偿模块则是通过运动平滑模块提供的抖动参数,直接利用图像处理的方法使图像上的像素按被检测出的运动偏移反方向做等量移动,实现图像的像素重组,从而达到稳定视频的目的。本文在现有研究的基础上,综合考虑了算法的鲁棒性和智能性两个问题,设计并实现了一种基于特征集合匹配的鲁棒性稳像算法,所完成主要工作如下:

(1)为保证稳像系统的鲁棒性,算法在进行特征匹配时融入亮度变化自适应模型并利用特征匹配误差分析和运动一致性原则对特征初步匹配结果作有效性验证以提高算法对光线变化和局部运动物体的鲁棒性。

(2)为提高稳像系统的智能性,提出一种基于特征集合匹配关系的抖动检测方法。该方法通过对帧间运动参数进行分析确定视频是否有抖动发生,进而确定是否需要做进一步的运动补偿处理,从而避免在视频没有发生抖动时产生由补偿引起的系统效率下降。

(3)在以上研究工作的基础上,成功开发一套数字图像稳定系统,该系统在拍摄场景具有一定的纹理信息时,即使场景中发生光线变化和存在产生局部运动的前景物体时,仍具有较好的稳像效果。

**关键词** 视频稳定; 抖动检测; 鲁棒性; 全局运动估计

---

## Abstract

With the prevalence of camera device, video shot can be easily obtained. However, such video obtained in moving carrier has the shaking problem resulted from camera's shaking. Video shaking not only leads to bad video effect but also leads to false judgment or missing judgment, which are harmful to the next image object trailing. Thus, the require of stabilization of video is more and more important.

The most common video stabilization framework including three modules: global motion estimation, motion filter and motion compensation. The object of global motion estimation module is to calculate the relative motion vector of adjacent frames' background, that is, the global motion vector. Global motion includes two parts, intentional motion and camera shaking. The motion resulting from cameraman's will is intentional motion, and the difference of global motion and intentional motion is camera shaking. For smooth visual effect, motion filter should be introduced to alleviate shaking. The function of motion compensation module is to translate, rotate and scale the video image according to the opposition of the shaking parameters so as to achieve smooth and stable visual effect. In this paper, a robust and intelligent video stabilization algorithm based on feature matching is put forward and realized. The main works in this paper are:

(1)The traditional Shi-Tomasi method is improved by introducing an adaptive illumination independent model to realize robust feature tracking. Then feature matching errors and motion consistence principle are used to improve feature reliability. Experiment results show that the proposed method is robust to both moving object and illumination change.

(2)The correspondence between matched features is used to shaking detection.

(3)We developed an automatic video stabilization system based on our proposed techniques. The relative motion vector of adjacent frames' background is used to make sure whether shaking is exist. If shaking does not exist then motion compensation is not need. So, system efficiency reduce by motion compensation is avoid.

**Key words:** video stabilization; shaking detection; robustness; global motion estimate

---

# 目 录

摘 要 .....	I
Abstract .....	II
第 1 章 绪 论 .....	1
1.1 视频稳定的研究背景及意义 .....	1
1.2 国内外发展概况 .....	3
1.2.1 全局运动估计 .....	4
1.2.2 运动平滑 .....	5
1.2.3 运动补偿 .....	5
1.3 课题的来源和内容的组织 .....	7
1.3.1 课题的来源 .....	7
1.3.2 本文研究的主要内容和意义 .....	7
1.3.3 论文组织结构 .....	8
第 2 章 全局运动估计 .....	10
2.1 摄像机的运动与图像运动的关系 .....	10
2.1.1 摄像机运动与图像运动 .....	10
2.1.2 图像运动模型 .....	12
2.2 全局运动估计的基本方法 .....	14
2.2.1 块匹配法 .....	14
2.2.2 灰度投影算法 .....	16
2.2.3 特征匹配算法 .....	18
2.3 由粗到精的运动估计思想 .....	19
第 3 章 基于特征匹配的全局运动估计 .....	21
3.1 基于特征匹配的算法 .....	21
3.1.1 特征提取 .....	21
3.1.2 特征匹配 .....	24
3.1.3 存在问题 .....	26
3.2 鲁棒性特征匹配算法 .....	26
3.2.1 亮度自适应特征匹配 .....	27
3.2.2 特征有效性判定 .....	27

---

3.3 全局运动参数求解 .....	29
3.3.1 相邻帧运动参数求解 .....	30
3.3.2 获取全局运动参数 .....	31
3.4 抖动检测 .....	32
3.5 小结 .....	34
第 4 章 运动平滑与运动补偿 .....	35
4.1 运动平滑 .....	35
4.2 运动补偿 .....	36
4.2.1 抖动补偿 .....	36
4.2.2 无定义区域重构 .....	38
4.3 小结 .....	39
第 5 章 稳像算法实现及实验结果分析 .....	40
5.1 稳像算法验证平台 .....	40
5.2 稳像算法流程 .....	41
5.3 稳像算法评估标准 .....	42
5.3.1 精度评估 .....	43
5.3.2 实时性评估 .....	44
5.3.3 可处理范围评估 .....	44
5.4 实验结果 .....	45
5.4.1 亮度无关特征匹配 .....	45
5.4.2 特征有效性判定 .....	45
5.4.3 随机抖动检测 .....	46
5.4.4 视频稳定 .....	46
5.5 小结 .....	53
总结与展望 .....	54
致 谢 .....	56
参 考 文 献 .....	57
攻读学位期间发表的论文和参加的项目 .....	61

---

# 第 1 章 绪 论

## 1.1 视频稳定的研究背景及意义

随着摄像设备应用的日益广泛, 摄像平台变得越来越多样化。通过车载摄像系统、手持摄影设备、飞机或者舰船摄影平台等得到的视频信号往往会因为摄像机的运动而发生抖动。视频抖动不仅会影响视觉效果, 容易使观看者产生晕眩进而导致误判和漏判, 并且给进一步的图像处理带来困难。如何将这些视频信号转化为高质量的稳定视频具有重要的理论意义和实用价值。

通过机械或者电子设备, 获得运动视频的帧间运动, 并对其进行补偿, 以获得平滑稳定的视频图像的操作称为视频稳定(简称稳像)。稳像的目的就是通过消除或减轻这种抖动, 获得高质量的稳定视频。视频稳定技术无论在军事领域还是民用领域, 都具有非常广泛的应用需求。

在军事上, 若摄像设备位于运动载体上(如战场上行进中的坦克, 风浪中的轮船, 颠簸中的飞机), 视频图像会随载体运动而剧烈抖动, 使用户无法对目标进行观察和瞄准。出现这种情况时, 通过稳像技术, 可以提高军队的作战能力。

在民用上, 随着家用移动数码摄像设备的应用日益广泛, 手或者其他运动载体拍摄的视频往往带有抖晃, 稳像技术在这里拥有乐观的商业价值。还如在航空摄影和地形测绘仪器中, 为在仪器像平面上得到稳定的测量基准必须采用稳像技术, 以便在仪器的测量面上提供一个相对稳定的坐标系, 使测量结果准确无误<sup>[1]</sup>。

因此对于运动载体上的摄像系统, 稳像是一个十分重要的问题。目前常见的稳像方法按照作用机制大致可分为光学稳像、机械式稳像和电子稳像。

### 1. 光学稳像

光学稳像是指用光学系统中少数元件的补偿运动来实现视频的稳定。通过在光路中设置一些光学元件来对不稳定图像进行补偿。光学稳像方法无需对整个仪器或光学系统进行稳定, 因此重量较轻, 同时物镜和像面均是固定的, 便于对图像进一步处理。光学稳像在军事、医疗仪器等方面均有相对广泛的应用, 但是系统制造工艺复杂, 成本高昂<sup>[2]</sup>。

### 2. 机械式稳像

传统的机械式稳像是利用陀螺等传感器和伺服系统构成的稳定平台来补偿基座上的摄像系统的相对运动来实现稳像的<sup>[3]</sup>。在一些大的系统中, 如武器系统中, 往往将整个成像设备置于稳定平台上。而在一些小型的设备中, 如导弹导引头或

电视摄像头中，则只是将光学系统稳定起来。平台稳定方式是通过惯性元件来获取载体的姿态角的变化，其输出信号经过放大后，驱动电机或压电陶瓷来保持摄像机或反射镜、棱镜等的稳定，以便确保输出的图像序列的平稳。

机械式稳像方法目前主要应用于车载、机载和舰载等摄像设备上。但是，由于这类方法存在着价格昂贵、结构复杂、体积庞大、功耗巨大等问题，因此，其应用在某些特定的场合中受到一定的限制，如弹载、轻型飞机、外星球探测、家用 DV 等。

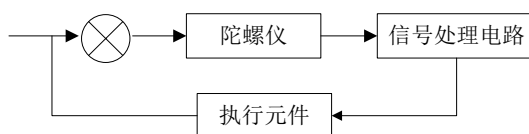


图1-1 传统的机械稳像装置

### 3. 电子稳像

随着传感技术、计算机技术的发展，稳像系统已经逐步成为应用光、机、电、算的综合性系统。应用计算机数字图像处理和电子技术相结合的方法来直接确定图像序列的帧间偏移并进行补偿的技术叫做电子稳像技术。电子稳像的前提条件是先获得图像的全局运动参数，根据获取图像运动参数方法的不同，主要分为硬件实现方法和软件实现方法<sup>[1]</sup>。

**硬件实现方法**是利用传感器检测摄像机的运动矢量。这种方法首先采用高精度的陀螺作传感器，获得图像的帧间运动参数，而后利用数字图像处理技术对图像进行运动补偿，消除因运动载体抖动对视频图像带来的不利影响(如来回摆动和歪斜等)，得到稳定的视频输出。例如 Oshima 等为家用摄录机设计的一种电子稳像系统即采用该技术。此方法易于实现，但是一方面其检测精度要依赖于高精度陀螺来达到，故对传感器的精度要求很高；另一方面由于检测到的运动量必须及时地转换成图像帧间运动参数，需要已知设备的许多参数，如摄像机镜头的焦距值等，还要增加一些其它辅助设备，因此系统的成本和重量仍然较高。

**软件实现方法**是通过稳像算法对视频序列进行分析处理，该方法直接利用图像信息确定摄像机的运动参数，然后对当前帧图像进行补偿。这种方法本质上是数字化的电子稳像。

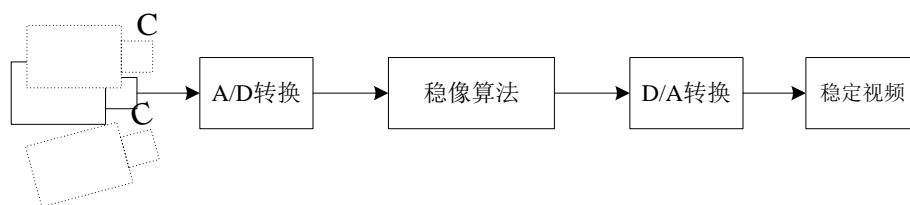


图1-2 数字稳像系统示意图



通过对各种稳像方法进行性能比较可知(如表 1-1 所示),光学稳像系统的功耗几乎为零、体积较小且稳像精度高,但是光学设备昂贵、使用寿命短、维护复杂等特点也限制了它的应用范围;机械式稳像方法受到控制方式、维护方式、成本等因素的影响,发展受到极大的制约,且由于其体积庞大、功耗很高,通常架设在对体积要求不高的大型设备上,发展前景不甚乐观;对于数字稳像技术来说:从自身特点分析,它与传统的光学稳像、机械式稳像方法相比,具有易于操作、便于集成、价格低、功耗小等特点;从外部条件考虑,手持数码摄像设备的普及,为数字稳像技术带来的巨大的市场需求;数字图像处理技术的发展,为它提供了有利的技术支持;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。因此,从整个稳像技术研究的历史来看,利用纯数字图像处理的方法进行视频稳定的研究是当前稳像技术研究发展的趋势。目前,人们的研究重点正渐渐从光学稳像、机械式稳像向数字稳像转移。

表1-1 各种稳像方法性能对比

稳像类别	稳像设备	体积	功耗	稳像精度	成本	使用场合
机械式稳像	机械(如陀螺等),机电结合	大	高	中	高	对成本参数要求不高的大型军用设备中
光学稳像	光学器件	中	低	高	高	小型军用设备,医疗器械
数字稳像	稳像算法,电子元器件	小	低	中	低	应用场合广泛,特别是对成本和体积参数要求较高的手持移动摄像设备中

## 1.2 国内外发展概况

视频稳像技术是国内外许多大学和研究所的热门课题之一。国外的数字稳像技术发展较早,目前已有二十余年的历史,九十年代初,随着图像处理技术和计算机技术的迅猛发展,以美国和加拿大为首的西方国家率先采用稳定算法和图像重组的方法实现视频稳定,使稳像系统向小型、实时和高精度的方向发展。这些成果主要应用在侦察车的侦察系统、目标跟踪系统和无人驾驶越野车中的导航系统等方面。其中,加拿大 DREV(Defense Research Establishment Valcartier)研究机构根据国防要求,成功研制了安装在侦察车 10 米高桅杆上的实时监控系统中的稳像

装置，其图像处理的精度达到像素级。而由美国陆军实验室(Army Research Laboratory)研制的应用在无人驾驶越野车上的稳像系统，其稳像精度达亚像素级<sup>[4]</sup>。

国内对于数字稳像技术的研究起步较晚，但随着稳像技术的需求日益广泛，近年来不少研究者致力于该技术的研究。中科院长春光机所<sup>[3,17,20,30]</sup>、清华大学<sup>[5,8]</sup>、北京理工大学<sup>[6]</sup>、国防科技大学<sup>[7]</sup>等都开始了对数字稳像技术的研究，并获得了阶段性的成果。其中清华大学研究人员提出了基于 2.5 维帧间运动估计和惯性运动模型并同运动滤波器方案相结合的视频稳定方法，具有较好的稳定效果<sup>[5]</sup>。

传统的稳像系统一般由全局运动估计(Global Motion Estimation)和运动补偿(Motion Compensation)两个功能模块组成。其中运动估计模块是通过各种不同的算法对图像进行分析，估算摄像机的全局运动参数；运动补偿模块则是通过运动估计模块提供的运动参数，直接利用图像处理的方法使图像上的像素按被检测出的运动偏移反方向做等量移动，实现图像的像素重组，从而达到稳定视频的目的。采用这种稳像流程，只能对定点拍摄获得的视频序列进行处理，而无法处理包含摄像机的主观运动(如水平、垂直方向的扫描运动)的视频序列。如果简单的利用包含扫描信息的全局运动参数进行补偿，会导致有用的视频信息损失严重。为了在保留主观运动的基础上去除视频抖动，近年来，一些研究者在稳像框架中加入运动平滑(Motion Smoothing)模块，如图 1-3 所示。下面按照这三个步骤，简单介绍稳像算法的研究现状及存在问题。

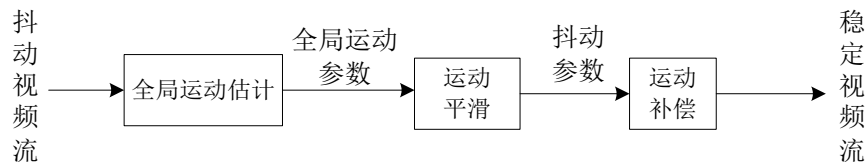


图1-3 稳像算法的基本框架

### 1.2.1 全局运动估计

稳像算法的关键问题是通过全局运动估计获得精准的全局运动参数。现有文献中，稳像的全局运动估计方法大致可分为三类：块匹配(BMA, Block Matching Algorithm)<sup>[10-16]</sup>、灰度投影(PA, Projection Algorithm)<sup>[17-19, 30]</sup>和特征匹配法(FMA, Features Matching Algorithm)<sup>[20-30]</sup>。

块匹配方法虽然由于没有信息损失，因此具有较好的鲁棒性，然而它却存在计算量大、耗时长等缺点。针对块匹配算法存在的问题，改进算法围绕着改进搜索策略和优化图像两方面提出改进方案。

灰度投影法通过信息约减降低了计算量，但是其算法特点决定了仅仅能够在图像的同一个小行、列上进行运动计算，亦即仅能处理平移问题，对于旋转运动无效。文献[19]提出将灰度投影与块匹配方法相结合的思想，相应文献[30]考虑了将灰度投影与特征匹配算法相结合的运动估计方法。

基于特征匹配的算法同时具有减少运算量和可处理旋转运动的优点，近年来受到了研究者的重视。但采用特征匹配方法时，会面临一个严重问题：真实场景中，外界环境往往很复杂，小物体的移动，树枝叶的摇动，以及外界光线突变、异物进入视场等都会对全局运动参数的估算产生不利影响。如何在尽量不增加算法复杂性的基础上进一步提高算法精度及鲁棒性，是目前基于特征的运动估计方法的研究热点问题之一。

### 1.2.2 运动平滑

如上所述，直接利用全局运动估计的结果进行运动补偿，会把摄像机的主观运动误判为抖动，不能达到全景扫描的效果。可见，如何在保留摄像机正常运动的同时去除不必要的抖动是稳像算法需要解决的问题之一。近年来，研究者提出采用平滑滤波或曲线拟和的方法获取抖动运动参数。Sung-Jea Ko 等人认为主观运动是当前帧以及其前一帧的全局运动的组合，为了区分出摄像者有意识的运动，他们采用了一个阻尼系数，认为当前帧的主观运动是前一帧的主观运动阻尼后和抖动参数的线性组合。这样当摄像机静止时，主观运动也趋于 0。其中阻尼系数的大小直接决定了对抖动的抑止程度，因此不同幅度的晃动需要不同的系数<sup>[31]</sup>。Yan 等人采用 Bezier 曲线拟合的方法，对运动轨迹进行平滑，该方法可以处理剧烈抖动，但是也存在过度平滑造成的视频失真现象<sup>[32]</sup>。其他很多研究者采用均值滤波器提取平滑运动参数<sup>[3,13]</sup>。

### 1.2.3 运动补偿

运动补偿是在完成运动平滑后，根据抖动参数，将图像校正到正确的位置以获得稳定的视频序列。它是通过抖动参数直接利用数字图像处理的方法，使图像中的像素按被检测出的抖动偏移反方向做等量移动，使帧序列位于同一坐标空间下。

然而在进行校正变换后，由于变换后的坐标空间和原图像的坐标空间不重合，若初步稳定后图像的某些像素点不在运算结果范围内时，这些像素点的灰度值为“空”，因此容易在变换后的图像的边界形成无定义区域，造成部分图像信息丢失，

同时又会影响人的视觉效果（如图 1-4-b 所示黑色区域）。针对稳像时产生的“空”区域，目前常见的处理方法大致有三种。



(a) 稳定前视频序列



(b) 初步稳定后视频序列

图 1-4 空白区域处理示意图

1. 对初步稳定后的图像进行边缘剪裁，而后放大成显示器输出大小。该方法会降低图像的清晰度，同时还会影响图像可处理的抖动幅度。当抖动幅度增大时所需要的剪切和放大也随之增大，因此，存在分辨率过低问题。

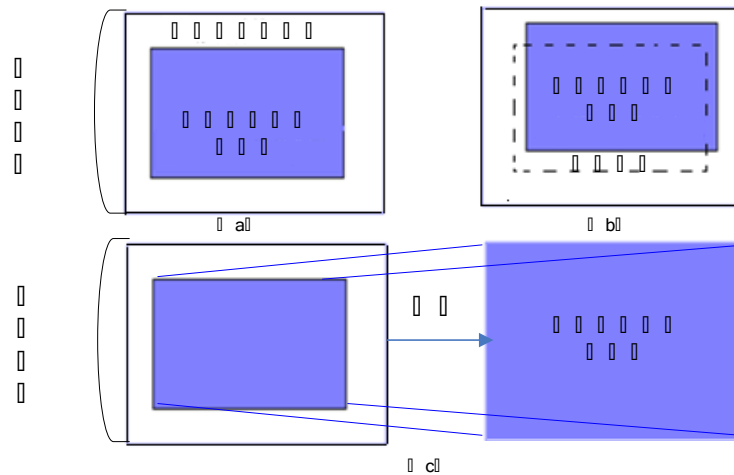


图 1-5 预留法空白区域处理示意图<sup>[3]</sup>

2. 文献[3]中预留一部分区域作为补偿区域。该方法认为，当成像系统晃动时，让成像区域大于输出到监视器上的信息区域，那么即使图像有晃动，根据图像晃动的参数来调整信息区域的位置，可保证有用的图像信息仍能在监视器上显示，

如图 1-5 所示。

3. 利用相邻帧图像信息填补图像边缘的缺失信息<sup>[9,23]</sup>。如图 1-6 所示, 该方法在拍摄场景边缘不存在发生局部运动的物体时, 有较好的拼接效果, 但是当存在局部运动物体时, 会使得拼接边缘无法吻合, 影响视觉效果。

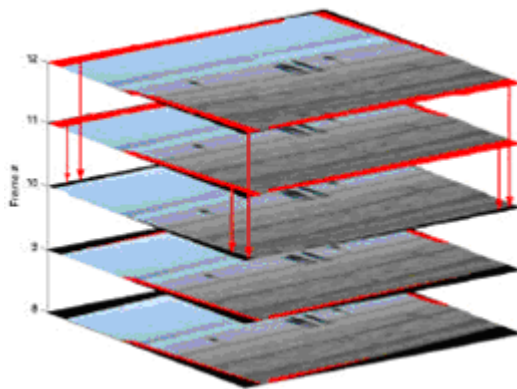


图 1-6 帧间相关法空白区域处理示意图<sup>[23]</sup>

## 1.3 课题的来源和内容的组织

### 1.3.1 课题的来源

本课题来源于国家自然科学基金项目“复杂环境下新型多目标动态检测、识别及跟踪算法研究”中的内容, 并得到中国科学院百人计划资金的支持。该项目主要利用多路视频图像实现 360 度全景下的动态目标检测和跟踪。考虑到实际应用中运动载体的抖动会造成视频序列的不稳定, 影响跟踪算法的正常运行。为扩大跟踪系统的应用场合, 需利用稳像技术减少视频序列的不稳定对跟踪算法的影响。作者以此为初衷开始对数字稳像系统的研究, 并将其扩展到包含扫描运动的非定点拍摄场景中。

### 1.3.2 本文研究的主要内容和意义

近年来, 虽然国内外许多大学和研究机构在稳像领域的研究工作已经取得较大进展, 但是目前大多数研究还处于实验室阶段, 仍有许多问题亟待解决。进一步开展稳像问题的研究具有潜在的经济价值和广泛的应用前景。因此, 我们亟需在此领域开展深入的研究。目前稳像算法的发展趋向主要有以下几个方面:

1. 全局运动估计的精度与速度的统一

全局运动估计的精度和速度，基本决定了稳像算法的精度和速度。而精度和速度总是一个对立问题，高的精度一般需要花费更多的处理时间。如何在精度和速度之间达到了一个较好的平衡是稳像算法需要关注的问题之一。

### 2. 提高复杂环境下算法的鲁棒性

由于外界光线变化、噪声点、和运动的前景物体等不利因素会对全局运动估计的精度产生影响，进而影响算法有效性，因此，提高复杂环境下算法的鲁棒性也是稳像算法需要关注的问题之一。

### 3. 提高系统的智能化

现有算法较少考虑系统的智能化问题，比如目前只有极少数算法涉及抖动检测过程，而其他算法基本采用直接补偿或手动触发方式。直接补偿在视频稳定时存在因资源浪费而造成的效率下降，而手动触发补偿给用户操作带来不便。以家庭摄像为例，由于摄像者的非专业操作，视频中的某些片断可能是抖动的，但是，一般来讲，视频中的大部分片断可能是稳定的。对于这类实际生活中的视频，我们不可能也没有必要对视频从头至尾地进行稳定化处理，因为这不仅会降低处理的效率，甚至有可能破坏原有视频中质量较好的片段(稳定片断)的视觉质量。因此，为了实现智能化的视频稳定化处理，就需要能够自动地检测出视频中的抖动，并触发稳像。再者，由于摄像机抖动本身具有随机性，一般需要针对不同类型的抖动视频(如抖动频率、幅度不同)，由用户手动设定一些参数，进行针对性的稳定化处理。以上这些都给用户的操作带来不便，降低了算法的实用性。因此，需要在保证算法正确运行的基础上，进一步提高算法的智能化。

在本文中，作者在综合考虑鲁棒性和智能化两个问题的基础上，设计出一种适用于移动摄像设备获取的视频序列的基于特征匹配的鲁棒性视频稳像算法。该算法在运动估计部分将亮度自适应模型融入特征匹配算法中，而后利用特征误差分析和运动一致性原则对特征集合进行有效性验证，通过以上方法提高特征匹配的鲁棒性。同时为了避免视频无抖动时的稳像误差，方便用户操作，增加系统的智能化，算法在运动估计部分加入抖动检测功能，提出一种基于运动往复特性的抖动检测方法。最后，给出去除抖动和处理“空”区域的实现方法，最终获得稳定的视频序列。

## 1.3.3 论文组织结构

本文主要工作是设计并实现一种基于特征匹配的鲁棒性视频稳像算法。全文内容安排如下：

第一章：绪论。介绍数字稳像技术的研究背景和意义以及稳像技术的主要研

---

究现状和存在的问题，最后给出本文的课题来源及所研究的主要内容。

第二章：全局运动估计。首先在分析摄像机的运动与图像运动关系的基础上介绍了几种常见的运动模型，然后对计算全局运动参数的几种算法原理进行分析阐述，最后简单介绍了图像金字塔在算法中的应用。

第三章：基于特征匹配的全局运动估计。对基于特征匹配的运动估计方法做了详细介绍，并且针对现有方法存在的鲁棒性较差的问题进行改进。最后为了提高稳像算法的智能化，提出了基于往复特性的抖动检测方法。

第四章：运动平滑与运动补偿。给出利用运动平滑获取抖动参数，并进行视频稳定化的过程。

第五章：稳像算法实现及实验结果分析。对实际拍摄的视频图像序列进行稳像实验，并用通用的评定方法对稳像结果进行评定，实验结果表明了算法的有效性。

总结与展望：全文的总结。对本文的主要研究工作进行了简要的阐述，并探讨和展望了在未来时间内应当完善的问题。

---

## 第 2 章 全局运动估计

全局运动是指在视频序列中占有较大比例的像素运动，是由摄像机的运动造成的。通常视频图像由前景和背景构成，如果在拍摄过程中摄像机是运动的，同时前景物体也有其自身的运动，那么在视频序列中所表现出来的就是背景、前景都有各自的运动：背景的运动是由摄像机运动造成的，即全局运动；而前景物体所表现出来的运动是前景物体相对于摄像机的运动，称为局部运动。全局与局部运动最后都反映在视频图像信息上。稳像算法的关键问题就是利用帧间图像的差别获得全局运动参数，可以说全局运动估计模块的优劣决定了整个稳像系统的运行效果。

在进行全局运动估计时，首先需要建立描述摄像机运动的参数模型。本章在分析摄像机的运动与图像运动关系的基础上介绍了几种常见的参数模型，而后对当前主流的估计运动参数的方法进行比较分析。最后，为保证稳像算法的效率和精度，本文利用图像金字塔分层参数估计的思想，由粗到精细化运动参数。

### 2.1 摄像机的运动与图像运动的关系

#### 2.1.1 摄像机运动与图像运动

假定任意一种摄像机的运动形式都可以用旋转、平移和变焦表示。其中，摄像机变焦运动包括载体向前、向后行进造成的景深变化和摄像机参数改变(如推镜头、拉镜头)而引起的视频变化。由此，摄像机的运动可用式(2-1)描述，设 $(X, Y, Z)^T$ 是摄像机坐标系下景物上的一点， $(X, Y, Z)^T$ ， $(X', Y', Z')^T$ 分别为一点在不同时刻的三维坐标，则 $(X, Y, Z)^T$ 和 $(X', Y', Z')^T$ 的关系为：

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \quad (2-1)$$

摄像机的成像系统将三维场景变换成二维灰度或者彩色图像，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：

$$\begin{aligned} f: R^3 &\longrightarrow R^2 \\ (X, Y, Z)^T &\longrightarrow (x, y)^T \end{aligned} \quad (2-2)$$



其中 $(X, Y, Z)^T$ 是摄像机坐标系下景物上的一点， $p=(x, y)^T$ 为该点在二维图像坐标系下的位置。利用透视投影法或者正交投影法将三维空间映射到二维空间，我们就可以获得摄像机坐标系与图像坐标系的对应关系<sup>[35]</sup>。

### 1. 透视投影

透视投影是最常用的成像模型，可以用针孔成像模型来近似表示。根据透视投影原理，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与图像平面(成像平面)的直线，点 $(X, Y, Z)^T$ 的视线、点 $(X, Y, Z)^T$ 到 $Z$ 轴的垂直线段以及 $Z$ 轴形成了两个三角形，如图 2-1 中虚线所示。这两个三角形是相似三角形，由此可以得到透视方程：

$$\frac{x}{X} = \frac{y}{Y} = \frac{f}{Z} \quad (2-3)$$

则有摄像机坐标系下景物上的一点 $(X, Y, Z)^T$ 与图像平面中对应点 $p$ 的位置关系：

$$p = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f \frac{X}{Z} \\ f \frac{Y}{Z} \end{pmatrix} \quad (2-4)$$

其中 $f$ 为像机焦距。

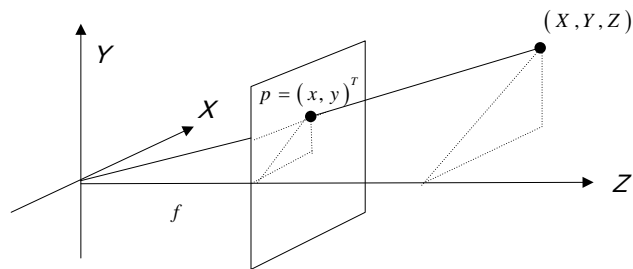


图 2-1 透视投影几何示意图

### 2. 正交投影

正交投影是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，如图 2-2 所示。正交投影是透视投影的一个特例，对透视投影进行简化，即假设透视模型的焦距 $f$ 很大且物体距投影中心很远时，透视投影可以用正交投影来近似。正交投影方程为：

$$x = X \quad y = Y \quad (2-5)$$

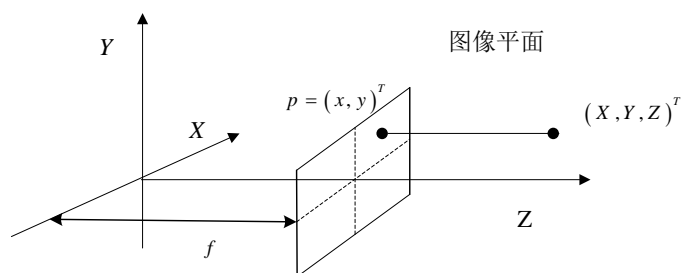


图 2-2 正交投影几何示意图

由以上分析可知, 图像的二维运动(也称为投影运动)反应了摄像机的三维运动信息, 它是三维运动在图像平面上的透视或正交投影。由此给出八参数的透视模型:

$$\begin{aligned} x' &= \frac{a_1x + a_2y + a_3}{a_7x + a_8y + 1} \\ y' &= \frac{a_4x + a_5y + a_6}{a_7x + a_8y + 1} \end{aligned} \quad (2-6)$$

其中 $(x, y)^T$ 和 $(x', y')^T$ 为点 $p$ 在不同图像的对应位置坐标。利用正交投影关系式(2-5)和摄像机的运动式(2-1)可推导出图像平面运动模型中的六参数仿射模型:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \quad (2-7)$$

### 2.1.2 图像运动模型

由于图像的二维运动反应了摄像机的三维运动信息, 在进行运动估计时, 首先要选取合适的图像运动模型来描述摄像机的运动信息。常见的三种图像运动模型为: 平移模型, Similarity 模型和 Affine 变换模型。

#### 1. 平移模型

图像的平移运动是指图像仅在两维空间中发生了 $x$ 方向和 $y$ 方向的位移, 如图 2-3 所示。如果摄像机仅发生了平移运动, 则可以采用平移模型:

$$p' = W(p, d) = p + D = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \quad (2-8)$$

其中 $p = (x, y)^T$ ,  $p' = W(p, d) = (x', y')^T$ , 分别表示某个像素点在参考帧和当前第 $k$ 帧的坐标,  $d_1$ 和 $d_2$ 分别为 $p'$ 相对于 $p$ 在 $x$ 轴和 $y$ 轴的偏移量。

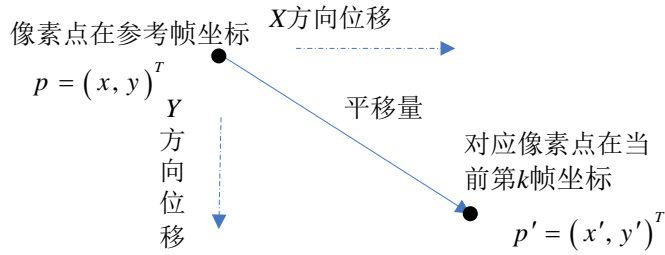


图 2-3 平移运动模型

### 2. Similarity模型

摄像机本身除平移运动外还可能发生旋转运动，例如：在车载摄像系统中，摄像机固定在车体上，在车辆行驶过程中，因为路面不平坦，车辆可能发生左右摇摆，从而使图像旋转。目前一般对绕光轴或者平行光轴的旋转运动进行研究。同时，在存在场景的缩放时，利用缩放因子  $s$  对变焦运动进行描述。由此，当图像可能发生平移、旋转、变焦运动时，可采用 Similarity 运动模型。

$$p' = W(p, D, \theta) = s \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} p + \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \quad (2-9)$$

其中：  $d_1, d_2, p', p$  含义同上，  $\theta$  为旋转角度，  $s$  为缩放因子。

### 3. Affine变换模型

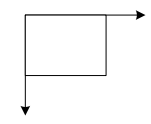
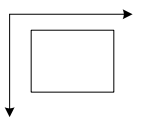
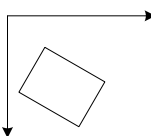
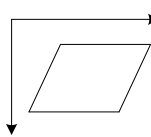
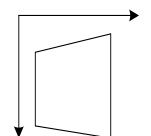
Affine 模型是一种六参数线性变换模型，即具有平行线变换成平行线，有限点映射到有限点的一般特性。具体表现可以是各个方向尺度变换系数一致的均匀尺度变换或变换系数不一致的非均匀尺度变换及剪切变换等，可描述平移运动，旋转运动及小范围的缩放和形变。其过程可以描述如下：

$$p' = W(p, A, D) = Ap + D = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} p + \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \quad (2-10)$$

其中  $d_1, d_2, p', p$  含义同上，  $a_1 \sim a_4$  是旋转参数，也包含了小范围的缩放。

总的来说运动模型的参数个数越多，对摄像机运动的描述更加精确，相应计算复杂性越高，如表 2-1 所示。由于帧间图像变化较小，在精度要求不高时可近似采用平移模型，该模型计算复杂性低，但是在对连续帧间的摄像机运动参数进行叠加时，所造成的累计误差也更严重，非相邻帧的摄像机运动不能简单的用平移模型描述。八参数透视变换模型尽管能更精确的描述摄像机的运动，但它的计算复杂度更大。综合考虑精度与实时性的要求，本文采用六参数仿射模型对造成帧间场景变化的摄像机运动建模。当场景相对深度变化不大时，六参数仿射模型能够很好的描述摄像机的旋转、摇镜头和平移运动。

表 2-1 图像运动模型

运动模型	坐标变换关系	变换参数	图示
原始图	---	---	
平 移	$p' = p + D = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$	$d_1 \sim d_2$ : 平移	
Similarity	$p' = s \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} p + \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$	$s$ : 变焦 $\theta$ : 旋转 $d_1 \sim d_2$ : 平移	
Affine	$p' = Ap + D = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} p + \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$	$a_1 \sim a_4$ : 旋转, 小范围缩放和形变。 $d_1 \sim d_2$ : 平移	
透 视	$x' = \frac{a_1x + a_2y + a_3}{a_7x + a_8y + 1}$ $y' = \frac{a_4x + a_5y + a_6}{a_7x + a_8y + 1}$	$a_1 \sim a_8$	

## 2.2 全局运动估计的基本方法

稳像算法中的运动估计模块是通过各种不同的算法对图像进行分析, 估算出图像序列的运动偏移。它占去了稳像系统的主要工作时间, 是稳像技术的重要环节。近年来许多研究者对全局运动估计算法展开了深入的研究, 提出了多种方法, 但总的来说主要分为三类: 块匹配法、投影算法、特征量匹配算法。下面对各个方法所具有的不同的优势和特点进行分析。

### 2.2.1 块匹配法

#### 1. 匹配思想

将图像帧为划分大小相同的一系列子块，每个子块大小为 $N \times N$ 。对于第 $k-1$ 帧中的一个子块，在第 $k$ 帧中寻找与其大小相同的最相似子块，这个过程称为寻找匹配块。图2-4给出了待匹配块与搜索区的几何位置关系。

该算法假定子块内的每个像素都具有同一个位移值，假定运动物体在相邻帧时间内水平和垂直方向的最大位移都为 $M$ ，则搜索区面积为 $(N+2M+1) \times (N+2M+1)$ 。如果图像变化不大，为了提高搜索速度，可以用小的搜索窗口；反之，如果处理的序列有很多的激烈场面，帧与帧之间的变化很大，则必须用大的窗口进行搜索。

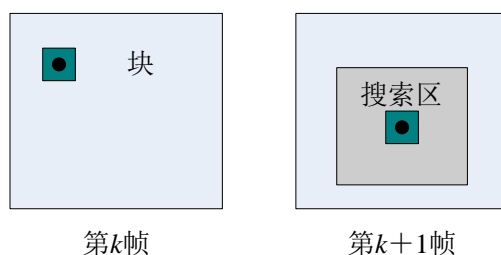


图2-4 块匹配运动估计法示意图

## 2. 匹配准则

块匹配可依据各种准则进行。常用的有最小均方误差准则、最小绝对差准则和最大匹配像素统计准则等<sup>[3]</sup>。

**最小均方误差准则：**

$$MSE(d) = \frac{1}{N \times N} \sum_{x=1}^N \sum_{y=1}^N [I_k(x, y, d) - I_t(x, y)]^2 \quad (2-11)$$

其中， $I_k(x, y)$ 表示第 $k$ 帧图像中点 $(x, y)$ 处的灰度值， $I_t(x, y)$ 表示参考图像在点 $(x, y)$ 处的灰度值， $d=(d_1, d_2)$ 为平移参数，运动参数估计结果为 $MSE(d)$ 最小时的 $d$ 值。

**最小绝对差值准则：**

$$MAD(d) = \frac{1}{N \times N} \sum_{x=1}^N \sum_{y=1}^N |I_k(x, y, d) - I_t(x, y)| \quad (2-12)$$

各参数意义同(2-11)。运动矢量估计为 $MAD(d)$ 最小时的 $d$ 值。

**最大匹配像素统计准则：**

图像匹配实际上是像素点的匹配，所以，如果能在一定的匹配误差范围内，检测图像之间的匹配像素个数，也可描述图像的匹配程度。定义匹配像素个数如下：

$$Q(x, y, d) = \begin{cases} 1 & |I_k(x, y, d) - I_t(x, y)| < T \\ 0 & \text{其他} \end{cases} \quad (2-13)$$

其中 $T$ 为阈值,当像素点的匹配误差小于阈值 $T$ 时,表示该点满足匹配条件, $Q(x, y, d)$ 值为1,否则为0。 $MPC(d)=\sum Q(x, y, d)$ 表示满足匹配条件的像素个数。最大匹配像素统计准则求取匹配像素累加个数最大时的 $d$ 值。

由于匹配准则对匹配的精度影响不大,所以上面不含乘除法的最小绝对差准则成为最常用的匹配准则。

基本的块匹配思想对搜索区内的所有子块进行搜索,因此能够得到搜索区内与当前块最为匹配的块,但是由于算法需要对搜索窗口内每个块进行匹配,因此运算量大、效率低。针对基本块匹配方法计算较复杂等问题,近年来国内外同行提出很多种改进方法,总体来说从优化图像、改进搜索策略两个方面对块匹配方法进行改进。

### 优化图像

在优化图像方面是通过直接减少图像中不影响匹配精度的信息量来降低计算复杂性,如代表点匹配法,位平面匹配算法<sup>[31,34]</sup>等。其中,代表点匹配法通常把整个图像分成若干区域,每个区域有若干个代表点,在进行匹配时,仅考虑代表点的信息,该方法的优点是算法简单、计算量小,但是由于代表点是确定的,并非是图像上真正有明显特征的点,有时存在对图像的变化不敏感的问题。位平面匹配算法是由Sung-Jea Ko等提出的,其思想是利用位平面图像来代替八位的灰度图像来实现图像块的快速匹配。利用该算法图形块的匹配仅用二进制函数就能实现,因此运算量和计算复杂程度能够大大的降低,但是如何进一步提高匹配精度仍需探讨<sup>[31]</sup>。

### 改进搜索策略

改进搜索策略法通过减少搜索次数来降低计算复杂性。由于直接对搜索策略进行优化不需要增加额外的运算,所以该方法是目前有关块匹配中涉及较多的部分。改进的搜索算法相对于全搜索算法,其运算量级较低,是实现高效运动估计的有利手段。针对这一点,在早期,三步法等很多快速搜索算法被提出,以提高运动估计的速度。而后来提出的新三步法<sup>[12]</sup>、四步查找算法<sup>[15]</sup>和自适应块匹配算法<sup>[16]</sup>利用运动矢量的中心偏置分布特性来设计搜索样式,提高了快速运动估计算法的速度和精度。两步搜索法<sup>[33]</sup>利用多候选值搜索策略解决陷入局部最小值的问题。

## 2.2.2 灰度投影算法

美国ARL(Army Research Laboratory)采用此方法在Demo 1号自控目标跟踪系统中实现图像的实时稳定<sup>[17]</sup>。该算法利用图像的行列灰度投影曲线做一次相关运算

获取图像的运动矢量。算法主要由图像映射，投影滤波和相关运算三部分组成。

### 1. 图像映射

对每一帧输入的二维图像使用直方图均衡化进行预处理后，映射成两个独立的一维波形。下式是以列投影计算为例说明图像映射的方法。

$$Col_k(j) = \sum_i Cur_k(i, j) \quad (2-14)$$

其中，式 $Col_k(j)$ 为第 $k$ 帧图像第 $j$ 列的灰度值， $Cur_k(i, j)$ 是第 $k$ 帧图像上 $(i, j)$ 位置处的像素的灰度值。

### 2. 投影滤波

通过对投影值进行滤波可提高算法精度，这是因为图像移动量大时，边缘信息在每一幅图像上是唯一的，因此边缘信息在互相关计算时会对互相关的峰值产生不利的影响。为了解决这个问题要将边界处的投影值去除，方法是通过余弦滤波器进行滤波，此滤波方法可保留中间区域的幅值，降低边界信息的幅值。

### 3. 相关计算

将第 $k$ 帧图像投影波形与参考图像的投影波形做互相关计算。参考图像可以是前一帧图像，也可能是存储器中存储的标准图像。根据两条曲线的波谷值即可确定当前帧图像相对于参考帧图像的行、列位移矢量值。下式为进行列相关运算的计算公式：

$$C(w) = \sum_{j=1}^N [Col_k(j+w) - Col_r(M+j)]^2 \quad 1 \leq w \leq 2M+1 \quad (2-15)$$

式中 $Col_k(j)$ ， $Col_r(j)$ 分别为第 $k$ 帧图像和参考帧图像的第 $j$ 列的灰度投影值， $N$ 为列的长度， $M$ 为位移矢量相对于参考帧在一侧的搜索宽度。设 $W_{min}$ 为 $C(w)$ 最小时的 $w$ 的值，则第 $k$ 帧图像相对于参考帧图像在垂直方向的位移矢量为：

$$\delta_c = M + 1 - W_{min} \quad (2-16)$$

灰度投影算法不需要对图像的每一像素点做相关运算，它是利用图像的灰度投影曲线做一次互相关运算来求其图像序列的帧间运动参数，因而具有较高的计算速度，是实现只含平移运动或旋转运动较小的图像序列稳定的较理想算法。但是投影算法设计时就只能补偿图像的平移运动，且稳像精度不能达到亚像素精度。

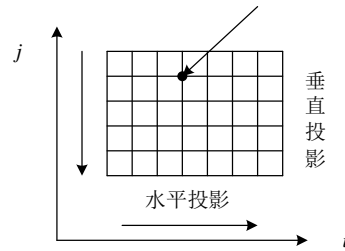


图2-5 投影原理示意图

### 2.2.3 特征匹配算法

由上文可知，基于块匹配和灰度投影的全局运动估计算法分别存在，计算复杂性高和不能处理旋转等缺陷。为了弥补这些缺点，有时可以使用基于特征匹配的方法。基于特征匹配的全局运动估计算法是以从图像中提取出来的某些特征作为匹配基元。一幅图像经过特征提取之后，由有限的特征模式表示，每个特征模式具有位置信息与特征描述量。特征的选择一般是基于一些纹理或者角点测度，因为具有丰富的纹理模式的区域在图像中通常是独特的，能够代表性的描述图像信息，而一致的或者线性的灰度区域却很普遍。为了进行后续处理，这些特征可以通过具有代表性的像素点（重心、线段端点、特殊点等）表示，这些点通常在文献中称为控制点(Control Points)。基于特征的匹配算法具有以下优点：

1. 图像的特征比图像的像素点要少很多，因此大大减少了计算量。
2. 特征点代表性的描述了图像的信息，可以提高匹配的精确程度。

如图2-6所示，基于特征匹配的运动估计算法首先在参考帧中选取特征，然后在后续帧中对特征进行定位，最后利用特征集对应关系，求取全局运动参数。算法基本步骤描述如下：

1. 特征检测：对图像序列中某一帧图像进行处理，提取图像特征，特征数目为 $m$ 。
2. 特征匹配：使用特征标识符、相似度测量以及特征间的空域关系，确定特征在当前帧图像的位置信息。
3. 变换模型估计：特征量的对应位置关系代入运动模型中联立方程，解超定方程组，求出全局运动矢量。



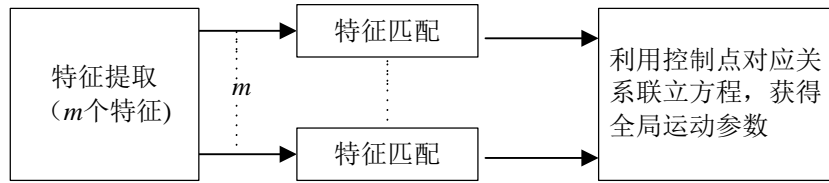


图 2-6 特征匹配法求全局运动估计参数

基于以上考虑本文选用基于特征匹配的运动估计方法，获取帧间全局运动参数，在第三章中将对基于特征匹配的运动估计方法做详细介绍。

### 2.3 由粗到精的运动估计思想

本文利用由粗到精的算法思想，采用金字塔分层技术进行帧间运动估计，以提高运算速度。Gaussian图像金字塔是一个图像序列，序列中的每一级图像均是前一级图像低通滤波的复制图像。图像的低通滤波由图像和高斯核卷积完成，将结果隔行降采样便得到了上层低分辨率的数据。以 $5 \times 5$ 大小的离散滤波核和4倍降采样系数为例，定义如下算子：

$$REDUCE(I)[i, j] = \sum_{m=1}^5 \sum_{n=1}^5 w(m, n) I[2i+m, 2j+n] \quad (2-17)$$

由上式，高斯图像金字塔如下定义：

$$G_0 = I \quad (2-18)$$

$$G_{i+1} = REDUCE(G_i) \quad (2-19)$$

其中 $I$ 为原始图像。采用图像金字塔技术可以提高运动估计算法的计算效率，有效的避免迭代计算时陷入局部极值。



图 2-7 图像三层高斯金字塔示意

分层参数估计的过程为：从最上层即最低分辨率级开始，在每一层进行运动

---

估计，将最上层的运动参数估计中的平移结果扩大一倍，作为第二层的运动参数估计的初始值，然后再在第二层进行更精细的匹配，重复与最上层计算相同步骤，直至计算出原始图像的运动参数。

---

## 第 3 章 基于特征匹配的全局运动估计

关于稳像中的运动估计问题, 研究人员提出了许多算法, 其中基于特征匹配的运动估计方法因具有运算量小和可以处理旋转运动的优点受到了研究人员的广泛关注。基于特征匹配的运动估计方法的难点在于提高特征的选取和匹配算法的精度与算法在复杂环境下的鲁棒性, 本章首先介绍本文所使用的特征窗匹配算法的基本原理和算法流程, 然后讨论特征检测和特征匹配这两个重要模块, 并且对该方法所存在的以下两方面的问题进行研究:

1. 特征匹配对于光线变化和运动前景的鲁棒性问题。光线变化可能引起有效特征的丢失, 而前景物体的运动会对全局运动估计造成误差。

2. 传统的基于特征丢失比率的抖动检测方法仅适用于定点拍摄。因此现有的稳像系统较少涉及抖动检测过程而基本采用直接补偿或手动触发方式。直接补偿在视频稳定时存在因资源浪费造成的效率下降, 而手动触发补偿给用户操作带来不便。

### 3.1 基于特征匹配的算法

图像的特征是图像场景中可作为标志的属性。利用特征匹配法估计连续视频图像序列的帧间运动参数, 首先要在参考图像中确定一组特征作为图像标识, 并对当前图像进行搜索, 以寻找到对应的特征结构, 从而获得图像序列的帧间运动参数。利用图像特征法估计视频图像序列的帧间运动参数的基本步骤为:1) 对图像序列中第一帧图像进行处理提取图像特征, 特征数目为  $m$ ; 2) 读入下一帧视频图像, 确定特征在下一帧图像的位置信息; 3) 将控制点的对应位置关系代入图像运动模型中联立方程, 解超定方程组, 求出全局运动参数; 4) 若部分特征匹配失败造成特征数目减少为  $n$  个, 需重新提取  $m-n$  个特征, 使特征个数始终为  $m$ , 转到步骤 2)。

#### 3.1.1 特征提取

目前, 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, 其中角点是应用最广泛的一种图像特征。特征匹配结果的好坏, 很大程度上取决于角点的提取。在已发表的文献中, 提取

角点的方法大致可分为两类：1) 基于图像边缘的角点提取算法<sup>[38]</sup>，2) 直接基于图像灰度的角点检测算法<sup>[25],[39-43]</sup>。

基于图像边缘的角点检测算法认为角点首先是一种边缘上的点，是一种特殊的边界。这类算法先检测出图像边缘，然后再检测出边界上出现突变的点作为检测到的角点。基于边缘的方法，由于要先提取边缘，然后再进行角点检测，因此，计算复杂、运算时间较长，并且对边缘提取算法依赖性较大。如果提取的边缘发生错误，或是边缘线发生中断(在实际中经常会遇到这种情况)，则会对角点的提取结果造成很大影响。

直接基于图像灰度的角点检测算法选取局部范围内灰度和梯度变化剧烈的极大点，这类算法所应用的手段主要是通过计算曲率及梯度来检测角点。其中较著名的有：Kitchen 发现以局部梯度乘以梯度方向的变化可以很好的提取角点<sup>[40]</sup>，Moravec 提出利用灰度方差提取点特征的方法<sup>[41]</sup>，Harris 和 Stephens 采用了相同的思想对 Moravec 算子进行改进，提出了著名的 Plessey 角点探测算子<sup>[42]</sup>，在 Plessey 角点提取思想的基础上 Shi-Tomasi 提出检测便于跟踪特征窗的思想<sup>[25]</sup>，Smith 和 Brady 提出的一种 SUSAN(Smallest Univalued Segment Assimilating Nucleus)探测算子<sup>[43]</sup>。

### 1. Plessey算子

Plessey 算子是 C.Harris 和 M.J.Stephens 在 1988 年提出的一种基于信号的点特征提取算子。这种算子受信号处理中自相关函数的启发，给出与自相关函数相联系的矩阵  $M$ 。 $M$  矩阵的特征值是自相关函数的一阶曲率，如果两个曲率值都高，那么就认为该点是特征点。图像梯度定义为：

$$g = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{bmatrix} \quad (3-1)$$

特征矩阵为：

$$Z = \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{\partial I}{\partial x} & \frac{\partial I}{\partial y} \end{bmatrix} = \begin{bmatrix} g_x^2 & g_x \cdot g_y \\ g_x \cdot g_y & g_y^2 \end{bmatrix} \quad (3-2)$$

角点响应函数为：

$$f_{resp}(x, y) = Det(Z) - kTrace^2(Z) \quad (3-3)$$

其中  $I(x, y)$  代表点  $(x, y)$  的灰度值， $Det$  为矩阵的行列式， $Trace$  为矩阵的迹， $g_x$  和

$g_x$  分别为参考图像的每个像素点在横向和纵向的一阶导数,  $k$  为参数, Plessey 建议取  $k$  为 0.04。上式含义为: 对灰度区域中的每个像素点, 计算其在横向和纵向的一阶导数以及二者的乘积, 得到 3 幅新的图像。3 幅图像中的每个像素对应的属性值分别代表  $g_x^2$ ,  $g_y^2$ ,  $g_x \cdot g_y$ 。Plessey 算法认为, 特征点是局部范围内的极大兴趣值对应的像素点。因此, 在计算完各像素点的兴趣值后, 要提取原始图像中所有局部兴趣值最大的点。实际操作中, 可依次从以每个像素为中心的  $3 \times 3$  的窗口中提取最大值, 如果中心点像素的兴趣值就是最大值, 则该点就是特征点。

Plessey 算子具有提取的角点分布均匀而且合理, 及特征稳定等优点。即使在有图像的旋转, 灰度的变化, 它也是一种较稳定的点特征提取算子。

## 2. Shi-Tomasi特征

著名的 Shi-Tomasi 特征利用图像的灰度信息, 选取含有丰富空间频率内容或者具有显著灰度变化的正方形特征区域, 这些区域具有便于匹配的特性。该方法提取的特征是一个图像方块, 用每一个图像方块的中心点描述特征的位置信息。

首先将图像划分成  $M$  个大小相同的互不重叠区域, 一般特征窗大小为  $7 \times 7$  像素。在每个特征区域  $\Omega_i$  内对式(3-2)做卷积, 由此可得到该区域的特征矩阵:

$$Z = \sum_{x \in \Omega_i} \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial I}{\partial x} & \frac{\partial I}{\partial y} \end{bmatrix} = \sum_{x \in \Omega_i} \begin{bmatrix} g_x^2 & g_x \cdot g_y \\ g_x \cdot g_y & g_y^2 \end{bmatrix} \quad (3-4)$$

设  $Z$  的两个特征值分别为  $\lambda_1$ ,  $\lambda_2$ , 则若:

$$\text{Min}(\lambda_1, \lambda_2) = 0.5 \left( g_x^2 + g_y^2 - \sqrt{(g_x^2 - g_y^2)^2 + 4(g_x \cdot g_y)^2} \right) > A \quad (3-5)$$

就认为该区域具有便于匹配的纹理模式, 可选为特征窗口, 其中  $A$  为阈值, 这里取  $A=1$ 。算法得到的特征窗是含有丰富空间频率或者具有多于一个方向的显著灰度变化的纹理模式的区域。

在实际应用中通常会提取过多的特征, 这时, 可根据需要对特征数目加以限制。例如, 将特征窗根据  $\text{Min}(\lambda_1, \lambda_2)$  进行排序, 选出较大的若干区域作为最后的结果。

同 Plessey 算子相比较 Shi-Tomasi 特征提取算法具有思路简明, 特征便于跟踪的特性, 因此本文实验中选用 Shi-Tomasi 特征提取方法选取特征窗。

### 3.1.2 特征匹配

本文采用Shi-Tomasi特征提取方法获取特征，而后利用特征匹配确定已选取的特征在后续视频图像中对应的位置信息（如图3-1所示）。

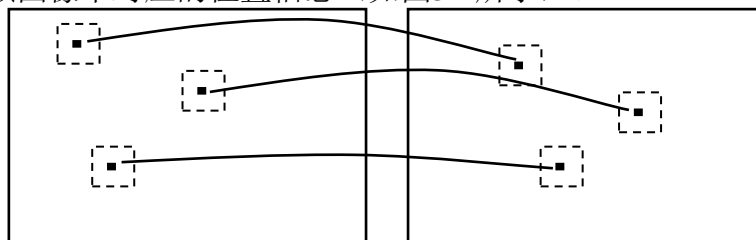


图3-1 两帧图像中的特征及其对应关系示意图

在视频序列首帧中自动检测并选取一组特征窗后，就需要在后续帧中匹配这些特征窗。匹配的基本原则就是搜索图像序列中连续两幅图像的相似之处，即对两幅图像中的特征进行匹配。由于使用彩色图像会提高算法的复杂度，并且对于匹配精度也没有明显改善，因此，为了简单起见，在特征匹配中我们使用八位灰度图像。考虑到一般视频的相邻帧时间间隔很短，可假设相邻帧图像公共部分的灰度值保持不变，如式(3-6)所示：

$$I_{k-1}(X) = I_k(X') = I_k(W(X, P)) \quad (3-6)$$

其中 $X = (x, y)^T$ ， $X' = (x', y')^T$ 分别为特征的中心坐标在参考帧第 $k-1$ 帧和当前第 $k$ 帧的坐标， $I_k(X')$ 为第 $k$ 帧图像坐标 $X'$ 的像素灰度值， $W$ 表示当前帧相对于参考帧的运动， $P$ 为运动参数。匹配的任务就是计算每个特征的 $P$ 值。我们采用Affine运动模型的齐次形式描述相邻帧的图像运动，如式(3-7)所示：

$$X' = W(X, P) = \begin{pmatrix} a_1 & a_2 & d_1 \\ a_3 & a_4 & d_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (3-7)$$

由于图像运动模型不够完善，还有图像噪声等的影响，等式(3-6)只能近似成立，问题转化为通过最小化匹配残差即误差平方和(SSD, Sum of Squared Differences), 求取运动参数 $P$ 。

$$SSD = \sum_{X \in \Omega_i} [I_k(W(X, P)) - I_{k-1}(X)]^2 \quad (3-8)$$

这里 $\Omega_i$ 表示第 $i$ 个特征窗，特征窗大小可根据需要设定，我们取 $7 \times 7$ （像素）。利用Newton-Raphson优化方法和泰勒公式可求得(3-8)式最小时的参数 $P$ ，进而获得特征在当前帧坐标值。具体步骤如下：

假定 $P$ 的初始值 $P_0$ 已知，将求解 $P$ 问题转换为不断地迭代求解 $P$ 的增量 $\Delta P$

的过程。获得 $\Delta P$ 后更新 $P$ 值 $P \leftarrow P + \Delta P$ ，则式(3-8)转化为:

$$SSD = \sum_{X \in \Omega_i} [I_k(W(X, P + \Delta P)) - I_{k-1}(X)]^2 \quad (3-9)$$

将 $I_k(W(X, P + \Delta P))$ 泰勒展开并省略高阶分量，有:

$$I_k(W(X, P + \Delta P)) = I_k(W(X, P)) + \frac{\partial I}{\partial W} \frac{\partial W}{\partial P} \square P \quad (3-10)$$

将式(3-10)代入(3-9)得到:

$$SSD = \sum_{X \in \Omega_i} [I_k(W(X, P)) + \frac{\partial I}{\partial W} \frac{\partial W}{\partial P} \square P - I_{k-1}(X)]^2 \quad (3-11)$$

其中 $\frac{\partial I}{\partial W}$ 是当前帧点 $W(X, P)$ 的梯度值，实际计算过程中，首先计算当前帧在 $x$ 方向和 $y$ 方向的梯度，而后利用变形参数映射到参考图像的坐标平面下。 $\frac{\partial W}{\partial P}$ 是 $W(X, P)$ 的 $Jacobi$ 矩阵，若 $W(X, P) = (W_x(X, P), W_y(X, P))$ 则有:

$$\begin{aligned} \frac{\partial W}{\partial P} &= \begin{pmatrix} \frac{\partial W_x}{\partial a_1} & \frac{\partial W_x}{\partial a_2} & \frac{\partial W_x}{\partial d_1} & \frac{\partial W_x}{\partial a_3} & \frac{\partial W_x}{\partial a_4} & \frac{\partial W_x}{\partial d_2} \\ \frac{\partial W_y}{\partial a_1} & \frac{\partial W_y}{\partial a_2} & \frac{\partial W_y}{\partial d_1} & \frac{\partial W_y}{\partial a_3} & \frac{\partial W_y}{\partial a_4} & \frac{\partial W_y}{\partial d_2} \end{pmatrix} \\ &= \begin{pmatrix} x & 0 & y & 0 & 1 & 0 \\ 0 & x & 0 & y & 0 & 1 \end{pmatrix} \end{aligned} \quad (3-12)$$

当(3-11)式值最小时，应满足对 $\Delta P$ 偏导数为0，则有:

$$\sum_{X \in \Omega_i} \left( \frac{\partial I}{\partial W} \frac{\partial W}{\partial P} \right)^T [I(W(X, P)) + \frac{\partial I}{\partial W} \frac{\partial W}{\partial P} \square P - I_{k-1}(X)] = 0 \quad (3-13)$$

解方程得:

$$\square P = H^{-1} \sum_{X \in \Omega_i} \left[ \frac{\partial I}{\partial W} \frac{\partial W}{\partial P} \right]^T [I_{k-1}(X) - I(W(X, P))] \quad (3-14)$$

其中 $H$ 是:

$$H = \sum_{X \in \Omega_i} \left[ \frac{\partial I}{\partial W} \frac{\partial W}{\partial P} \right]^T \left[ \frac{\partial I}{\partial W} \frac{\partial W}{\partial P} \right] \quad (3-15)$$

不断重复求解 $\Delta P$ ，更新 $P$ 值 $P \leftarrow P + \Delta P$ ，直到 $\Delta P$ 小于临界值 $\varepsilon$ 或者迭代次数过大时，停止计算，此时的 $P$ 值即为所求。类似的可以采用 Steepest-Descent, Newton, Levenberg-Marquardt 等优化方法进行迭代求解。

### 3.1.3 存在问题

在基于特征匹配的稳像方法中，特征提取、特征匹配是后续操作的基础，对最终结果有较大影响。所以，提高特征匹配精度和鲁棒性是实现高精度算法的基础，经典的特征匹配稳像算法主要存在以下几个问题：

1. 在真实场景中，各特征所处的环境往往很复杂。由于旋转、遮挡、变形、环境遇到超出假设条件等因素的影响，在长时间匹配过程中会产生不可信特征。为了保证特征集的可信性，提高稳像精度，在实际应用中需要在匹配过程中对特征可信性进行监控。

2. 由于外界光线变化造成的有效特征丢失。在光线骤然变化时，Shi-Tomasi 算法假定亮度不变的前提将不再成立，按照原有模型处理会造成大量有效特征丢失。

3. 当特征位于发生局部运动的物体（比如晃动的树叶等）上时，会给全局运动估计造成偏差。因此，需要去除局部运动物体对全局运动估计的影响。

4. 没有涉及抖动检测。

针对以上问题，Yan<sup>[32]</sup>等人提出了基于单特征的抖动检测算法，但是作者在算法中没有考虑第一个问题的影响，因此特征匹配鲁棒性较差。Shi和Tomasi利用图像残差进行特征有效性验证<sup>[25]</sup>，Sinisa等人在特征中抽取可信特征支持<sup>[36]</sup>，Michele等人在反射模型的基础上减少了亮度剧变时小特征窗的匹配丢失现象<sup>[37]</sup>，提高了特征匹配的鲁棒性。可以看出，以上几个研究都只是针对抖动检测或者特征匹配算法鲁棒性中的某一方面进行研究。对于两个方面的问题都进行处理的算法目前还很少看到。

在本文中，作者综合考虑了以上两个问题，设计出一种基于特征集合匹配的全局运动估计方法。该算法在进行特征匹配时融入亮度变化自适应模型并利用特征匹配误差分析和运动一致性原则对初步匹配结果作有效性验证以提高算法鲁棒性。同时，在运动估计模块中加入基于特征集合匹配关系的抖动检测算法。该方法通过对帧间运动参数进行分析确定视频是否有抖动发生，进而确定是否需要作进一步的运动补偿处理，从而避免在视频没有发生抖动时产生由补偿引起的系统效率下降。

## 3.2 鲁棒性特征匹配算法

针对基于特征匹配的运动估计算法中存在的问题，本文提出一种简单高效的

---



方法增加全局运动估计的鲁棒性。

### 3.2.1 亮度自适应特征匹配

经典的 Shi-Tomasi 算法在匹配过程中假定特征亮度不发生变化, 在这种前提下, 当外界光线发生突变时会引起特征的匹配距离突然增大, 进而造成大量有效特征丢失。本文结合亮度变换的均匀特性, 对算法加以改进, 使之对于亮度具有自适应性。在算法中, 因为每一个特征包含的图像面积较小, 通常可近似假定特征区域内像素点亮度变化具有统一性, 由此可利用特征在相邻帧中的整体亮度信息描述特征内像素点的亮度变化。根据上述假定条件, 对(3-8)式作出改进。

在目标函数中加入对比度及亮度调整因子 $\lambda(k)$ 、 $\delta(k)$ , 其中 $\lambda(k)$ 用特征 $\Omega_i$ 在第 $k$ 帧图像和参考帧图像间亮度的对比度进行计算:

$$\lambda(k) = \sqrt{\frac{\sum_{X \in \Omega_i} [I_{k-1}(X)]^2}{\sum_{X \in \Omega_i} [I_k(X')]^2}} \quad (3-16)$$

$\lambda(k)$ 反映了某个特征 $\Omega_i$ 从参考帧到第 $k$ 帧图像的亮度整体变化率。 $\delta(k)$ 则表示该特征整体亮度残差, 计算为:

$$\delta(k) = \overline{I_{k-1}} - \lambda(k) \cdot \overline{I_k} \quad (3-17)$$

$$\overline{I_k} = \frac{\sum_{x \in \Omega_i} [I_k(X')]}{L \cdot L}, \overline{I_{k-1}} = \frac{\sum_{x \in \Omega_i} [I_{k-1}(X)]}{L \cdot L} \quad (3-18)$$

其中 $L$ 表示特征的边长(单位: 像素)。在得到亮度调整因子以后, (3-8)式可转化为:

$$SSD = \sum_{X \in \Omega_i} [I_{k-1}(X) - \lambda(k) \cdot I_k(W(X, A, D)) - \delta(k)]^2 \quad (3-19)$$

(3-8)转化为(3-19)以后, 特征匹配的总目标函数中对整体亮度变化进行了线性调整, 从而避开了亮度变化对于目标函数的影响。实验结果也表明, 引入亮度及对比度因子可较大程度减少由于外界亮度突变造成的有效特征丢失和误匹配。

### 3.2.2 特征有效性判定

匹配处理过程中由于特征被遮挡或者消失及运动模型与实际运动差异等因素会造成特征的误匹配。同时, 当场景中存在发生局部运动的前景物体时, 由于局

部运动不能真正反映摄像机的运动形式，会给全局运动估计带来偏差。为了提高精度，本文分别采用特征误差分析方法与运动一致性规则进行特征有效性判定，以提高算法的精度及鲁棒性。

### 1. 有效性判定准则

为了保证匹配特征的鲁棒性，本文使用了一种特征有效性判定规则对提取的特征进行筛选。如果定义  $\varepsilon_i$  为特征  $\Omega_i$  在第  $k$  帧和参考帧  $t$  的误差平方和

$$\varepsilon_i = \sum_{X \in \Omega_i} [I_t(X) - I_k(W(X, P))]^2 \quad (3-20)$$

根据观察和实验，我们对于误差变量  $\varepsilon$ ，可以假定其为正态分布，也就是其出现概率可以使用如下公式计算：

$$P(\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\varepsilon - \mu)^2}{\sigma^2}\right) \quad (3-21)$$

其中  $(\mu, \sigma)$  表示期望和方差，可以由所有特征的误差集合进行参数估计得到<sup>[44]</sup>。由高斯分布的数学特性可知区间  $[\mu - 3.5\sigma, \mu + 3.5\sigma]$  上的数据占高斯分布总数据的 99.9%。基于高斯分布的思想，区间  $[\mu - 3\sigma, \mu + 3\sigma]$  以外的数据出现的概率非常小，是小概率事件。因此我们认为，当特征  $\Omega_i$  的误差  $\varepsilon_i$  在区间以内时，即

$$|\varepsilon_i - \mu| < 3\sigma \quad (3-22)$$

此特征为有效特征，否则为无效特征。至此，我们根据特征匹配误差集合自身的分布规律，提出了特征的有效性判定准则。

### 2. 运动前景特征剔除

前景物体的运动会影响整体的运动参数估计，进而影响稳像结果。为了避免位于运动前景上的特征参与后续计算，根据运动一致性原则对通过初步验证的特征进行分析。运动一致性原则可叙述如下：当拍摄场景中存在运动前景物体时，该物体运动向量的模及方向同场景全局运动向量相比有一定偏差。根据运动一致性原则，可以首先获得当前第  $k$  帧中的某个特征相对于参考帧的运动向量，而后确定单特征的运动向量和平均运动向量间模及方向角的偏差。若偏差超出某个预定的阈值，则标识该特征为无效特征。以上过程的数学描述如下：

算法以  $V(k, t)$  作为样本， $\|V(k, t)\|$  和  $ang(V(k, t))$  为其样本值。其中  $\|\cdot\|$  表示向量模、 $ang$  表示向量方向， $V(k, t)$  代表当前第  $k$  帧相对于参考帧  $t$  的运动向量，类似  $V_i(k, t)$  描述某一特征  $\Omega_i$  的运动向量。由此运动向量模的标准差：

$$\sigma\|(V(k, t))\| = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (\|V(k, t)\| - \|V_i(k, t)\|)^2} \quad (3-23)$$

其中：

$$\|\overline{V(k,t)}\| = \frac{1}{n} \sum_{i=0}^{n-1} \|V_i(k,t)\| \quad (3-24)$$

$n$  为匹配的特征数量。我们的目的是将那些和平均运动向量相差比较大的特征判定为前景特征，进而从特征集合中剔除。综合考虑运动向量方向的标准差，可以定义如下判定函数：

$$S(i) = \begin{cases} 1, & \|\overline{V(k,t)} - V_i(k,t)\| < 3\sigma \cdot \|V_i(k,t)\| \cap |\overline{ang(V(k,t))} - ang(V_i(k,t))| < 3\sigma \cdot ang(V_i(k,t)) \\ 0, & otherwise \end{cases} \quad (3-25)$$

当  $S(i)$  为 1 时表示特征  $\Omega_i$  有效，为 0 时特征  $\Omega_i$  无效。

由于算法采用的特征有效性判定方法是基于统计原则的，根据  $3\sigma$  准则<sup>[44]</sup>，这种校验方法不能检验样本数量较小（样本数  $\leq 10$ ）的情况。在相邻帧匹配处理过程中由于特征被遮挡或者消失及运动模型与实际运动差异等因素，会使得匹配特征数量  $n$  有可能小于选取特征数量  $m$ ，因此，为保证算法的有效性，实验中选取特征数量至少为 25，以保证匹配特征对数量大于 10。并且，为了避免特征数量过多造成的算法计算复杂性提高，效率下降，选取特征数量一般不超过 50，即选取特征数量范围为  $25 \leq m \leq 50$ 。

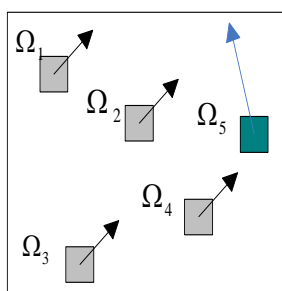


图 3-2 运动一致性示意图

### 3.3 全局运动参数求解

稳像算法的全局运动参数的计算方式按图像的匹配来源，基本上可以分为两种方式。一种是固定帧匹配，即在系统运行初期的某一帧图像就被采集下来作为以后图像的参考帧，以后的图像都作为匹配帧与它进行比较来计算相对运动参数；另一种则是相邻帧匹配，顾名思义，就是对每相邻的两帧图像都进行比较，来计算运动参数。两种方法各有优缺点，固定帧匹配方法由于不存在累计误差，所以参数估计较为准确，定点拍摄时视频稳定效果好，且相对简单易行，也是在运动估计算法的研究中较为常用的。这种方法的主要问题是一旦出现失稳状况，即图

像的运动超出了运动参数计算的临界值，或者是图像匹配程度偏低，运动参数变得不可靠以后，后续的处理非常困难，同时难以实现在扫描拍摄时的稳像。另一方面，相邻帧匹配相比于固定帧匹配，缺点是运算参数存在累计误差，对匹配精度有不利影响，但是由于始终是相邻的两帧进行匹配，所以在固定帧匹配当中不同参考帧带来的跳变误差，实际上被分散到了每一对相邻帧的运动估计过程中，因此不会因为参考帧相差较大而出现明显的图像跳变行为，适合于对视频稳定精度要求不高的普通摄像应用。本文以下的讨论即是基于相邻帧匹配的方法来进行的。

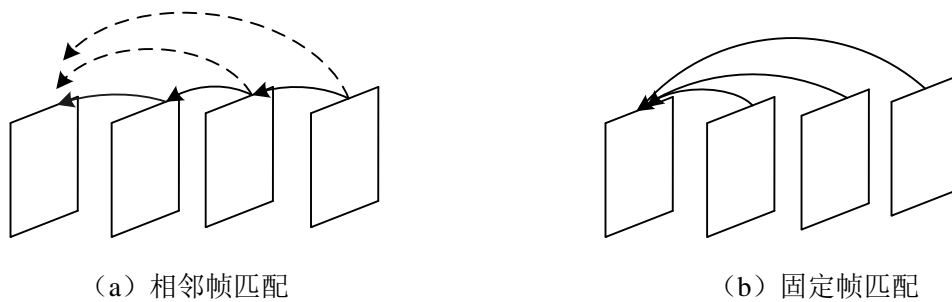


图 3-3 帧间运动参数关系示意图

### 3.3.1 相邻帧运动参数求解

在采用改进后特征匹配算法的基础上，本文利用由粗到精的运动估计思想，采用金字塔分层技术确定特征在下一帧图像的位置信息，以提高系统运算速度和精度，具体流程如下：

1. 首先对金字塔第三层进行全局运动参数估计。由于该层图片较小，计算量小，此时选用灰度投影法估计平移参数，这主要是针对大平移运动，初步确定特征粗略位置。

2. 将最上层的运动参数估计结果扩大一倍，作为第二层的运动运动参数估计的初始值。对第二层图像选取特征窗，然后采用 Affine 运动模型，用 Newton-Raphson 方法进行运动参数估计，确定特征位置。对初步匹配上的特征进行有效性验证，去除无效特征。

3. 将计算结果(平移参数扩大一倍)映射到第一层金字塔，采用 Affine 运动模型，用 Newton-Raphson 方法进行运动参数估计和有效性验证，获得各个特征在当前帧的位置信息。

将一特征在相邻帧的位置代入 Affine 模型中，可得线性系统方程：

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \quad (3-26)$$

若存在  $n$  对匹配特征, 则可以得到一个实数域上的  $(2n \times 6)$  阶线性方程组, 方程组的解即为相邻帧全局运动参数。而一般情况下所获得的方程个数  $2n > 6$ , 即超定方程组。这种方程组没有通常意义下的解, 为此需求方程组的一个最小二乘解。定义能量函数  $\rho_x, \rho_y$ :

$$\begin{aligned} \rho_x &= \sum_{i=0}^N (a_1 x_i + a_2 y_i + d_1 - x')^2 \\ \rho_y &= \sum_{i=0}^N (a_3 x_i + a_4 y_i + d_2 - y')^2 \end{aligned} \quad (3-27)$$

则有:

$$\frac{\partial \rho_x}{\partial a_1} = \frac{\partial \rho_x}{\partial a_2} = \frac{\partial \rho_x}{\partial d_1} = \frac{\partial \rho_y}{\partial a_3} = \frac{\partial \rho_y}{\partial a_4} = \frac{\partial \rho_y}{\partial d_2} = 0 \quad (3-28)$$

得到一组线性方程如式(3-29), 求解此线性方程, 可求得参数  $a_1 \sim a_4, d_1 \sim d_2$ 。

$$\begin{aligned} a_1 \sum_{i=0}^n x_i^2 + a_2 \sum_{i=0}^n y_i x_i + d_1 \sum_{i=0}^n x_i - \sum_{i=0}^n x_i x_i' &= 0 \\ a_1 \sum_{i=0}^n x_i y_i + a_2 \sum_{i=0}^n y_i^2 + d_1 \sum_{i=0}^n y_i - \sum_{i=0}^n y_i x_i' &= 0 \\ a_1 \sum_{i=0}^n x_i + a_2 \sum_{i=0}^n y_i + N d_1 - \sum_{i=0}^n x_i' &= 0 \\ a_3 \sum_{i=0}^n x_i^2 + a_4 \sum_{i=0}^n y_i x_i + d_2 \sum_{i=0}^n x_i - \sum_{i=0}^n x_i y_i' &= 0 \\ a_3 \sum_{i=0}^n x_i y_i + a_4 \sum_{i=0}^n y_i^2 + d_2 \sum_{i=0}^n y_i - \sum_{i=0}^n y_i y_i' &= 0 \\ a_3 \sum_{i=0}^n x_i + a_4 \sum_{i=0}^n y_i + N d_2 - \sum_{i=0}^n y_i' &= 0 \end{aligned} \quad (3-29)$$

### 3.3.2 获取全局运动参数

利用多个相邻帧运动参数累计, 即可获得当前第  $k$  帧到参考帧  $r$  的运动参数  $P_k^r$

$= (A_k^r, D_k^r)$ ,  $P_i = (A_i, D_i)$  表示第  $i$  帧到第  $i-1$  帧的运动参数, 亦即  $P_i^{i-1}$ , 其中  $A = \{a_1, a_2, a_3, a_4\}$ ,  $D = \{d_1, d_2\}$ , 含义同式(2-10)。根据 Affine 运动模型则:

$$X_{k-1} = A_k X_k + D_k \quad (3-30)$$

其中  $X_k$  和  $X_{k-1}$  描述了某像素点在第  $k$  帧和第  $k-1$  帧的坐标信息。假设已知第  $k-1$  帧到第  $r$  帧的变换参数, 则相应的第  $k-1$  帧某点坐标  $X_{k-1}$  在第  $r$  帧坐标系下对应的位置  $X_r$  可以表示为:

$$X_r = A_{k-1}^r X_{k-1} + D_{k-1}^r \quad (3-31)$$

(3-30)式中的  $X_{k-1}$  带入式(3-31)中, 可以得到:

$$\begin{aligned} X_r &= A_{k-1}^r (A_k X_k + D_k) + D_{k-1}^r \\ &= (A_{k-1}^r A_k) X_k + A_{k-1}^r D_k + D_{k-1}^r \end{aligned} \quad (3-32)$$

由上式可知, 第  $k$  帧至第  $r$  帧的变换参数可由第  $k-1$  帧至第  $r$  帧的变换参数及第  $k$  帧到第  $k-1$  帧的变换参数累计推出, 依次类推, 第  $k-1$  帧至第  $r$  帧的变换参数可由第  $k-2$  帧至第  $r$  帧的变换参数及第  $k-1$  帧到第  $k-2$  帧的变换参数累计求得, 由此类推, 最终可由相邻帧运动参数累计, 求得当前帧到任意参考帧的变换参数。

$$A_k^r = \left( \prod_{i=0}^k A_i \right) X_r \quad (k \geq 2, k > r) \quad (3-33)$$

$$D_k^r = \sum_{i=1}^{k-1} \left( \prod_{j=i+1}^{j=k} A_j \right) D_i + D_k \quad (k \geq 2, k > r) \quad (3-34)$$

类似可得:

$$A_k^r = \left( \prod_{i=k}^r (A_{i+1})^{-1} \right) X_r \quad (r \geq 1, k < r) \quad (3-35)$$

$$D_k^r = \sum_{i=k}^{r-1} \left( \prod_{j=i+1}^{j=r} (A_{j+1})^{-1} \right) D_i + D_k \quad (r \geq 1, k < r) \quad (3-36)$$

### 3.4 抖动检测

从用户的角度分析, 摄像机的运动可分为两种: 主观运动与非主观运动 [3,13,29,32]。主观运动在短时间内一般向着同一个方向, 具有平滑的特性, 需要保留。非主观运动又可称为抖动, 具有高频特性, 需要被检测并且去除。传统的稳像算

法仅包含运动估计与运动补偿两部分，并未对主观运动和非主观运动进行区分。一般采用特征数目丢失率或者图像变化程度进行运动检测，但是，由于这些检测方法不能对主观运动和非主观运动进行区分，因此，它们实际上不适用于一般的稳像系统。针对以上问题，文献<sup>[32]</sup>利用某一特征  $\Omega_i$  在相邻帧运动向量夹角进行抖动检测。该算法认为若摄像设备在短时间内做往复运动，可表现为相邻帧运动向量夹角大于 90 度的出现频率较高。这种方法较好地解决了主观运动和抖动的区分问题，但仍存在以下两个问题：

1. 没有考虑运动前景物体对抖动检测的不利影响。
2. 当主观运动位移较大，抖动位移较小时，相邻帧间的向量夹角不能反映往复运动，算法可能无法检测出抖动。

针对以上两个问题，本文提出一种新的抖动检测方法。

根据抖动的往复特性，可以认为在视频发生抖动时，运动方向会短时间内频繁变化。若第  $k+1$  和第  $k$  帧相比在  $x$  方向发生了运动方向变化，则会有： $V(k+1, k)_x \cdot V(k, k-1)_x < 0$ 。类似的若在  $y$  方向发生了运动方向变化，有  $V(k+1, k)_y \cdot V(k, k-1)_y < 0$ 。其中  $V(k, t)$  代表当前第  $k$  帧相对于参考帧  $t$  的运动向量， $V(k, k-1)_x$  和  $V(k, k-1)_y$  分别为第  $k$  帧相对于前一帧在  $x$  方向和  $y$  方向的位移。如图 3-4 所示视频在  $x$  正方向运动，没有发生方向的变化，在  $y$  方向存在由抖动引起的方向变化，在图上表现为在  $y$  方向同时存在正负方向的相邻帧位移。分别对  $x$  方向和  $y$  方向进行抖动检测，当某一方向上发生方向变化的视频帧数目  $l$  超出特定阈值时，认为视频在该方向存在抖动，需要进行抖动补偿。实验中帧频为 15 帧/秒时，每 30 帧做一次抖动检测，当 30 帧中方向变化次数  $l$  超过 2 时，认为可能存在抖动，需进行抖动补偿，如式 (3-37) 所示：

$$Shaking = \begin{cases} 1, & \text{for each 30 frames, } l > 2 \\ 0, & \text{otherwise} \end{cases} \quad (3-37)$$

本文算法通过有效特征集获得相邻帧全局运动向量，增加了抖动检测算法对运动前景物体的鲁棒性。同时分别在两维方向上进行抖动检测，避免了某方向主观运动过大时对抖动检测的不利影响。

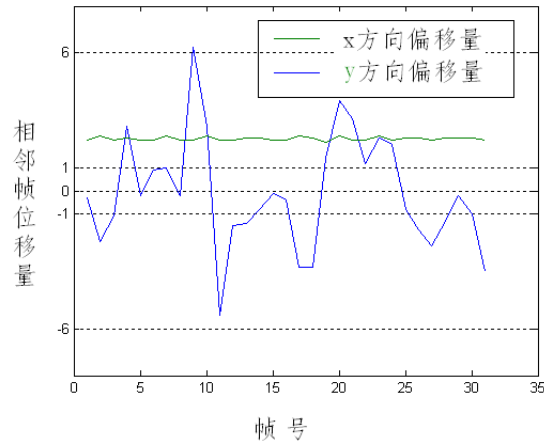


图 3-4 视频抖动示意图

### 3.5 小结

本章采用特征匹配法进行全局运动估计，并且对该方法存在的问题作了以下改进：

1. 匹配模型中加入亮度及对比度因子以减少由于亮度变化造成的有效特征丢失。
2. 由于特征残差符合高斯分布，利用高斯统计特性对特征进行优选。
3. 利用运动一致性原则，剔除位于局部运动物体上的特征，减少局部运动对全局运动估计的不利影响。
4. 为了增加稳像算法的智能性，在运动估计模块加入基于特征集对应关系的抖动检测算法，该检测算法具有对复杂外界环境的鲁棒性。
5. 结合使用图像金字塔技术减少迭代计算时陷入局部极值的可能性。



## 第 4 章 运动平滑与运动补偿

全局运动是由主观运动和抖动两部分叠加得到。为了在保留主观运动的基础上去除抖动，需要对抖动分量和主观运动进行分离。近年来，很多研究者认为，抖动本身具有高频特性，而主观运动具有慢速平滑的特性，可以通过低通滤波的方式获得主观运动参数，进而获取需去除的抖动分量。在实现稳像系统时，作者吸取该思想，在算法中加入运动平滑模块。利用运动平滑模块获得的抖动参数对视频中的各帧图像进行抖动补偿，并根据相邻帧间的像素相关性，重构校正变换后图像中出现的无定义区域，最终获得高质量的稳定化视频。图 4-1 为本章算法的流程：

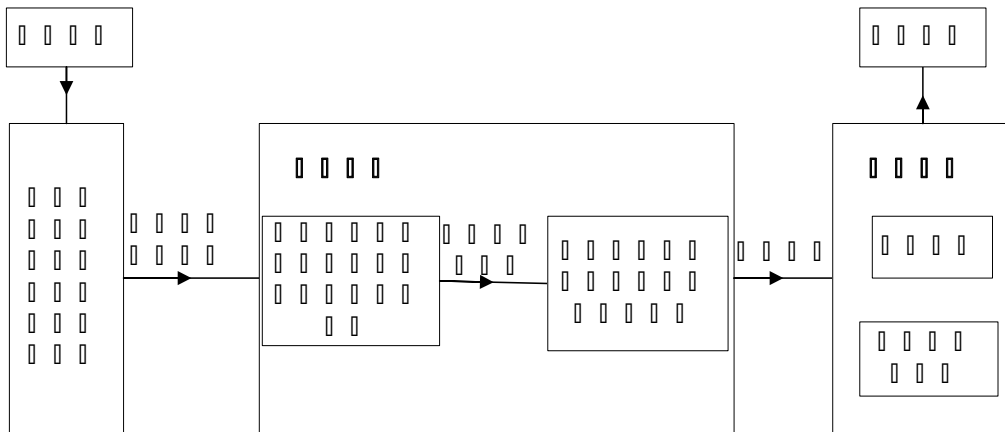


图 4-1 本章算法流程

### 4.1 运动平滑

在处理包含摄像机主观运动的非定点拍摄场景时，本文采用简单高效的均值滤波去除高频抖动。以水平平移分量为例，设平滑长度为  $n$ ， $D_i(x)$  表示第  $i$  帧与第  $i-1$  帧间的水平位移量，则当前第  $k$  帧与第  $k-n+1$  帧间的水平位移累计量  $D_k^{k-n+1}(x)$  为：

$$D_k^{k-n+1}(x) = \sum_{i=k-n+2}^k D_i(x) \quad (4-1)$$

该视频段的平均位移为：

$$\bar{D}(x) = \frac{1}{n-1} \sum_{i=k-n+2}^k D_i(x) \quad (4-2)$$

当前帧在水平方向上的位移补偿量为  $J_i(x)$ :

$$J_i(x) = \bar{D}(x) - D_i(x) \quad (4-3)$$

该滤波方法在某些特定的场景中能有效地平滑摄像机运动。但是,实际的摄像机运动具有不可预知性,抖动剧烈程度在不同场合也有所差别,实验表明该算法在采用固定的滤波器长度  $n$  时不能应用在所有的拍摄场景中。当  $n$  过小时达不到平滑的效果,在  $n$  过大的时候,则可能产生平滑过度的情况,都不能获得满意的平滑结果。为了提高算法适应性,降低对用户的要求,我们希望在以后的研究工作中,利用一定的先验知识,根据抖动频率的大小对基于均值滤波平滑方法中的滤波器长度  $n$  自动设置初值,并在稳定过程中根据抖动变化,自适应的调整滤波器长度。

## 4.2 运动补偿

运动补偿是根据运动估计及运动平滑获得的抖动参数重构视频图像,重新生成一个只存在摄像机主观运动的稳定视频序列。在完成运动平滑后,可以根据抖动运动参数,对视频中的每一帧图像进行校正变换,获得初步稳定的视频序列。同时根据相邻帧间的像素相关性,对稳像时产生的“空”区域进行拼接,最终获得高质量的视频输出。

### 4.2.1 抖动补偿

实验平台选用均值平滑方法获得抖动运动参数  $J_i$ 。设视频图像序列中的某一帧  $I_k$  变换到参考帧  $r$  的坐标系后的图像为  $I_k'$ ,  $X_r$  为当前第  $k$  帧中的点  $X_k$  在参考帧  $r$  坐标系下对应的位置,利用抖动运动参数,可得

$$X_r = J_k^r \cdot X_k \quad (4-4)$$

根据式(4-4)可得补偿后图像  $I_k'$  中某点在原始第  $k$  帧图像中的对应位置  $X_r^k$ :

$$X_r^k = (J_k^r)^{-1} \cdot X_r \quad (4-5)$$

$I_k'$  中某点  $X_r$  的灰度值即为  $I_k(X_r^k)$ 。如图 4-2 所示,由于  $(J_k^r)^{-1} X_r$  所计算值可能并不是整数,而像素点位置信息应由两个整数值描述。这时需要选择插值方法获取

像素值，常见的方法有最近邻法、双线性插值法、三次内插法<sup>[45]</sup>。

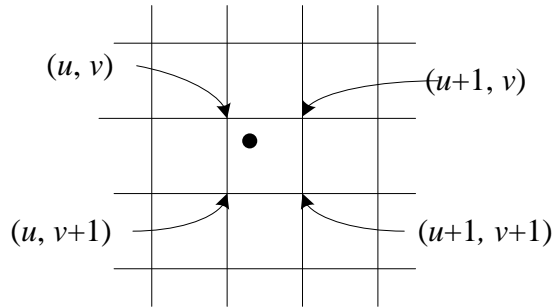


图 4-2 插值法示意图

### 1. 最近邻法

最近邻法即将与 $(x_0, y_0)$ 点最近的整数坐标 $(x, y)$ 点的灰度值取为 $(x_0, y_0)$ 点的灰度值。在 $(x_0, y_0)$ 点各相邻像素点灰度变化较小时，这种方法是一种简单快速的方法，但是当 $(x_0, y_0)$ 点相邻像素间灰度相差很大时，这种灰度估值方法会产生较大的误差，形成锯齿现象。

### 2. 双线性插值法

双线性插值法是对最近邻法的一种改进，即用线性内插方法，根据 $(x_0, y_0)$ 点的四个相邻点的灰度值，插值计算出 $f(x_0, y_0)$ 值，具体计算过程如下：

(1) 先根据 $f(u, v)$ 及 $f(u+1, v)$ 插值求 $f(u_0, v)$

$$f(u_0, v) = f(u, v) + \alpha(f(u+1, v) - f(u, v)) \quad (4-6)$$

(2) 再根据 $f(u, v+1)$ 及 $f(u+1, v+1)$ 插值求 $f(u_0, v+1)$

$$f(u_0, v+1) = f(u, v+1) + \alpha(f(u+1, v+1) - f(u, v+1)) \quad (4-7)$$

(3) 最后根据 $f(u_0, v)$ 及 $f(u_0, v+1)$ 插值求 $f(u_0, v_0)$

$$\begin{aligned} f(u_0, v_0) &= f(u_0, v) + \beta[f(u_0, v+1) - f(u_0, v)] \\ &= (1-\alpha)(1-\beta)f(u, v) + \alpha(1-\beta)f(u+1, v) + \\ &\quad (1-\alpha)\beta f(u, v+1) + \alpha\beta f(u+1, v+1) \end{aligned} \quad (4-8)$$

在实际计算时，若对任意 $s$ 值，规定 $\lfloor s \rfloor$ 表示其值不超过 $s$ 的最大整数，则上式中 $u = \lfloor u_0 \rfloor$ ， $v = \lfloor v_0 \rfloor$ ， $\alpha = u_0 - \lfloor u_0 \rfloor$ ， $\beta = v_0 - \lfloor v_0 \rfloor$ 。上述 $f(u_0, v_0)$ 的计算过程，实际是根据 $f(u, v)$ ， $f(u+1, v)$ ， $f(u, v+1)$ ， $f(u+1, v+1)$ 四个整数点的灰度值做二次线性插值得到的，则上述 $f(u_0, v_0)$ 可以改写为：

$$\begin{aligned} f(u_0, v_0) &= [f(u+1, v) - f(u, v)]\alpha + [f(u, v+1) - f(u, v)]\beta + \\ &\quad [f(u+1, v+1) + f(u, v) - f(u, v+1) - f(u+1, v)]\alpha\beta + f(u, v) \end{aligned} \quad (4-9)$$

该方法由于已经考虑到了 $(u_0, v_0)$ 的直接邻点对它的影响，因此一般可以得到令

人满意的插值效果，但是这种方法具有低通滤波性质，使高频分量受到损失，造成图像轮廓模糊。如果要得到更精确的灰度插值效果，可采用三次内插法。结合考虑精度与实时性，本文采用二次线性插值方法。

在一般情况下，采集图像为 RGB 格式，图像为  $m \times n \times 3$  矩阵形式，这时需要对 R, G, B 层分别进行插值处理，进而完成图像的重采样。

#### 4.2.2 无定义区域重构

对视频进行稳像处理后，当稳定后图像的像素点  $(J_k^i)^{-1} \cdot X_k^i$  不在参考图像坐标系范围内时，该像素点灰度值将为“空”，在图像边缘形成空白的待补偿区域。空白区域的存在，不仅丢失部分图像信息，同时又会影响人的视觉效果。为了获得高质量的视频输出，本文根据视频帧间具有很强的相关性的特性，利用图像拼接技术，进行图像合成以弥补图像中的空白区域。

令  $N_k = \{j: k-l \leq j \leq k+l\}$  为与当前第  $k$  帧相邻的  $2l$  帧图像，根据抖动运动参数，可以获得当前帧空白像素点在这  $2l$  帧中的对应点。使用拼接技术可从这  $2l$  个对应点中获得空白点位于图像坐标轴内的对应点，如时序中值过滤<sup>[9]</sup>，时序加权求和等方法。本文实验中使用时序中值过滤法，空白点  $X_k$  的像素灰度值为  $median(I_j(J_k^i \cdot X_k))$ ，实验结果如图 4-3 所示。



(a) 拼接前初步稳像结果



(b) 重构后稳定图像序列

图 4-3 图像重构后结果

---

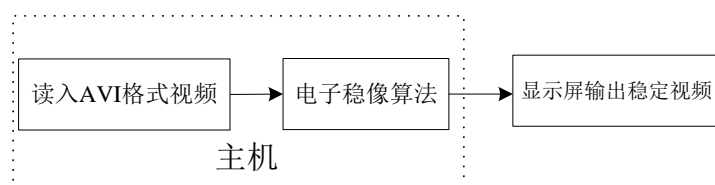
### 4.3 小结

本章简要介绍了均值滤波的实现方法，而后通过分析和实验指出均值滤波方法存在的适应性较差的问题，并给出未来的改进的方向。最后，对运动补偿中采用的插值和拼接方法作了简单描述。

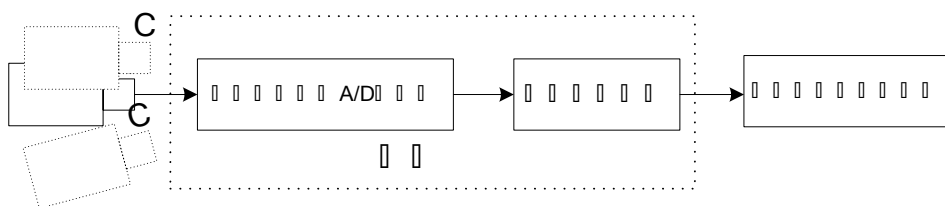
## 第 5 章 稳像算法实现及实验结果分析

### 5.1 稳像算法验证平台

由于本文所采用的实验平台需要对大量的视频图像数据进行处理，故选择了 Microsoft 公司开发的 Visual C++ 6.0 (VC) 作为软件开发工具。另外，OpenCV 开源类库提供了大量的封装良好的类，可以轻松高效的实现对数字图像的编程，因此，我们结合使用 OpenCV 以提高编程效率。



(a) 以AVI格式视频为处理源



(b) 对实时采集视频流进行处理

图 5-1 视频稳定平台操作需求

本文所采用的视频稳定操作平台的操作需求如图 5-1 所示。当输入视频流为计算机本身存储的 AVI 格式文件时，稳像算法直接对 AVI 视频流处理，将稳定后的图像序列送到 PC 机的显示器显示，此时只需要 PC 机即可获得稳定视频流；在对实时采集的视频流进行稳像处理时，需要将摄像设备与计算机上所安装的图像采集卡相连，把摄像设备所拍摄得到的视频信号送入图像采集卡。在图像采集卡内由 A/D 转换器将模拟视频信号转化为计算机可处理的数字图像，然后采用本文设计的稳像算法处理数字图像序列。最后将稳定后的图像序列送到 PC 机的显示器，输出稳定视频序列。系统主要用到的硬件设备有摄像设备，图像采集卡和计算机。

#### 1. 摄像设备

本实验所用摄像设备为 1/3 SONY CCD 摄像头，其自动快门档范围如下：  
EIA:1/60~12000 秒，CCIR:1/50~10000 秒，信号制式为 NTSC/PAL。

## 2. 图像采集卡

图像采集卡可以将摄像机拍摄的视频流量化成 2 维图像阵列。每个像素量化为 0-255 之间的整数，并保存在存储空间中等待处理。实验中采用 M404T 图像采集卡。M404T 图像采集卡是一种可采集标准/非标准、单色/RGB 的模拟/数字视频信号采集卡。可采集标准的模拟彩色/黑白视频信号，并将强大的函数扩展性能集成到采集卡上。它支持 NTSC/PAL/CCIR/RS170 等多种制式的图像信息，并且可以实时地把接收到的图像经数字化处理（A/D 转换）后经 PCI 或 PC/104-Plus 送往计算机内存等待进一步处理。主要特点有：可采集多种制式的视频信号，可处理 12 路视频输入。

## 3. 计算机

本系统所用计算机 CPU 为 Pentium IV/3.0 GHZ，内存为 512M，显示卡为 G-Force II，监视器为 Acer 77e。

本实验平台可对 AVI 视频和采集卡所获得的视频流进行稳定处理，程序运行后的主界面如图 5-2 所示：



图 5-2 平台界面

其中左窗口显示稳定后视频序列，右窗口显示稳定前的视频序列，可实现对源视频的播放显示。稳像系统默认读取采集卡上的视频，若找不到采集卡则等待点击打开按钮。在点击打开按钮时，选择读入 AVI 格式视频文件路径，视频源切换为 AVI 视频流。

## 5.2 稳像算法流程

本文的提出的算法流程可以用图 5-3 表示，其基本步骤可以描述如下：

1. 对输入图像序列进行高斯平滑去除噪声；

2. 利用 Shi-Tomasi 提出的特征提取算法选取利于匹配的特征;
3. 结合亮度变化自适应模型对 Shi-Tomasi 匹配器进行优化, 利用高斯金字塔由粗到精得到特征初步匹配结果;
4. 通过特征误差分析和运动一致性规则对特征集进行有效性验证以减少因遮挡和前景运动物体产生的无效特征, 获取有效特征集合;
5. 基于有效特征集在相邻帧间的对应关系, 求得运动参数;
6. 根据运动参数判断视频序列是否存在抖动, 若图像无抖动, 直接输出当前图像, 并读取下一帧, 否则转向抖动运动平滑模块;
7. 对图像采用均值平滑的方法去除高频运动得到平滑运动参数, 利用摄像机原始运动参数与平滑运动之间的差值得到需要补偿的高频抖动参数;
8. 利用高频抖动参数对当前帧图像做坐标变换获得初步稳定视频;
9. 利用相邻帧信息对空白区域进行填充;
10. 输出稳定图像序列。

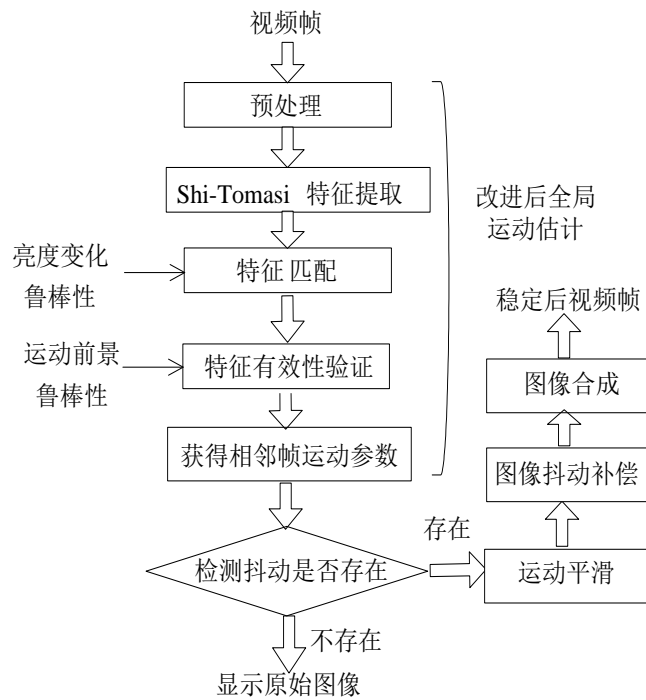


图 5-3 视频稳像算法框架

### 5.3 稳像算法评估标准

稳像系统的评估一般从运动矢量的精度、速率和可处理范围三个主要指标来对算法效果进行评估。



### 5.3.1 精度评估

在理想情况下，如果摄像设备只发生了需要去除的随机抖动，经稳定后的相邻两帧图像应完全重合。而在摄像机存在正常扫描运动或者视频流中存在运动的前景物体时，即使去除抖动后，相邻两帧图像间仍不能完全重合。但是无论何种情况，在对随机运动进行补偿后，获得的图像序列稳定程度将会有所提高，相邻帧重合度也随之变高。同时由于噪声、算法估计误差以及运动模式设置不精确等原因也会使得重合度不为 1。为了更准确地衡量算法效果，提出用算法的准确度来评价算法。目前常见的准确度的品质评估方法有三种：

#### 1. 均方差

均方差<sup>[3]</sup>(*MSE*, Mean Square Error)是两帧图像间每个像素的偏差值平方的和。它反映了图像序列变化的快慢和变化量的大小。*MSE* 越低，视频稳定效果越好。*MSE* 定义如下：

$$MSE(I_k, I_t) = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N [I_k(x, y) - I_t(x, y)]^2 \quad (5-1)$$

其中  $I_k(x, y)$  表示第  $k$  帧图像在  $(x, y)$  处的灰度值， $I_t(x, y)$  表示参考图像在  $(x, y)$  处的灰度值， $M$  和  $N$  分别表示图像的长和宽。*MSE* 作为评价稳定算法准确度的指标之一，反映了图像序列经过稳像以后和参考图像的对应像素匹配程度，*MSE* 值越大稳定程度越低。

#### 2. 峰值信噪比

基于峰值信噪比(*PSNR*)<sup>[3]</sup>的评定标准本质上和均方差相同，它所反应的是参考图像和当前图像间的峰值信噪比，本质与 *MSE* 相同。*PSNR* 越高，视频稳定效果越好。*PSNR* 定义如下

$$PSNR = 10 \log \frac{255^2}{MSE(I_k, I_t)} \quad (5-2)$$

参数含义同式(5-1)。

#### 3. 差影法

差值图像是图像的相减运算。对于定点拍摄如果能够将相邻两帧图像间的相对变化量完全补偿了，那么稳定后两帧图像差值图上相对应的像素值应为 0。然而，通常由于噪声、算法误差、量化误差等因素的存在，通常导致这个差值不为 0。偏差越大，重合度越低，稳定性越差。如图 5-4 所示

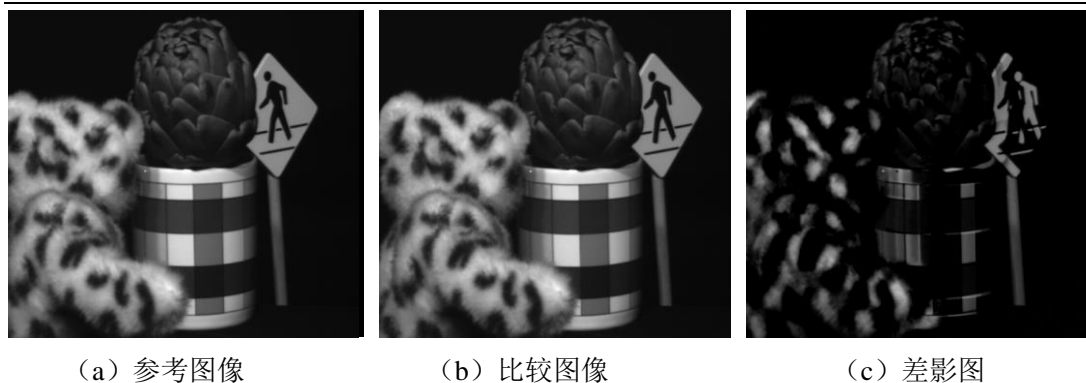


图 5-4 差影示意图

考虑到峰值信噪比的评定标准本质上和均方差相同而峰值信噪比法计算相对比较繁琐，我们这里采用均方差和差影法作为衡量算法稳定性的标准。

### 5.3.2 实时性评估

算法的速度是稳像算法的一个重要指标，它关系到系统能否实现对动态图像序列的实时处理，一般由帧处理率即每秒处理的帧数来衡量。传统算法中仅包含运动估计和运动补偿两部分，其中运动估计模块占用了大量的处理时间。近年来，研究者为了在保留摄像机的主观运动的基础上去除抖动，在算法框架中增加了运动平滑模块，使得计算复杂性进一步提高。因此，目前稳像算法的帧处理率主要是由运动估计模块和运动平滑模块决定，其中运动估计模块又起了决定作用。

算法的速度和精度经常是一对矛盾体，实现高精度的算法经常是以牺牲稳像速度为代价的，而片面的追求速度则会带来算法精度的下降。在实际应用中应该在满足基本应用要求的情况下，尽可能地提高速度或精度。

### 5.3.3 可处理范围评估

稳像系统可处理的最大偏移量是评价系统性能的一个重要指标。与速度指标类似，系统的可处理范围越大，估计运动参数模块可能消耗的时间越多。因此，在设计系统时，可设定部分可调参数以增加系统的适应性。例如在进行图像金字塔匹配时，调整金字塔的层数就可避免算法陷入局部极值点，进而影响增大算法可处理范围。但是金字塔层数越多，算法的处理复杂性也就越高。因此需要根据实际需要，选择合适的参数。

## 5.4 实验结果

下面分别是对室内、室外抖动视频稳定结果分析。

### 5.4.1 亮度无关特征匹配

图 5-5 所示为融入亮度变化自适应模型后和经典 Shi-Tomasi 匹配器的特征匹配结果对比。图 5-5-a 所示为参考图特征提取结果，特征数为 22。若采用经典 Shi-Tomasi 算法进行特征匹配，由于亮度变化造成的特征残差过大，仅匹配到 7 个特征，如图 5-5-b 所示。算法中加入亮度及对比度因子后，可匹配到 19 个特征，如图 5-5-c 所示。

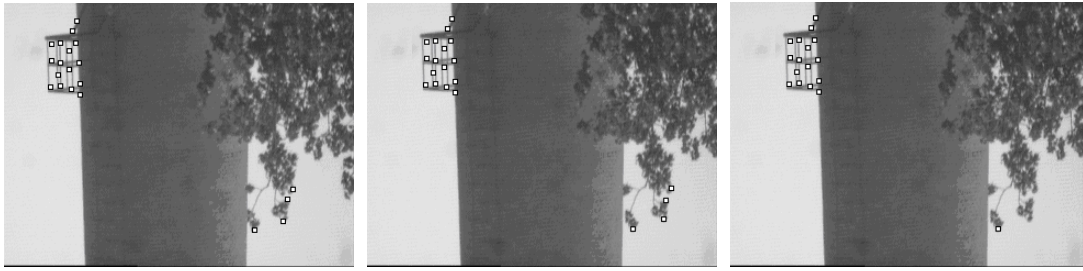


(a) 特征提取结果                      (b) 经典 Shi-Tomasi 特征匹配      (c) 本文算法特征匹配

图 5-5 亮度变化时的鲁棒性特征匹配结果和对比

### 5.4.2 特征有效性判定

运动前景特征验证实验结果如图 5-6 所示，视频序列中的烟囱本身是静止的物体，然而，摄像机的运动造成了烟囱在图像中左右晃动。同时视频中还存在发生局部运动的树叶，树叶随风摇动，与主观运动不一致，会对全局运动估计产生不利影响。我们希望通过特征进行有效性判定剔除位于晃动树叶上的前景特征。图 5-6-a 描述了特征在参考图像的位置，图 5-6-b 是利用亮度模型相结合算法获得的特征初步匹配结果，图 5-6-c 是对其进行有效性验证后的结果，位于晃动树叶（前景运动物体）上的前景特征基本被剔除。这说明了特征验证算法的有效性。在实验过程中每 50 帧的特征验证耗费时间为 1ms。



(a) 参考图像的特征提取结果 (b) 特征匹配初步结果 (c) 前景鲁棒性特征匹配结果

图 5-6 存在运动前景物体时特征匹配算法效果比较

### 5.4.3 随机抖动检测

对图 5-6 视频序列进行抖动检测，发现视频序列前 30 帧在  $x$  方向出现了 4 次方向变化如图 5-7 所示。根据判定条件，认为视频发生抖动。

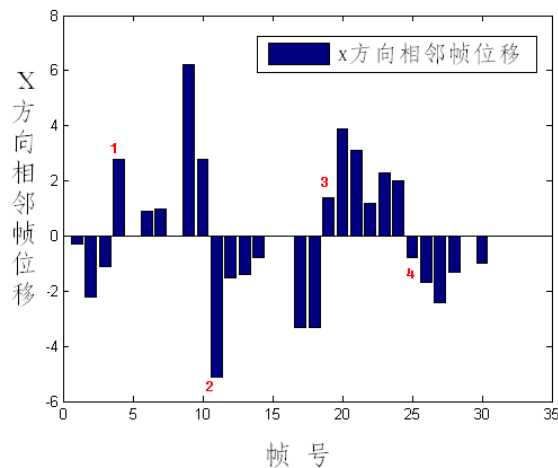


图 5-7 某方向相邻帧位移

### 5.4.4 视频稳定

示例 1: 对室内环境下实验室自拍抖动视频进行处理，图 5-8-a 分别为原始图像序列的第 30、40、50、60 帧，根据横向辅助线可以看出图像有明显的晃动。根据图 5-8-b 中的横向辅助线可以看出输出图像序列比较稳定，根据标识的小箭头可以看出在第 40 帧和第 50 帧之间基本消除了摄像机的平移运动，在第 60 帧基本消除了明显的旋转运动。灰色区域是由运动补偿造成的待拼接空白区域。



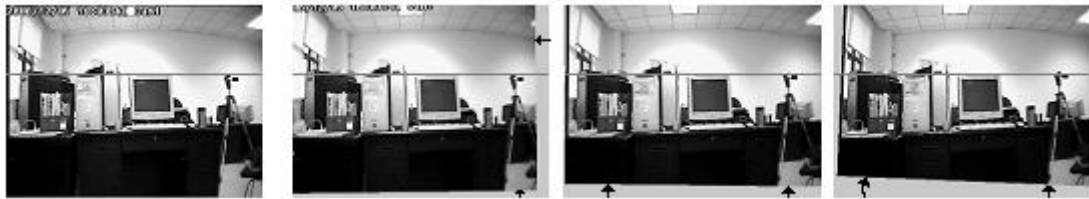
第 30 帧

第 40 帧

第 50 帧

第 60 帧

(a) 原始图像序列



第 30 帧

第 40 帧

第 50 帧

第 60 帧

(b) 初步稳定后图像序列

图 5-8 稳像算法实验结果示例一

**示例 2:** 对自然抖动情况下某一彩色视频进行处理, 给出初步稳定后视频序列, 及对无定义区域 (图像黑色区域) 进行处理后的图像序列, 实验表明处理后的视频可保持被摄物体位置基本不变, 如图 5-9 所示 (该抖动视频由哈尔滨工业大学提供)。



(a) 原始视频第 1 帧



(b) 原始视频第 31 帧



(c) 原始视频第 34 帧



(d) 原始视频第 55 帧



(e) 初步稳定后视频第 1 帧



(f) 初步稳定后视频第 31 帧



(g) 初步稳定后视频第 34 帧



(h) 初步稳定后视频第 55 帧





(i) 最终稳定视频第 1 帧



(j) 最终稳定视频第 31 帧



(k) 最终稳定视频第 34 帧



(l) 最终稳定视频第 55 帧

图 5-9 稳像算法实验结果示例二

**示例 3:** 本例对室外环境下包含摄像机主观扫描运动和存在运动前景物体的视频进行处理。由实验结果可知采用鲁棒性的运动估计算法基本上克服了行进中的汽车对全局运动估计造成的不利影响，获得了较精确的全局运动参数。由于摄像机本身包含主观运动，直接进行补偿，会造成图像信息的大量丢失，实验给出了直接处理和加入运动平滑算法后处理的对比结果，如图 5-10 所示。



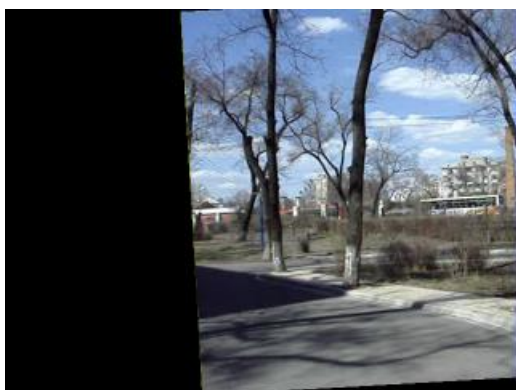
(a) 原始视频第 118 帧



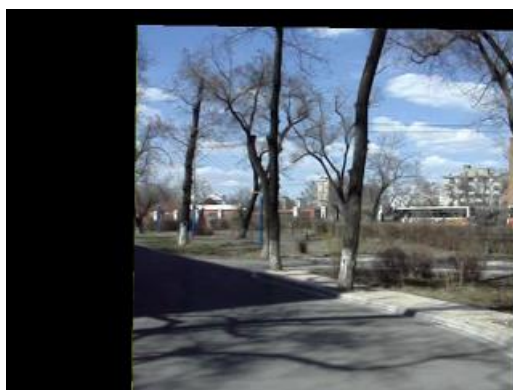
(b) 原始视频第 127 帧



(c) 原始视频第 132 帧



(d) 原始视频第 139 帧



(e) 未采用运动平滑处理后第 118 帧



(f) 未采用运动平滑处理后第 127 帧



(g) 未采用运动平滑处理后第 132 帧



(h) 未采用运动平滑处理后第 139 帧





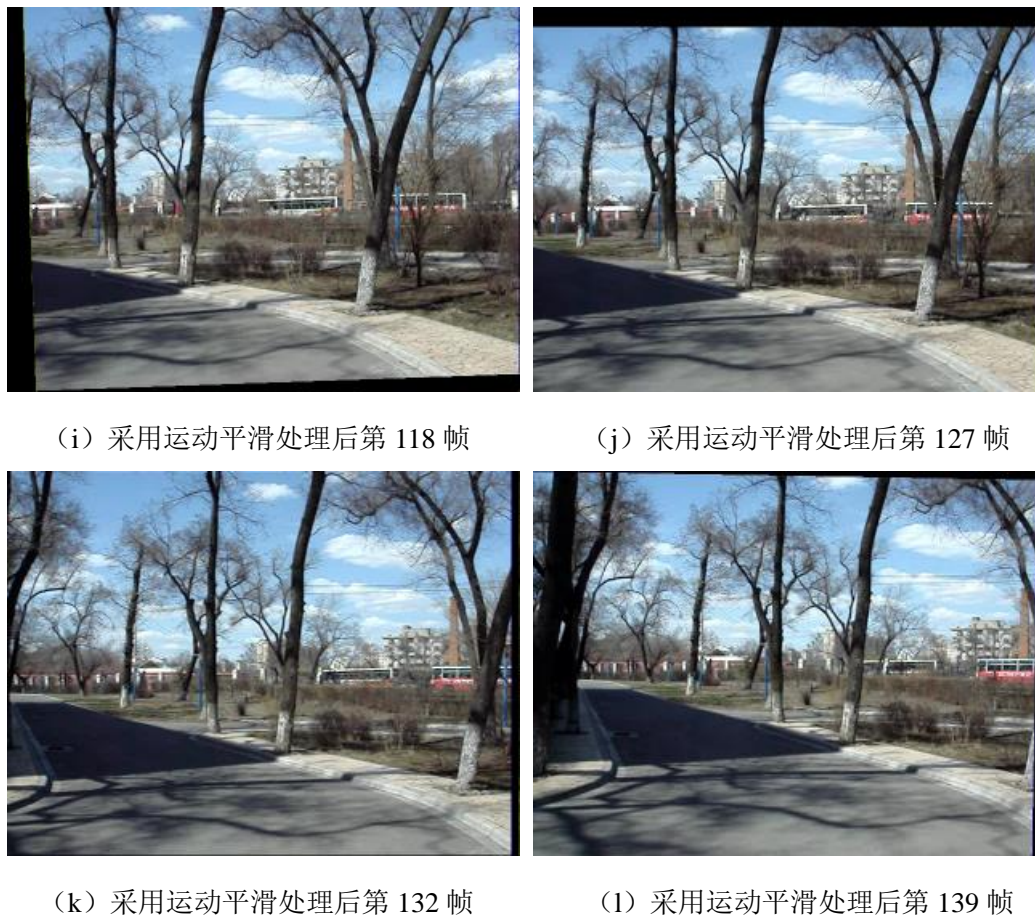


图 5-10 稳像算法实验结果示例三

**示例 4:** 对室外环境下的定点抖动视频进行稳定。实验结果表明，算法克服了发生局部运动的树叶的不利影响，稳像结果如图 5-11 所示，图 5-11-a 是原始图像序列的第 51, 65, 75 帧图像，可以看出图像有明显的左右抖动；图 5-11-b 为视频补偿后的初步稳定图像序列，基本没有抖动，图像黑色部分是由补偿造成像素缺失；图 5-11-c 和 5-11-d 分别为补偿前后的差影图像（反色后），通过对差影图像分析可知补偿后图像公共区域重合度较高（该抖动视频由北方车辆研究所提供）。

表5-1 运动估计部分结果与真值对比

摄像机运动方式	实际值	测量值（场景包含局部运动物体）
平移	69	69.34
平移	73	73.10
平移	92	92.87
平移	104	102.80
旋转	3 度	3.12 度



第 51 帧

第 65 帧

第 75 帧

(a) 原始图像序列

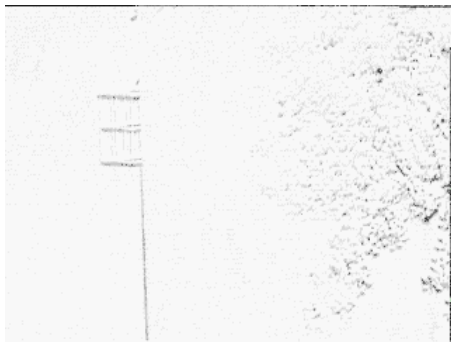


第 51 帧

第 65 帧

第 75 帧

(b) 稳定后图像序列



(c) 稳定前第 46 帧同 45 帧图像差影 (反色) (d) 补偿后第 46 帧同 45 帧图像差影 (反色)

图 5-11 稳像算法实验结果示例四

为了定量衡量稳像效果，我们对当前帧同参考帧之间的均平方误差  $MSE$  进行了计算。各算法  $MSE$  计算结果如图 5-12 所示。从图可见经过本文算法稳定后图像序列的  $MSE$  曲线远小于稳定前序列，且同特征有效性验证前相比本文算法  $MSE$  值也有所下降。 $MSE$  计算结果表明，本算法的稳像结果要比单纯采用传统的 Shi-Tomasi 的稳像算法质量提高 5~8%。

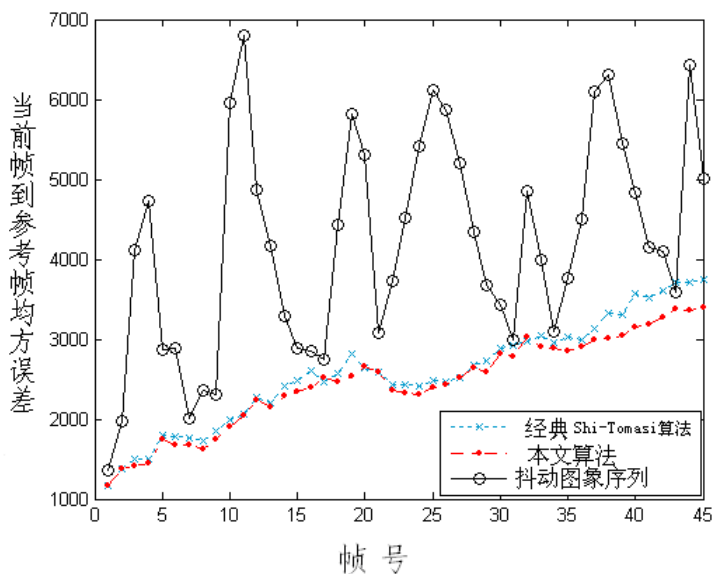


图 5-12 稳像结果均方差比较

## 5.5 小结

对实际拍摄的视频图像序列进行了稳像实验，并用定量评定方法对稳像结果进行了评定。实验结果表明稳像算法可以正确检测出视频抖动，同时算法对光线变化、小范围的局部运动具有较好的鲁棒性。需要注意的是，本文算法的适用范围包含如下条件：

1. 背景占据视频画面大部分面积；
2. 背景具有一定的边缘纹理；
3. 摄像设备的主要运动模式是水平平移，垂直平移，平面旋转或较小的前后平移；
4. 拍摄景物的深度变化不大。

## 总结与展望

本文设计了一种基于特征匹配的鲁棒性稳像算法。实验结果表明，算法对于外界光线变化和场景内部物体运动具有较好的鲁棒性。同时，我们利用抖动检测，减少了人工干预，提高了系统的智能性。主要研究工作和结论主要包括：

### 1. 提高复杂环境下全局运动估计的鲁棒性

全局运动估计的精度和速度基本决定了稳像算法的准确度和速度。本文采用简单高效的原则，在保证算法速度的同时，极大程度地提高了复杂环境下算法的鲁棒性。

首先针对外界光线变化造成的有效特征丢失和特征误匹配问题，引入自适应亮度变化模型，及时调整亮度与对比度因子以适应外界光线的变化。

利用特征残差所满足的统计原则对特征进行优选，去除无效特征。

采用运动一致性原则，去除位于局部运动物体上的外点，减少局部运动物体对全局运动估计产生的不利影响。

综合采用以上方法，可减少外界光线变化、噪声点和运动前景物体等不利因素对全局运动估计的精度所产生的影响，提高了算法的鲁棒性。由于全局运动估计是视频运动信息分析的基础，因此本文提出的全局运动估计算法在性能上的提高对进一步的视频运动信息分析具有重要的意义。

### 2. 提高系统智能性

在运动估计模块加入一种基于运动往复特性的抖动检测方法。该检测算法避免视频无抖动时的稳像误差，方便用户操作。

### 3. 实现视频稳定系统

建立了视频稳定系统模型，并用软件实现了该系统。系统能对获取图像进行实时稳定处理，并取得了初步稳定效果。处理速度可达到 15 帧/秒。

本文虽然针对基于特征的视频稳像算法中的若干主要问题进行了研究，并取得了相应的研究成果，但是在下列几方面还需要进一步研究：

#### 1. 拼接不吻合问题

文中根据帧间相关性，利用拼接技术，在对由于抖动补偿造成的空白区域进行重构，减少了图像信息缺失，获得了较高质量的稳定视频。但是若空白区域中存在局部运动物体时，会产生拼接失真现象。在以后的研究中，需在拼接时考虑局部运动物体的不利影响。

#### 2. 抖动检测的阈值设定

---

为了更好的进行抖动检测，需在大量的实验基础上，对各种应用需求进行统计分析，得到更可靠的经验阈值，以避免漏检和误检。

### 3. 自适应滤波

由于显示场景中的抖动模式变化多样，采用固定的滤波器长度不能应用在所有拍摄场景中。为了提高算法适应性，我们希望在以后的研究工作中，根据抖动频率的大小自适应的调整滤波器长度。

## 致 谢

时光飞逝，转眼间三年的硕士学习生涯即将告一段落。借此论文完成之际，向许许多多给予我指导、帮助、关心与支持的师长、同学和朋友以及亲人表示我最诚挚的谢意！

首先，我要特别感谢我的恩师陈志国教授！三年来，他从学业上和生活上都一直给予我悉心的指导和关心，为我创造宽松的研究环境。他渊博的学识，对问题实质的洞察入微和高屋建瓴，以及严谨求实的治学精神和宽厚坦荡的为人令我永志难忘，这在我今后的工作学习中必将使我受益匪浅。在生活上，陈老师也给我提供了无微不至的关怀，督促我不断成长进步。同时，衷心感谢我的另外一位导师焦建彬教授，在研究生院工程应用研究中心学习的近一年半时间里，焦老师在学习、生活上都给了我很多精心的指导和耐心的帮助，他对事业孜孜以求的精神、对生活乐观进取的态度永远是我学习的榜样。两位老师严谨求实的治学风格、诲人不倦的师表风范和精益求精的学术精神将使我终生受益。

感谢计算机与信息工程学院的郭拯危副教授、李玉玲老师、宋相法老师等。在三年研究生学习期间，他们在教学工作、研究工作过程中给予的指导和帮助，使我受益颇多。

感谢中科院研究生院叶齐祥老师、河南省招生办公室的王永刚老师，两位老师广博的知识、丰富的编程经验、忘我的工作精神、严谨的做事态度都给我留下了深刻印象。通过两位老师的教导，我的动手能力得到较大提高。

感谢三年来和我一起学习的各位同学和各位师兄师弟、师姐师妹们，我们大家互相鼓励一起度过了这三年的珍贵时光，同窗之谊弥足珍贵。

最后，感谢我的家人及好友，在我多年的学习生活中，是他们一直鼓励我向自己的目标前进。

---

## 参 考 文 献

- [1] 赵红颖, 金宏, 熊经武. 电子稳像技术概述[J]. 光学精密工程. 2001, 9(4): 354-358
  - [2] 谷素梅, 张经武. CCD摄像机光学惯性稳像跟踪头[J]. 光学技术. 1995, (2): 21-47
  - [3] 钟平. 机载电子稳像技术研究. 博士学位论文. 中国科学院长春光学精密机械与物理研究所. 2003
  - [4] Jean Maheux, James Cruickshank. Video-rate image stabilization system[J]. SPIE, 1998, 34-14
  - [5] 杨雨东, 徐光佑, 朱志刚. 2.5维数字图像序列稳定方法[J]. 清华大学学报(自然科学版). 1997, 37(9): 78-81
  - [6] 刘明, 赵跃进, 周渝斌. 电子稳像中的运动补偿矢量处理方法的研究[J]. 光学技术. 2005, 31(3): 457-459
  - [7] 罗诗途, 张王己, 王艳玲. 一种基于特征匹配的实时电子稳像算法[J]. 国防科技大学学报. 2005, 27(3): 46-48
  - [8] 孟龙, 林行刚, 王贵锦. 视频抖动补偿系统的运动滤波[J]. 清华大学学报自然科学版). 2005, 45(1): 41-45
  - [9] Yasuyuki, Matsushita, Eyal Ofek, Xiaou Tang. Full-frame video stabilization[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. 2006, 1150-1163
  - [10] C Zhu, X Lin, L P Chau. Hexagon-based search pattern for fast block motion estimation[A]. IEEE Trans on Circuits System Video Technology[C]. 2002, 12(5): 349-355
  - [11] Lidong Xu, Xinggang Lin. Stabilization based on circular block matching[A]. Proceeding of the 17th International Conference ICPR'04 Digital Image[C]. 2006,(5)
  - [12] Renxiang Li, Bing Zeng, M L Liou. A new three-step search algorithm for block motion estimation[A]. IEEE Trans on Circuits System Video Technology[C]. 1994, 4(2): 438-442
  - [13] Li Song, Yuanhua Zhou, Jun Zhou. Robust video stabilization based on motion vectors[J]. Journal of Shanghai University (English Edition). 2005, 9(1): 46-51
-

- 
- [14] 阳方林, 杨风暴, 韦全芳. 一种新的快速图像匹配算法[J]. 计算机工程与应用. 2005, (5): 51-53
- [15] A K Liu, E Feig. A block-based gradient descent search algorithm for block motion estimation in video coding[A]. IEEE Trans on Circuits System Video Technology[C]. 1996, 6(4): 419-423
- [16] 杨琛, 白波. 自适应运动估计搜索算法[J]. 计算机应用. 2006, 26(6): 41-43
- [17] 钟平, 于钱洋, 王明佳. 提高用于电子稳像的灰度投影算法精度的方法[J]. 光电子 激光. 2003, (14): 183-188.
- [18] 赵红颖, 晏磊, 张罡. 灰度投影拟和算法稳定船载侦察系统获得的视频图像[J]. 仪器仪表学报. 2004, 25(4): 280-293
- [19] 陈跃庭, 冯华君, 徐之海. 多参考点三步搜索法快速稳像算法[J]. 光电子 激光. 2006, 17(6): 746-749
- [20] 钟平, 于钱洋, 金光. 基于特征点匹配技术的运动估计及补偿方法[J]. 光电子 激光. 2004, (15): 74~80
- [21] Bruce D. Lucas, Takeo Kanade. An iterative image registration technique with an application to stereo vision[A]. International Joint Conference on Artificial Intelligence[C]. 1981, 674-679
- [22] Carlos Morimoto, Rama Chellapa. Fast electronic digital stabilization for off-road navigation[J]. IEEE Journal of Real-time Imaging. 1996, 285-296
- [23] A.Litvin, J.Konrad, W.C.Karl. Probabilistic video stabilization using kalman filtering and mosaicking[A]. Proc. of IS&T/SPIE Symposium on Electronic Imaging[C]. 2003, 663-674
- [24] Erturk S. Image Sequence stabilization based on kalman filtering of frame positions[J]. Electronic Letters. 2001,137(20): 1217-1219
- [25] Shi J B, Tomasi C. Good features to track [J]. IEEE Conference on Computer Vision and Pattern Recognition CVPR. 1994, 593-600,
- [26] G.D Hanger, P.N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination[A]. IEEE Trans on Pattern Analysis and Machine Intelligence[C]. 1998,20(10): 1025-1039
- [27] H. Jin, S. Soatto, P. Favaro. Real-time feature tracking and outlier rejection with changes in illumination. In IEEE Int. Conf. on Computer Vision. 2001: 684 - 689
- [28] T. Tommasini, A. Fusiello, E. Trucco, V. Roberto. Improving feature tracking with robust statistics[A]. Pattern Analysis & Applications[C]. 1999: 312 - 320,.
- [29] 张敏, 赵猛, 贾云得. 基于自适应高斯混合模型的图像稳定方法[J]. 北京理工
-



- 大学学报. 2004, 24(10): 888-900
- [30]赵红颖, 晏磊, 熊经武. 一种稳定船载电视侦察图像序列的灰度投影拟合算法[J]. 光学技术. 2004, 30(1): 83-88
- [31]Sung-Jea Ko, Sung-Hee Lee, Kyung-Hoon Lee. Digital image stabilizing algorithms based on bit-plane matching[A]. IEEE Transactions on Consumer Electronics[C]. 1998,44: 617-622
- [32]Weiqi Yan, Mohan S, Kankanhalli. Detection and removal of lighting & shaking artifacts in home videos[A]. Proceedings of the tenth ACM international conference on Multimedia[C]. 2002, 107-116
- [33]F H Cheng, S N Sun. New fast and efficient two-step search algorithm for block motion estimation[A]. IEEE Trans on Circuits System Video Technology[C]. 1999, 9(7): 977-983
- [34]Yeoumin Yeh, Sheng-Jyh Wang, Huang-Cheng Chiang. A digital camcorder image stabilizer based on gray coded bit-plane block matching[J]. SPIE, 2000
- [35]Carlos. Electronic digital image stabilization design and evaluation with applications. University of Maryland. 1997.
- [36]Sinsisa Segvic, Anthony Remazeilles, Francois Chaumette. Enhancing the point feature tracker by adaptive modeling of the feature support[A]. European Conference on Computer Vision [C]. 2006, 112-124
- [37]Mich`ele Gouiff`es, Christophe Collewet, Christine Fern. Feature points tracking: robustness to specular highlights and lighting changes[A]. European Conference on Computer Vision [C]. 2006, 82-93
- [38]郑宁. 数字视频稳像技术. 硕士学位论文. 北京理工大学. 2006
- [39]史阳, 高新波. 一种基于特征跟踪的视频序列稳像算法[J]. 光子学报. 2005, 34(7): 1099-1101
- [40]Kitchen L, Rosenfeld. A gray level corner detection[J]. Pattern Recognize, 1982,(1): 95-102.
- [41]Moravec HP. Towards automatic visual obstacle avoidance[A]. Proceeding of International Joint Conference Artificial Intelligence[C]. 1977
- [42]Harris C, Stephens M J. A. combined corner and edge detector[A]. Proc of 4<sup>th</sup> Alvery Vision Conference[C]. 1988, 189-192
- [43]S M Smith, M Brady. SUSAN-a new approach to low level image processing[J]. International Journal of Computer Vision. 1997, 23(1): 45-78
- [44]肖明耀. 实验误差估计与数据处理[M]. 科学出版社. 1981
-

[45] Linda G, Shaprio George C, Sockman. 计算机视觉[M]. 机械工业出版社. 2005

---

## 攻读学位期间发表的论文和参加的项目

### 一、发表的论文

1. 黄亚博, 翁梅, 李玉玲, 陈志国. 基于 Web Services 技术的跨地域应用系统的开发. 河南农业大学学报. 2007 年第 1 期.
2. 黄亚博, 焦建彬, 叶齐祥, 陈志国. 一种鲁棒性的视频稳像算法. 光电子 激光(已录用).

### 二、参加的科研项目

1. 面向服务的架构技术在高招在线系统中的应用研究 (科技成果登记号: 9412005Y0515)
  2. 国家自然科学基金项目: 复杂环境下新型多目标动态检测、识别及跟踪算法研究.
  3. 中科院百人计划资助项目: 复杂环境下动态目标跟踪技术.
-