

山东大学

硕士学位论文

基于点特征的图像配准算法研究

姓名：赵辉

申请学位级别：硕士

专业：信号与信息处理

指导教师：陈辉

20060407

## 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科研成果。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律责任由本人承担。

论文作者签名：赵辉 日期：2006.4.15

## 关于学位论文使用授权的声明

本人完全了解山东大学有关保留、使用学位论文的规定，同意学校保留或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；本人授权山东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、缩印或其他复制手段保存论文和汇编本学位论文。

(保密论文在解密后应遵守此规定)

论文作者签名：赵辉 导师签名：陈辉 日期：2006.4.17

## 摘 要

图像配准是指找出场景中同一物体表面的结构点在不同图像上的投影像素点之间的对应关系。图像镶嵌指的是将多幅互有重叠的图像组合形成一幅宽视角图片的技术。目前图像配准和镶嵌广泛应用于虚拟现实、视频压缩、图像复原、图像数据库检索等技术中。图像配准的研究是计算机视觉中最困难也是最重要的任务之一。

不同的图像配准方法总是对应于某种适用的图像变换模型,总体可以分为三类方法:直接像素亮度差优化的方法、基于特征匹配的方法和变换域求解的方法。图像配准算法研究的核心问题是提高配准的速度、精度和算法的稳健度。

本文针对同一场景在不同角度下拍摄的重叠图像配准及镶嵌问题进行了研究,包括研究了窄基线条件下基于频域信息的配准技术和基于点特征匹配的配准技术;研究了图像配准与镶嵌技术中的一系列相关问题,包括摄像机运动模型、透视变换矩阵参数的鲁棒估计和求解方法、插值技术、图像变形技术、流形的概念等,并分析了影响合成图像清晰度的主要因素。

本文提出了一种快速、稳健的基于角点匹配的图像配准算法。该算法提出了两次导向匹配的思想,第一次使用相位相关法估计的全局粗平移运动参数指导局部的角点匹配,以降低无效搜索、提高匹配的速度和稳健性,第二次使用改进的RANSAC算法鲁棒估算的精确变换模型参数再次指导角点匹配,以获得更多的有效匹配角点;算法使用奇异值分解最小二乘法估算变换模型参数,并使用LM非线性优化方法进一步降低配准误差。这种全局参数指导局部匹配的思路提高了现有算法的匹配速度,并较好的降低了错配,提高了算法的整体稳健性和配准精度。实验结果表明,与现有算法相比,对于图像重叠区域较小、图像中含有运动物体、重复性纹理等较难自动匹配情况,使用本文算法可以有效提高算法的成功率。

本文还提出了一种基于亮度与白平衡自动调整的图像镶嵌算法。该算法中,首先使用本文提出的基于角点匹配的图像配准算法对图像进行几何配准,然后在对应点位置处利用各个色彩通道的灰度信息估算图像色彩变换函数以进一步实现图像的光度配准,最后进行无缝的图像镶嵌。实验结果表明该算法可以较好的融合

存在一定亮度和白平衡差异的图像。

另外本文还介绍了宽基线条件下的特征匹配技术，并研究了基于尺度空间和不变量技术的 SIFT 特征点匹配算法及其应用。

**关键词：**图像配准；相位相关法；点匹配；图像镶嵌；SIFT 特征匹配

## ABSTRACT

Image Registration is a technique that relates or aligns different images taken from different viewpoints, at different times, or by different image sensing devices. Image mosaic technology uses the result of image registration to build large view panorama-like image from small adjacent image series, which could enlarge the user's view scope and increase image resolution as well.

Till now, image registration and mosaic technology is widely used in the field of virtual reality(VR), video compression, image super-resolution, intelligent surveillance system, etc, which came to be an active research area in computer vision during recent years.

Feature-based, direct pixel difference optimization based, and Fourier based method are three typical ways of image registration, which have its own appropriate application area separately. The key problems focus on how to increase the algorithm speed, to increase the registration precision and to enhance the robustness of these registration methods.

This paper did research on overlaid-image registration and mosaicing technique and presented some new effective algorithms based on point-matching.

Firstly, the author did some research on frequency based and corner-matching based image registration technology used in short baseline correspondence condition. Then a new fast and robust image auto registration algorithm is presented. In this method, there is a new twice guided-matching scheme that first use phase correlation to estimate translational parameters to roughly guide local corner matching, and second use an improved RANSAC algorithm to estimate 8 parameters of perspective transformation to guide local corner matching again, with more precision. So much more valid corner matches could be quickly and stably generated. SVDLS (Singular Value Decomposition Least Square) is used to estimate transform parameters and Levenberg-Marquardt optimization is used to decrease registration error. The thought of guided matching mechanics is experimented to show better results than existing methods on matching speed, registration error, and the whole robustness of the algorithm.

Secondly, a simple, practical image auto stitching method based on brightness and white-balance automatic tuning is presented. Both of the geometric and photometric

registration techniques are employed in this method to efficiently estimate six linear parameters of the color transform model, after which the two images are well prepared for the next mosaicing step.

Thirdly, the invariant based registration technique is researched in this paper, and the author introduces the popular SIFT (Scale Invariant Feature Transform) based matching algorithm in the condition of wide baseline correspondence point matching technologies.

The algorithms in this paper are testified to be effective by lots of interesting, promising experimental results too.

**Keywords:** Image Registration; Phase Correlation; Point Matching; Image Mosaic; SIFT Feature Matching

## 第一章 绪论

本章首先阐述了图像配准的概念和图像配准的几类方法，然后分析了图像配准技术的应用领域及国内外研究现状，最后对本文的研究内容、研究重点和全文章节安排予以介绍。

### 1.1 课题背景介绍

近年来，随着数字化技术的不断发展，计算机处理能力不断提高、硬件装置价格的不断下降，各种先进实用的图像图形算法层出不穷，这使得计算机视觉技术变得更加切实可行。虚拟现实、模式识别、视频压缩等实用技术得到了广泛的关注。更多的算法开始由实验室走向市场。

我们使用光学成像系统可以从三维世界获得数字化的图像序列。对同一场景拍摄的图像序列实际上是真实的三维世界在不同时间向成像平面的一系列投影，图像帧与帧之间具有较大的相关性和信息冗余。这是图像配准研究的理论依据。

#### 1.1.1 图像配准技术简介

图像配准(Image Registration)是图像处理的一个基本问题，它源自于多个领域的很多实际问题，如不同传感器获得信息的融合；不同时间、不同条件下获得的图像的差异监测；成像系统和物体场景变化情况下获得的图像的三维信息获取；图像中的目标识别；多媒体数据库的检索等等。

简单来说，图像配准是将同一场景拍摄的不同图像进行对齐的技术，即找到图像之间的点对点映射关系，或者对某种感兴趣的特征建立关联。在中文文献中，图像配准也常被翻译作图像对齐、图像注册。

以同一场景拍摄而成的两幅图像为例。假如实际的三维世界点  $P$  在两幅图像中分别对应着  $p_1$  和  $p_2$  两个二维图像点。图像配准要做的就是找到  $p_1$  和  $p_2$  的映射关系，或者  $p_1$ 、 $p_2$  跟  $P$  的关系。 $p_1$  和  $p_2$  被称为对应点(Correspondence Points)、匹配点(Matching Points) 或控制点(Control Points)。

### 1.1.2 图像配准技术分类

Barbara Zitova<sup>[1]</sup>从应用的角度将图像配准问题概括为下面四类情况:

- (一) 同一场景从不同角度拍摄形成的不同图像的配准问题。其研究目的是获得更宽阔视野的图片, 获得立体信息、进行三维模型等。
- (二) 不同时间拍摄的不同图像的配准问题。其研究目的是检测并定位场景中的变化部分, 比如在遥感图像处理技术中检测地理环境是否发生变化, 在医学图像处理技术中检测患者的局部部位是否发生病变, 在自动视频监控系统中智能化的检测是否有入侵等。
- (三) 不同传感器所拍摄图像的配准与融合问题, 比如在 CT、SPECT、MRS 等医学图像系统中的应用。
- (四) 二维场景的图像跟三维模型的配准问题, 比如在 GIS 系统、目标识别、图像数据库检索系统中的应用。

无论所处理的图像是发生何种形式的变化、是用何种传感器去拍摄, 我们总是力求使用不变的部分、共性的信息去完成配准, 然后再根据需要进行处理变化的部分。这就是图像配准技术的核心思想。

### 1.1.3 图像配准系统的组成

一个典型的图像配准系统包括四个部分<sup>[1]</sup>:

#### 1.) 特征检测(Feature Detection)

这里的特征指的是广义的特征, 包括图像灰度、色度特征; 角点、边缘、轮廓等结构性特征; 频域信息、小波系数等<sup>[2]</sup>。

#### 2.) 特征匹配(Feature Matching)

即使用特征描述算子(即特征向量)和相似性策略对特征进行匹配。

#### 3.) 变换模型参数估算(Transform Model Estimation)

变换模型又称映射模型(Mapping Model), 即将输入图像向参考图像映射的坐标变换函数<sup>[1]</sup>。

#### 4.) 图像重采样与变换(Image Resampling and Transformations)

这一步就是我们由输入图像经变换模型向参考图像进行对齐的过程。变



换后图像的坐标将不再是整数，这就涉及到重采样与插值的技术。

## 1.2 图像配准应用领域与研究现状

图像配准是立体视觉、运动分析、数据融合等实用技术的基础，在导航、地图与地形配准、自然资源分析、天气预报、环境监测、生理病变研究等许多领域有重要的应用价值。国内外学者针对不同的图像配准应用问题进行了大量的研究工作，早在 1992 年英国剑桥大学的 Lisa Gottesfeld Brown 在文献<sup>[3]</sup>中就总结了图像配准的主要理论及图像配准在各个领域的应用。当时他讨论的图像配准技术主要还是着眼于医学图像处理、遥感图像处理等传统应用领域。

时隔十四年，图像配准技术已经有了飞跃的发展。下面本文作者将综述介绍一下当前图像配准技术在虚拟现实、视频压缩、图像复原等领域的应用，并着重介绍图像配准及静态图像镶嵌技术的国内外研究近况。

### 1.2.1 虚拟现实与静态图像镶嵌

虚拟现实(Virtual Reality)技术是一系列多媒体新技术的汇集，其目标是用计算机来生成一个逼真的三维世界以给观众如同真实世界的体验。近年来随着计算机技术的发展，虚拟现实技术已经深入到游戏设计、建筑设计、医学仿真、军事仿真等诸多应用领域<sup>[4]</sup>。虚拟现实的实现方法分为传统的基于三维建模的方法和基于图像绘制的方法。基于三维建模的方法首先要对场景中的所有物体建立三维几何模型和光照模型，然后利用透视投影原理（将在第二章介绍）将三维几何模型映射到二维成像空间。基于三维建模的方法的缺点是：对硬件要求较高，设备复杂；渲染速度受场景复杂度的限制；生成的模拟世界真实感不佳等。

基于图像的虚拟现实技术(Image-Based Virtual Reality, 简称 IBVR)的出现较好的解决了传统虚拟现实技术的这些不足。首先利用图像处理技术得到场景的全景图<sup>[5]</sup>，然后采用几何变换和插值的方法就可以生成新的任意角度的视图。

获取全景图的方法有两种，一种方法是基于硬件实现的系统，比如使用鱼眼镜头、扫帚式摄像机等设备直接获取全景图。这种方法造价昂贵，获取的全景图含有较大的变形失真，而且图像分辨率也受到限制。另一种思路就是使用图像镶

嵌的方法。图像镶嵌是指将多幅图像进行组合、拼接的技术：首先对周边场景进行拍摄以得到一系列的重叠图像，然后利用重叠区域的信息计算出图像间的变换模型关系，并将其镶嵌起来以形成全景图<sup>[5-8,10]</sup>。

图像镶嵌要考虑三个最基本的问题，即首先要确定图像坐标变换的数学表示模型，或者说摄像机运动模型；其次要根据某种方法来估算、求解该模型的参数；最后将多幅图像统一投影、变形到一个新的坐标系，并使用某种图像融合算法进行拼合。

图像配准是图像镶嵌技术的核心问题。微软研究院的 Richard Szeliski 在 1996 年 SIGGRAPH 上提出了基于运动模型的全景图拼接算法<sup>[5]</sup>。Szeliski 采用了非线性优化的方法来最小化像素两幅图像的亮度差以确定变换参数。该方法使用了全部像素进行优化处理，所以配准精度较高，但是计算速度较慢，且稳健性不佳。

国内的赵向阳、杜立民在 2004 年提出了一种基于特征点匹配图像自动拼接算法<sup>[8]</sup>，其中使用了 Harris 算法<sup>[9]</sup>提取角点并进行匹配。赵的算法采用了鲁棒变换估计技术，在一定程度上提高配准算法的稳健性，但是计算速度依然较慢，且无法配准重叠区域较小、运动物体较多的图像。M.Brown 在 2003 年 ICCV 大会上发表了一篇名为 Recognising Panoramas 的文章<sup>[10]</sup>，文中使用了基于不变量技术<sup>[11]</sup>的 SIFT 算法<sup>[12]</sup>进行图像配准，算法完全自动完成且效果较好。M.Brown 的大会发言再次掀起了全景图拼接技术研究的热潮。

块匹配算法是用于运动图像编码的一种常用算法，其实质也是对视频图像序列的相邻帧进行配准<sup>[13]</sup>。块匹配算法中，每帧图像被分成二维的  $M \times N$  像素的子块，假定每个子块内的像素都作平移运动。当前帧的  $M \times N$  子块在上一帧对应的子块邻域窗口内搜索到与之最匹配的子块，当前子块与匹配块在二维平面上的位移即为运动估计得到的运动矢量。Dae-Hyun Kim<sup>[14]</sup>提出了一种改进的块匹配算法并将其用于全景图拼接技术中。

另外，B.Reddy 和 B.Chatterji<sup>[15]</sup>中提出了一种基于 FFT 的图像配准方法，可以处理包含平移、缩放和尺度变化在内的图像配准问题；李忠新等在 2004 年提出了一种基于频域相关的柱面全景图拼接技术<sup>[16]</sup>。这些方法都是在变换域进行图像配准的例子。

综上所述,面向图像镶嵌的图像配准算法主要可以分为三类<sup>[17]</sup>:基于特征匹配的方法;直接像素亮度差最小优化的方法;在变换域求解的方法。

### 1.2.2 动态镶嵌与视频压缩

图像镶嵌分为静态镶嵌(Static Mosaicing)和动态镶嵌(Dynamic Mosaicing)两种。1.2.1节介绍的属于前者。静态镶嵌是对一系列相邻帧图像的共性信息即背景的描述,但无法对每个帧独有的动态信息进行更新。

动态镶嵌则常见于视频压缩与编码等技术中<sup>[18,19]</sup>。基于动态镶嵌的视频压缩原理是:在一个视频图像序列中,相连续的帧之间重叠较大,体现出较强的延续性和相关性。我们可以将实际场景表示为前景物体和从序列图像中提取并镶嵌而成的背景,在存储和传输视频的时候我们只处理背景 Mosaic 和每一帧的增量。而在需要复现某一帧的时候,则该序列中的任一帧可以由背景和增量的 Mosaic 完全表示,从而降低了信息冗余度,实现了视频数据的压缩。

动态镶嵌技术推动了视频压缩、图像检索、视频编辑和交互技术的发展。创建视频的动态镶嵌需要三个步骤<sup>[19]</sup>: (1)相邻帧的图像配准; (2)生成包含相邻一系列帧共性的背景 Mosaic; (3)计算每一帧跟 Mosaic 相比的增量。

### 1.2.3 图像超分辨率复原

图像的超分辨率复原(Image Super-Resolution)<sup>[20]</sup>指的是从低分辨率图像中构造出高分辨率图像的技术。超分辨率图像复原技术最初是以单张图像重构的概念和方法提出的。对于单帧图像重构问题,主要的解决思路都是围绕重采样和插值算法来进行的。然而,通过这些单帧插值方法的图像经过放大后,会出现一些平滑效应,图像的边缘细节通常不能被很好的重构出来。

而多帧图像重构技术的出现弥补了这一不足。该技术是通过充分利用各幅低分辨率图像提供的不完全相同的信息来提高图像的分辨率的<sup>[21]</sup>。在由低分辨率图像重构高分辨率图像的过程中,建立重构模型时,必须要知道各低分辨率图像彼此之间的微位移量和微旋转角的大小。这就需要采用亚像素(Sub-Pixel)精度的精确图像配准技术。精确图像配准算法的好坏直接影响图像复原的最终效果。

图像超分辨率复原技术有两个重要用途，一是可以通过软件手段提高已有图像的分辨率，二是在不降低图像品质和细节的前提下实现降噪。目前图像超分辨率复原技术在军事、医学和公安等领域具有十分重要的应用价值和良好前景。

#### 1.2.4 目标辨识与检索

图像识别和检索属于广义的图像配准范畴。在进行识别时，我们希望能找到一种对平移、缩放、旋转、光照变化、噪声等因素保持一定不变性、稳定性的图像局部特征算子，从而可以作为辨识物体的重要依据。

Schmid 和 Mohr 最早在 1997 年就提出了基于尺度空间和不变量理论的图像特征描述算子<sup>[12,22]</sup>—SIFT 算子(Scale Invariant Feature Transform)的概念，其含义是使用 SIFT 技术得到的图像特征向量对尺度的变化保持不变性。后来随着该算法的发展，SIFT 算子已不仅仅是对尺度变化保持不变性，而是扩展到对旋转、尺度缩放、甚至部分的三维视角变化和光照变化保持不变，从而成为稳定性、适应性较强的局部特征匹配算子。

Schmid 成功的将 SIFT 技术用于图像检索<sup>[22]</sup>，而 David G Lowe 则将其用于目标辨识<sup>[23]</sup>，M.Brown 则将 SIFT 特征匹配算法用于全景图自动拼接系统中<sup>[10]</sup>，均取得了不错的效果。

### 1.3 本文的研究重点及章节安排

本文主要研究窄基线（窄基线的概念将在第二章介绍）条件下基于点特征的图像配准问题及在图像镶嵌中的应用。本文的研究重点着眼于提高当前配准算法的速度、配准精度和稳健性。另外本文对宽基线条件下的图像配准问题以及基于不变量技术的点特征匹配问题也作了一定研究。

本文研究内容概括如下：

- 1) 系统研究了摄像机成像模型、坐标变换与摄像机运动，并论述了单应性矩阵成立的条件；
- 2) 研究了基于角点匹配和基于相位相关法的图像配准；
- 3) 研究了鲁棒的变换估计技术；

4) 介绍了图像插值算法、图像变形和流形技术;研究了图像镶嵌中的白平衡与亮度自适应调整技术;

5) 介绍了不变量技术,并研究了基于 SIFT 算子的特征匹配。

全文共分为六章,各章节内容安排如下:

第一章为绪论,阐述了图像配准技术的背景,介绍了图像配准技术的定义及分类,并概述了图像配准技术的应用领域及国内外研究进展。

第二章介绍了针孔摄像机模型及摄像机运动、坐标变换,介绍了窄基线和宽基线的概念,论证了 8 参数透视变换矩阵的应用条件,并概述了透视变换矩阵参数的求解方法。

第三、四章研究了窄基线情况下基于点特征的重叠图像配准技术,其中第三章阐述了图像特征的概念,并介绍了当前主流的边缘检测、角点检测和匹配方法,介绍了相位相关法估算平移参数的方法,然后提出了两次导向匹配的思想和改进的 RANSAC 算法,并将其用于角点匹配和变换模型参数估计。本章最后将给出本文算法与现有主流算法的对比实验。

第四章为图像变换与镶嵌,概述介绍了图像插值、重投影和图像融合技术,用实验结果分析了配准误差对图像镶嵌效果的影响,同时提出了一种白平衡与亮度自动调整的图像合成算法,然后介绍了影响图像合成清晰度的几个因素。本章的实验验证了算法的有效性。

第五章介绍了宽基线情况下基于 SIFT 算子的特征匹配,并用实验仿真结果进行分析讨论。

第六章总结了本文所作的工作,提出了仍然存在的问题,并对下一步的研究工作做了展望。

## 第二章 成像几何与图像变换模型

本章将阐述摄像机成像几何的基本原理, 然后介绍摄像机运动与图像配准的关系、窄基线与宽基线、图像变换的知识, 同时对 8 参数平面透视变换矩阵的存在条件进行了论证, 并简介了几类求解参数的方法。

### 2.1 摄像机成像几何基础

#### 2.1.1 针孔相机模型及摄像机坐标系

在一个成像系统中, 二维图像每一点的灰度值反映了空间物体表面某点反射光的强度, 而该点在图像上的位置则与空间物体表面相应点的几何位置有关。这些位置的相互关系由摄像机成像几何模型所决定<sup>[23,24]</sup>。在三维计算机图形学的研究中, 为了简化问题的处理同时又满足应用的需要, 常用针孔相机模型来代替实际的摄像机。针孔相机的成像几何关系称为透视投影(Perspective Projection)<sup>[25]</sup>。

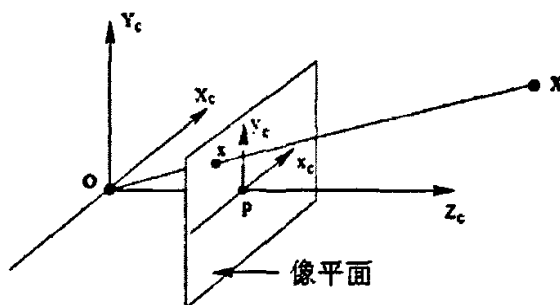


图 2-1 摄像机坐标系

首先定义摄像机坐标系  $O-X_cY_cZ_c$ , 如图 2-1 所示,  $O$  称为投影中心,  $OZ_c$  为薄透镜的光轴,  $p$  为像平面与光轴的交点,  $O-p$  的长度即为焦距  $f$ 。

三维空间中实际点  $X$  在摄像机坐标系内的空间坐标为  $(X_c, Y_c, Z_c)$ , 经过透视投影成像后形成二维像点  $x$  其坐标为  $(x_c, y_c)$ 。二者由透视投影公式关联:

$$\frac{x_c}{X_c} = \frac{y_c}{Y_c} = \frac{f}{Z_c} \quad (2-1)$$

式(2-1)中角标  $c$  表示这是在摄像机坐标系下的坐标。等号两边单位都是毫米。

$(x_c, y_c)$ 、 $(X_c, Y_c, Z_c)$ 属于欧氏坐标(Euclidean)。在欧氏坐标中，一个平面点由一个二维向量表示，一个三维空间点由一个三维向量表示。欧氏坐标因具备直观的特点而得到了广泛应用。为了更方便的表达透视投影公式及统一的描述平移、旋转等运动模型，我们引入齐次坐标的概念。齐次坐标增加了一个缩放因子  $w$ ，若一个二维点欧氏坐标为 $(x, y)$ ，则其对应的齐次坐标为 $(xw, yw, w)$ 。齐次坐标系下的坐标点具有一个很重要的性质就是缩放不变性。本章在描述变换公式时，如无特殊声明，则默认使用齐次坐标。齐次坐标下的透视投影公式如式(2-2)：

$$Z_c \begin{pmatrix} x_c \\ y_c \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} \quad (2-2)$$

### 2.1.2 世界坐标系与像平面坐标系

摄像机坐标系可以简明扼要的描述针孔相机模型的本质，但这还远远不够。完整的描述一个三维世界点到二维像素点的映射还需要知道摄像机内、外部参数，即需要知道摄像机坐标系跟像平面坐标系、世界坐标系的关系。

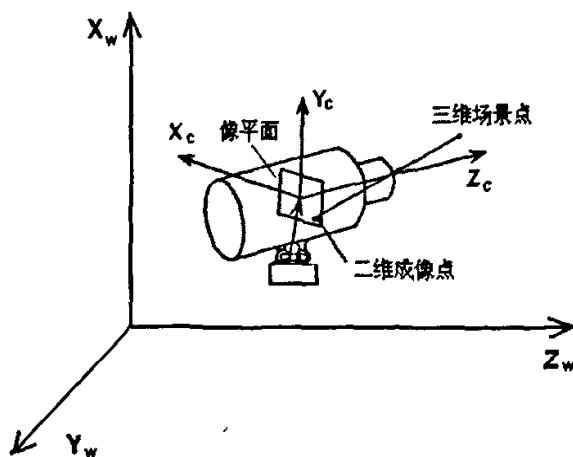


图 2-2 世界坐标系

世界坐标表示场景点在客观世界的绝对坐标。如图 2-2 所示，摄像机以某种姿态放置于一个实际的三维世界中。摄像机坐标系  $X_c Y_c Z_c$  与世界坐标系  $X_w Y_w Z_w$  不一定完全重合，而是具备一定的平移和旋转关系。

齐次坐标表示下坐标系的平移、旋转可以用一个  $4 \times 4$  的矩阵来表示:

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = \mathbf{M} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \quad (2-3)$$

式(2-3)中, 左边是摄像机坐标系下的齐次坐标, 公式右边是一个  $4 \times 4$  的变换矩阵乘以世界坐标系下的齐次坐标。其中  $\mathbf{R}$  是  $3 \times 3$  的正交单位矩阵, 代表旋转量,  $\mathbf{t}$  是  $3 \times 1$  的平移向量,  $\mathbf{0}=(0,0,0)$ 。

接下来作者来介绍一下像平面坐标系。像平面坐标系反映了摄像机的内部构造对最终成像结果的影响。假设摄像机坐标系下的以毫米为单位的图像点坐标  $(x_c, y_c)$ , 其最终形成的图像像素为  $(x, y)$ 。

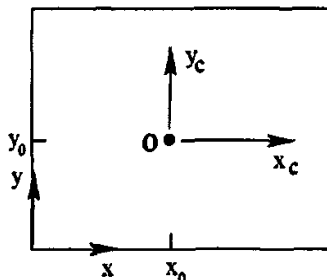


图 2-3 像平面坐标系

图 2-3 为像平面坐标系, 其中点  $\mathbf{O}$  定义在摄像机光轴和图像平面的交点。

$$\frac{x_c}{dx} = x - x_0 \quad \frac{y_c}{dy} = y - y_0 \quad (2-4)$$

像平面坐标系跟摄像机坐标系的关系如在式(2-4)所示, 其中  $dx, dy$  代表每毫米物理尺寸对应的像素个数。将(2-4)表示成齐次坐标形式如式(2-5)所示:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & x_0 \\ 0 & \frac{1}{dy} & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_c \\ y_c \\ 1 \end{pmatrix} \quad (2-5)$$

我们将式(2-2)、(2-3)、(2-5)联立可得式(2-6):



$$\begin{aligned}
 Z_c \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} &= \begin{bmatrix} \frac{1}{dx} & 0 & x_0 \\ 0 & \frac{1}{dy} & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \\
 &= \begin{bmatrix} a_x & 0 & x_0 & 0 \\ 0 & a_y & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}
 \end{aligned} \tag{2-6}$$

其中  $a_x = f/dx$ ,  $a_y = f/dy$ 。将式(2-6)简记为:

$$Z_c \mathbf{x} = \mathbf{K} [\mathbf{R} | \mathbf{t}] \mathbf{X}_w = \mathbf{P} \mathbf{X}_w \tag{2-7}$$

式(2-7)中,  $\mathbf{X}_w = [X_w, Y_w, Z_w, 1]^T$  是齐次坐标表示的三维世界点坐标,  $\mathbf{x} = [x, y, 1]^T$  是齐次坐标表示的二维平面点的像素坐标。  $Z_c$  为  $\mathbf{X}_w$  的深度, 为该点到焦平面的距离。  $\mathbf{P}$  是  $3 \times 4$  的三维投影矩阵, 由  $\mathbf{K}$  和  $[\mathbf{R} | \mathbf{t}]$  连乘得到, 其中:  $\mathbf{K}$  完全由  $a_x, a_y, x_0, y_0$  决定, 其只与摄像机内部结构有关 (包括摄像机焦距、镜头纵横比、图像中心坐标), 我们称这些参数为摄像机内参数(Camera Intrinsic Parameters);  $[\mathbf{R} | \mathbf{t}]$  完全由摄像机相对于世界坐标系的方向和位置决定, 称为摄像机外参数(Camera Extrinsic Parameters)<sup>[24]</sup>。摄像机内、外参数可以由摄像机标定(Camera Calibration)技术求得<sup>[26]</sup>。

至此, 根据式(2-7), 如果已知摄像机的内、外参数, 对于任何空间点  $\mathbf{X}_w$  就可以唯一确定它的图像点的位置  $\mathbf{x}$  (其中  $Z_c$  可以联立去掉)。这就是针孔相机几何成像原理的完整表述。

## 2.2 摄像机运动

### 2.2.1 摄像机运动的基本形式

一般摄像机运动可以分为以下几种基本运动形式<sup>[6,25]</sup>:

- ◆ 平移运动(Translation), 即摄像机的运动平行与成像平面。如图 2-4(a)
- ◆ 镜头缩放运动(Zooming), 即焦距发生变化, 或者相机作延光轴 (即  $z$  轴) 方向的运动。如图 2-4(b)

- ◆ 水平扫动(Panning), 即摄像机绕  $y$  轴转。如图 2-4(c)
- ◆ 垂直扫动(Tilting), 即摄像机绕  $x$  轴转。如图 2-4(d)
- ◆ 旋转运动(Rolling), 即摄像机绕光轴转。如图 2-4(e)

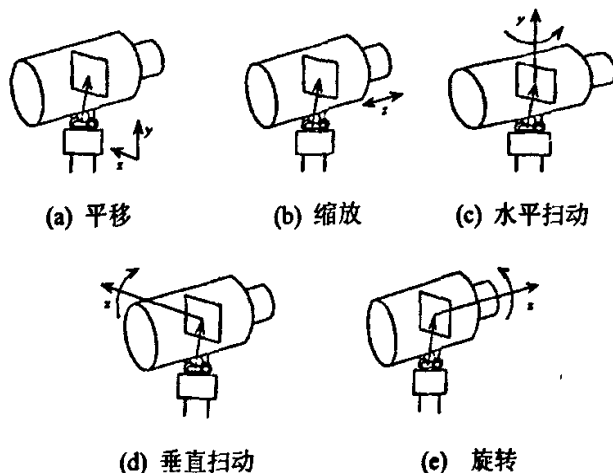


图 2-4 五种基本摄像机运动

### 2.2.2 摄像机运动与图像配准

实际的摄像机运动一般是上述五种基本运动的组合。图像配准问题其实就是研究在某种摄像机运动下拍摄的不同图像的图像模型参数求解问题, 或者说是摄像机运动参数估计问题。

假设场景点的三维世界坐标  $X_w$  及两次成像的透视投影矩阵  $P_1$ 、 $P_2$  已知, 则可分别确定成像点  $x_1$ 、 $x_2$ 。由于  $x_1$  和  $x_2$  对应于同一三维世界点  $X_w$ , 所以称之为 一对匹配点、对应点。两幅图像的匹配点是位于重叠区域内的像素点。

图像配准是上述过程的逆向过程, 即根据同一三维场景的两幅或多幅图像确定对应点并由此计算出图像变换模型, 甚至根据匹配点及其他约束条件进而重建出三维场景<sup>[27]</sup>。后者属于立体视觉的研究范畴。

## 2.3 图像变换模型

图像变换模型是指两幅二维图像所具备的坐标变换关系。在某种约束的摄像机运动条件下, 三维场景形成的两幅或多幅图像之间的关系可以完全由图像变换

模型描述。本文作者所做的研究都是围绕着某种特定的图像变换模型展开的。

### 2.3.1 透视变换模型的论证

本节将论证满足以下两个条件之一的情况下，两幅图像之间的对应关系可以由一个  $3 \times 3$  的平面透视变换矩阵(Planar Perspective Transform)来表示<sup>[24,28,29]</sup>：

➤ 1. 相机绕光心转动时拍摄的任意三维场景（如图 2-5(a)所示）

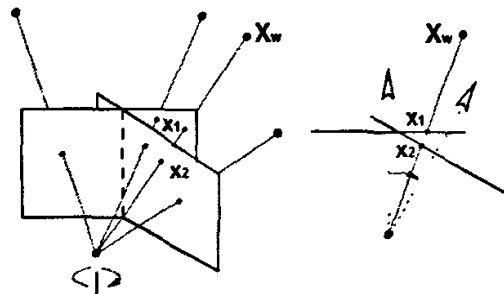
这里三维空间是可以任意分布的，但要求摄像机镜头严格绕光心旋转拍摄：即可以进行水平扫动、垂直扫动、旋转及镜头缩放，但不可以发生相机平移。

近似条件：将摄像机近似固定于一点进行拍摄。

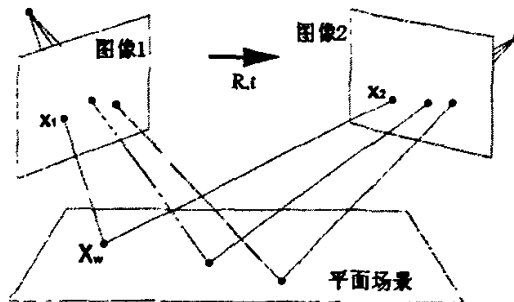
➤ 2. 相机以任意运动方式运动拍摄的平面场景（如图 2-5(b)所示）

平面场景是指不具备立体信息的类似于平面的场景，比如一堵墙、一幅油画等。

近似条件：场景深度远大于焦距时。



(a) 相机绕光心转动拍摄的任意三维场景



(b) 相机以任意方式拍摄的平面场景

图 2-5 透视变换矩阵成立的两个条件

下面将简单的证明一下这两个条件。

对于第一种情况，假设世界坐标系和第一幅图像的摄像机坐标系重合，则第二幅图像的摄像机坐标系与世界坐标系的关系可以用旋转矩阵  $R$  和平移向量  $t$  来表示。将两个摄像机的透视投影公式表示如下：

$$P_1 = K_1[I | 0] \quad (2-8)$$

$$P_2 = K_2[R | t] \quad (2-9)$$

式(2-8)和(2-9)中， $K_1$ 、 $K_2$  为  $3 \times 3$  摄像机内参数矩阵， $I$  为  $3 \times 3$  的单位阵， $R$  为  $3 \times 3$  旋转矩阵， $t$  为三维平移向量， $0=[0,0,0]^T$ 。由式(2-7)、(2-8)、(2-9)有：

$$Z_{c1}x_1 = K_1[I | 0]X_w \quad (2-10)$$

$$Z_{c2}x_2 = K_2[R | t]X_w = K_2R \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} + K_2t \quad (2-11)$$

$$\text{由(2-10)得: } Z_{c1}K_1^{-1}x_1 = [I | 0]X_w = \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} \quad (2-12)$$

$$\text{联立(2-11)和(2-12)得: } Z_{c2}x_2 = Z_{c1}K_2RK_1^{-1}x_1 + K_2t \quad (2-13)$$

根据式(2-13)，若三维场景静止，只要摄像机绕光心转动拍摄，就有  $t=0$ ，得：

$$Z_{c2}x_2 = Z_{c1}K_2RK_1^{-1}x_1 \quad (2-14)$$

记  $H=Z_{c1}K_2RK_1^{-1}/Z_{c2}$ ，则式(2-14)可以表示为：

$$x_2 = Hx_1 \quad (2-15)$$

上式中  $H$  是一个  $3 \times 3$  的满秩矩阵，称为平面透视变换矩阵，简称透视变换矩阵，又称单应性矩阵(Homography)。可见对一任意三维场景以除平移之外的运动方式进行拍摄所形成的图像像素点坐标满足式(2-15)的单应性矩阵变换关系。

对于第二种情况，假设所拍摄的三维场景点都位于同一平面  $\Pi$  上，其平面方程为  $n^T X_w = d$ ，其中  $n$  为该平面的单位法向量， $d$  为坐标原点到该平面的距离。由式(2-13)得：

$$x_2 = (H + Z_{c1}K_2t n^T K_1^{-1} / Z_{c2})x_1 = H_{\Pi}x_1 \quad (2-16)$$

由式(2-16)可见,对平面场景以任意运动方式进行拍摄所形成的两幅图像也满足单应性矩阵关系,这里不再要求平移向量  $t$  为零。

在实际拍摄过程中,当拍摄时有微小的相机平移时,或者当所拍摄的三维场景很远时(远大于焦距),都可以认为近似满足透视变换模型<sup>[29]</sup>。

### 2.3.2 透视变换矩阵求解方法概述

透视变换矩阵  $H$  各参数的求解是各种图像配准算法研究的关键内容,也是本论文研究的核心任务。求解  $H$  的方法可以分为三类,即直接像素亮度差优化的方法、基于特征的方法和基于频域的方法,这在绪论中已作了简要介绍。直接亮度差优化的方法通过优化技术来最小化像素的亮度差以求解  $H$ ,其特点是可以达到较好的精度,而且普遍适用于所有的成像条件,但是计算速度较慢,稳健性不佳,难以处理复杂的情况比如遮挡、光照变化等,另外很多时候往往需要手动选择参数以确定优化初始值;基于特征的方法则首先检测图像的特征,将特征进行匹配,然后利用特征对应关系(比如对应特征的坐标位置信息)来计算  $H$ ,其优点是计算速度较快,可以处理较大的几何变形,而且可以利用基于统计的去外点技术以去除错误的、不期望的特征匹配,对光照变化、运动物体、遮挡等复杂情况有较强的适应能力,从而可以实现鲁棒性较好的配准,而精确、稳健的特征匹配是该方法的难点所在,同时某些场景图像提取特征较为困难;基于频域的方法是将图像变换到频域进行配准,该方法速度较快,但其应用条件和配准精度往往受到限制。

本文主要研究基于点特征匹配的图像配准算法。若未作特别声明,则假设所拍摄的图像满足单应性矩阵成立的两个近似条件之一。

### 2.3.3 其他图像变换模型

如前面分析,单应性矩阵  $H$  为  $3 \times 3$  的矩阵,将其表示如式(2-17)所示:

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (2-17)$$

根据齐次坐标的性质  $h_{33}$  可以归一化为 1, 即  $H$  的自由度为 8, 其中  $h_{11}$ 、 $h_{12}$ 、 $h_{21}$ 、 $h_{22}$  是缩放、旋转因子;  $h_{13}$ 、 $h_{23}$  分别是水平、竖直方向的平移因子,  $h_{31}$ 、 $h_{32}$  是仿射变换因子。表 2-1 描述了各个参数的详细意义。表 2-1 中的各种图像变换都属于透视变换的子变换。

透视变换	$h_{11}$	$h_{12}$	$h_{13}$	$h_{21}$	$h_{22}$	$h_{23}$	$h_{31}$	$h_{32}$
平移变换	1	0	$x_0$	0	1	$y_0$	0	0
旋转变换	$\cos \theta$	$-\sin \theta$	0	$\sin \theta$	$\cos \theta$	0	0	0
缩放变换	$s_x$	0	0	0	$s_y$	0	0	0
相似变换	$r \cos \theta$	$-r \sin \theta$	$x_0$	$r \sin \theta$	$r \cos \theta$	$y_0$	0	0
仿射变换	$h_{11}$	$h_{12}$	$h_{13}$	$h_{21}$	$h_{22}$	$h_{23}$	0	0

表 2-1

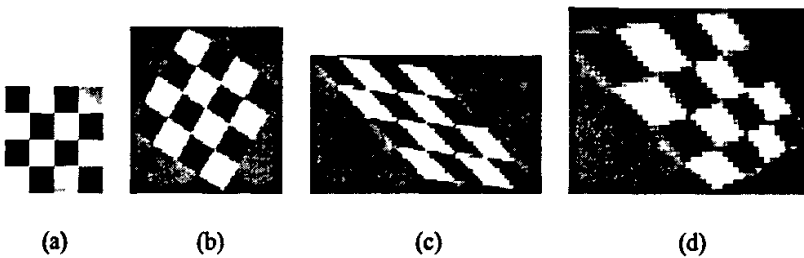


图 2-6 图像变换示例

下面作者以一幅简单的棋盘图像为例说明一下透视变换及其子变换。在图 2-6 中, 图(a)为棋盘的原图。图(b)是将该图进行相似变换的结果, 图(c)是仿射变换的结果, 图(d)是透视变换的结果。在透视变换下直线的平行关系不再保持。

相似变换指的是两幅图像只具备平移、旋转和等比例缩放的关系, 图像经相似变换后, 物体的形状保持不变而大小尺寸可以发生变化, 比如卫星拍摄的地球图像、扫描仪对一幅大图像多次扫描得到的图像均可认为满足相似变换关系。相似变换有 4 个自由度, 仿射变换是 6 个, 透视变换是 8 个。

## 2.4 宽基线与窄基线介绍

基线的本意是指立体视觉系统中两摄像机光心之间的距离。依据拍摄两幅图像的视点位置关系可将对应点匹配问题分为宽基线匹配(Wide Baseline)和窄基线

匹配(Short Baseline)。宽基线一词用于匹配时,泛指两幅图像有明显不同的情况下的匹配。产生这种情况的原因有可能为摄像机之间的位置相差很大,也有可能由于摄像机旋转或焦距的变化等因素产生的<sup>[30]</sup>。

宽基线匹配和窄基线匹配的分界不是很严格,但是在窄基线匹配中存在如下假设:摄像机焦距及其它内参数变化不大;摄像机位置不会相差很远,不会有大的转动,对应点的邻域是相似的。

宽基线匹配中则存在如下假设:对图像上的任意点,在另一图像上的对应点可以为任意位置;摄像机可以任意移动,且摄像机的焦距及其它内参数可以有较大的变化;一幅图像上的景物在另一幅图像上可能被遮挡;对应点的邻域有相似的地方,但由于摄像机位置的变化及光照的变化,单依靠邻域的相似不能得到正确的对应。

窄基线匹配中典型方法是利用邻域的互相关(Neighborhood Cross-Correlation)方法。但在宽基线的情况下,图像之间拍摄距离较远,成像条件存在较大差异,即使是空间同一特征,在图像中所表示出来的光学特性(灰度值,颜色值等)、几何特性(外形,大小等)及空间位置(图像中的位置,方向等)都有很大的不同,再加上噪声、遮挡等因素的存在,此时基于邻域互相关的匹配方法就失效了。

在宽基线匹配中,仅仅使用特征本身的信息(比如边缘、角点的位置信息)是难以正确匹配的,研究学者将多个特征尤其是结构性特征予以组合,以形成稳定的特征向量(称为特征描述符)。这种对于图像的几何变形、光照变化等因素保持一定稳定性的特征向量称为不变量(Invariant)。不变量技术是宽基线匹配应用中的重要技术。

本文第三、四章主要研究了窄基线条件下的图像配准。第五章则介绍了宽基线条件下基于不变量技术的点特征匹配问题。

### 第三章 基于角点特征的图像配准

本章第 3.1 节概要介绍了目前两类主流的图像配准算法, 即 Szeliski<sup>[5]</sup>基于最小像素亮度差优化的方法和赵向阳<sup>[8]</sup>基于特征的方法, 并分析了各自利弊, 总结了难点所在; 3.2 节介绍了图像特征检测的基础, 并着重介绍了 Harris 角点检测与匹配方法; 3.3 节介绍了作者提出的改进方法; 3.4 节为实验结果。

#### 3.1 现有算法概要

Szeliski<sup>[5]</sup>算法中, 假设图像  $I_1$  中的点  $(x_{1i}, y_{1i})$  其在图像  $I_2$  中的对应点为  $(x_{2i}, y_{2i})$ , 设  $E$  来表示两幅图像对应点像素的亮度差, 如式(3-1)所示:

$$E = \sum_i [I_2(x_{2i}, y_{2i}) - I_1(x_{1i}, y_{1i})]^2 = \sum_i e_i^2 \quad (3-1)$$

由式(2-15)可知,  $I_1$  和  $I_2$  中对应点之间的关联关系是平面透视变换矩阵。Szeliski 使用了高斯金字塔和 Levenberg-Marquardt 非线性优化算法 (简称 LM 优化)<sup>[31]</sup>将式(3-1)最小化处理就可以得到  $h_{11} \sim h_{32}$  这 8 个参数。该方法的本质是使用非线性优化方法对一个全局能量代价函数进行优化处理, 其优化目标是全部像素。该方法的优点是不需要提取特征点, 且配准精度较高。但其缺点也是致命的, 比如 LM 优化算法收敛速度较慢, 通常需要有良好的初始值并经过多次迭代才能得到一个趋于稳定的解, 在很多时候还面临陷入局部极值的危险。Szeliski 建议通过手工选取一系列匹配点来确定初始值的, 这就更加加重了算法的整体时间开销。另外该算法无法处理遮挡、几何形变、较多的运动物体等等复杂情况。

国内赵向阳<sup>[8]</sup>算法属于基于点匹配的图像配准方法。这种方法的典型思路为: 提取特征点; 通过特征点邻域的灰度互相关进行匹配; 利用图像之间的几何限制拒绝错误匹配; 使用最小二乘法估算变换模型参数。

赵的算法在一定程度上提高了以往配准算法的稳健性, 但其仅针对于重叠区域较大、易产生较多角点的静止建筑物图像做了研究, 其算法整体速度一般, 而且在处理重叠区域较小、含运动物体、含重复性纹理的图像配准时往往会失败。



同时赵的算法提取出的匹配角点数目不多，配准精度也不是很理想<sup>[8]</sup>。

作者总结了实际图像配准过程中的主要难点如下：

- ◆ 可能存在遮挡，一个三维点只出现于一幅图像中而在另一幅图像中没有；
- ◆ 如果使用不属于主运动的独立运动物体上的对应点来进行匹配则会导致变换模型参数计算错误；
- ◆ 图像中含有水波纹、草木等重复性较大的纹理特征时，会出现多个局部特征(Local Features)<sup>[32]</sup>相对应的情况，降低了局部特征匹配的可靠性；
- ◆ 当图像重叠区域较小时，特征匹配变得困难且不稳定，同时有效匹配点数量较少从而难以准确的估算变换参数；

## 3.2 图像特征概述

图像特征包括像素灰度特征、色彩特征、区域特征、纹理特征、轮廓特征、边缘特征、角点特征等<sup>[32]</sup>。边缘特征和角点特征是两种较常用的特征。

### 3.2.1 边缘特征

边缘指的是图像中在且仅在某个方向上灰度急剧变化，而在另一个与其正交的方向灰度变化平缓的像素<sup>[33-35]</sup>。边缘特征可以更好的保持图像的结构信息，同时很多角点提取算法也是基于边缘轮廓特征的。

在一维连续信号中，边缘定义为一阶导数最大或者二阶导数为零的地方。在二维连续图像 $f$ 中，边缘是用图像梯度来定义的。梯度公式如式(3-2)所示。

$$\nabla f = \left[ \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right] \quad (3-2)$$

边缘的方向与梯度方向垂直，边缘的强度决定于梯度的模值。

而在数字图像中，像素是离散的，此时我们使用两种思路来求梯度。第一种思路是用具有差分性质的算子来近似代替求导运算，如 Sobel、Prewitt、Roberts 边缘检测算子<sup>[33]</sup>。第二种方法是使用参数化模型方法，由离散图像拟和一个连续图像进而求梯度<sup>[35]</sup>，进而边缘位置所在。

下面介绍一下目前最常用的 Canny 边缘检测算法<sup>[34]</sup>。Canny 提出了著名的边

缘检测三大评判标准,同时推导出在连续信号情况下边缘检测的最优解,并开发了一种实用的算法。该算法大体思路是:先对图像进行高斯平滑以滤除噪声,然后沿着梯度方向进行非极大值抑制,并使用累计直方图计算两个阈值。边缘的最终判定:凡是大于高阈值的则认为是边缘;凡是小于低阈值的则认为不是边缘;如果大于低阈值但又小于高阈值,那就要看这个像素的邻接像素中有没有超过高阈值的边缘像素:如果有的话那么就认为是边缘,否则就不是边缘。通过修改高斯平滑窗口的大小和更改两个阈值可以实现不同的边缘检测和噪声抑制能力。当高斯平滑窗口较大时,噪声被较大程度上抑制,检测出的边缘较纯净,噪声少,但定位不准,同时损伤了细节信息;而当高斯平滑窗口较小时正相反。

### 3.2.2 角点特征

角点是图像的另一个重要的局部特征,其直观定义是指在至少两个方向上图像灰度变化均较大的点。在实际图像中,轮廓的拐角、线段的末端等都是角点。角点特征因具有信息量丰富、便于测量和表示、能够适应环境光照变化、尤其适用于处理遮挡和几何变形问题等优点而成为许多特征匹配算法的首选<sup>[2]</sup>。

角点检测方法分为基于直接像素灰度比较的方法和求轮廓弧度极值的方法。Smith 于 1995 年提出了一种非线性的边缘、角点提取方法—SUSAN 算法<sup>[37]</sup>(即 Smallest Univalued Segment Assimilating Nucleus,在中文文献中被翻译作“最小同值吸收核”)。该算法包括五步:1,在像素周围放置一圆形模板;2,计算同值区域(USAN);3,设定一阈值并减去 USAN 区域大小以得到角点强度图;4,分析 USAN 区域的重心及连接性以去掉假角点;5,使用非极大值抑制方法来寻找最终角点。SUSAN 算法对噪声不敏感,但其计算较复杂。

角点检测的第二类方法是,首先对图像求边缘或轮廓,比如使用 Canny 算法;然后求该轮廓上的局部弧度最大值就可以确定角点的位置<sup>[38]</sup>。但这种方法往往具有定位不准的弊病,其角点定位能力非常依赖于边缘检测的结果。

下一节将介绍另一重要的角点检测算法:Harris 算法。

### 3.2.3 Harris 角点检测

Harris 算法的前身是 Morave 算法<sup>[9]</sup>。Morave 提出的角点检测公式为：

$$E(u, v)|_{(x, y)} = \sum_{x, y} w(x, y) [I(x+u, y+v) - I(x, y)]^2 \quad (3-3)$$

$E$  的含义是在点  $(x, y)$  处移动一个  $(u, v)$  小窗口所发生的亮度变化值。式(3-3)的本质其实就是二维信号的自相关。其中  $w(x, y)$  是高斯平滑因子。

将上式进行泰勒级数展开，并忽略高阶项得到：

$$E(u, v)|_{(x, y)} \cong [u, v] \mathbf{M} \begin{bmatrix} u \\ v \end{bmatrix} \quad (3-4)$$

$$\mathbf{M} = \sum_{x, y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (3-5)$$

这里  $I_x$ 、 $I_y$  分别代表图像中该点在水平、竖直方向上的导数。

将  $\mathbf{M}$  相似对角化处理后得：

$$\mathbf{M} \rightarrow \mathbf{R}^{-1} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \mathbf{R} \quad (3-6)$$

上式中， $\lambda_1$  和  $\lambda_2$  是四元矩阵  $\mathbf{M}$  的特征值。每个像素对应着一个这样的四元矩阵。把  $\mathbf{R}$  看成旋转因子，这样就可以仅由特征值  $\lambda_1$  和  $\lambda_2$  来分析亮度变化量：当二者都较小时，定义为平坦区域；当二者只有一个较大、而另一个较小时，定义为边缘；当二者都较大时，定义为角点。将这一思想具体用公式表述：

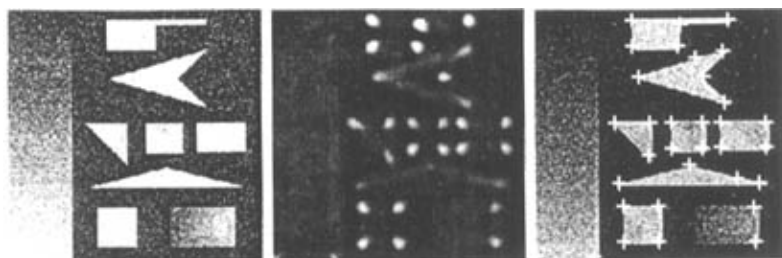
$$\text{Cornerness} = \det \mathbf{M} - k(\text{trace} \mathbf{M})^2 \quad (3-7)$$

其中：

$$\begin{aligned} \det \mathbf{M} &= \lambda_1 \lambda_2 \\ \text{trace} \mathbf{M} &= \lambda_1 + \lambda_2 \end{aligned} \quad (3-8)$$

上式  $\det$  和  $\text{trace}$  分别表示求行列式和迹。系数  $k$  一般取 0.04~0.2。当  $\text{Cornerness}$  大于某一阈值  $\text{threshold}$  且在某邻域内取得局部极值时，则标记该点为候选角点。

下图是 Harris 角点检测算法的效果，其中图 3-1(a)为加入强高斯噪声的标准角点测试图像，图 3-1(b)为  $\text{Cornerness}$  形成的图像，图 3-1(c)为角点检测结果。可见 Harris 算法可以检测出大部分主要的角点，具备一定的抗噪声能力。



(a) 加入高斯噪声的图像 (b) Cornerness 图像 (c) 最终角点检测结果

图 3-1 Harris 角点检测

### 3.2.4 角点匹配

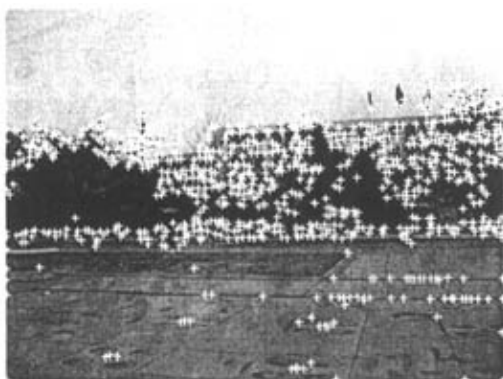
所谓角点匹配是指找出图像  $I_1$  和  $I_2$  中的唯一对应角点。窄基线条件下角点匹配常用的方法是归一化相关(Normalised Cross-Correlation, NCC)法。该方法依据角点邻域像素灰度值的相似性来进行匹配。设  $W_1$  和  $W_2$  是分别以  $I_1$  的角点  $p_1$  和  $I_2$  的角点  $p_2$  为中心的两个大小相同的窗口,  $u_1, u_2$  为相关窗口内像素灰度的均值,  $(x_i, y_i)$  的取值范围为窗口大小, 则定义 NCC 公式为:

$$NCC = \frac{\sum_i (W_1(x_i, y_i) - u_1)(W_2(x_i, y_i) - u_2)}{\sqrt{\sum_i (W_1(x_i, y_i) - u_1)^2 \sum_i (W_2(x_i, y_i) - u_2)^2}} \quad (3-9)$$

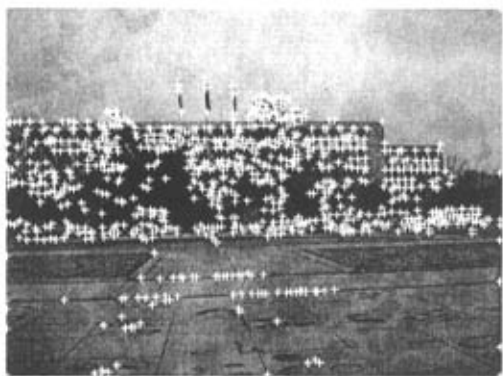
NCC 匹配方法去除了低频背景亮度信息的干扰。进行角点匹配时, 首先对于图像 1 中的任一角点, 寻找在图像 2 中和它相关性最大的角点; 然后对于图像 2 中的任一角点, 寻找在图像 1 中和它相关性最大的角点, 当双向搜寻到的最大相关性的角点彼此对应且相关性大于某一阈值时, 即认为找到一对候选匹配点。

下面是三组实验图像的角点匹配结果。作者在本文实验中使用的图像均为佳能 S50 数码相机手持拍摄, 其水平、竖直分辨率在 500~800 之间。拍摄时拍摄者位于固定位置手持相机水平扫动拍摄, 满足窄基线和透视变换模型的条件。

第一组图像是拍摄于校园的两幅建筑物图像, 即附图 (一) 中的前两幅图。



(a) 图 1 中检测出的角点



(b) 图 2 中检测出的角点



(c) 匹配点连线

图 3-2 校园照片的角点匹配

图 3-2(a)中检测出 916 个角点，图 3-2(b)中检测出 885 个角点，这样总共需要做 1621320 次 NCC 匹配，最终匹配的角点数目是 431 个。将每对匹配点的坐标

首尾连成一条直线，所有的连线画在同一坐标系中的，每一根连线代表一对匹配点，即得图 3-2(c)。匹配点连线的‘主流’代表了相机的水平扫动，称为主运动(Dominant Motion)。符合主运动的点被认为是有效匹配点，是内点(inliers)。同时图中也存在大量跟主运动相背的连线，这是由物体的独立运动(Independent Motion)和噪声等原因造成的，这些连线代表了错误的匹配，属于外点(outliers)。

在这组实验图片中，有效匹配点的比率约为 70%，大部分匹配都是正确的。

第二组图像是站在济南市经十路过街天桥拍摄的车流图像，即附图（二）中的图片。跟第一组图像不同，这两幅图像中存在大量的运动物体—车流和行人。

在这一组实验图片中，图 1 检测出 1339 个角点，图 2 检测出 1000 个角点，最终匹配点为 389 个，其中有效的匹配大约占 45%。由图 3-3 可见由于运动物体的存在而导致出现了大量错配，几乎淹没了主运动的方向。

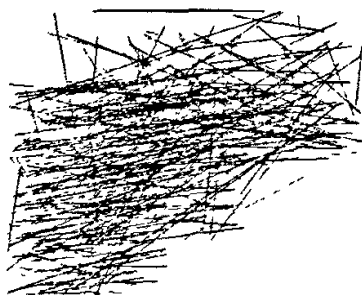


图 3-3 车流图片的角点匹配连线图

第三组实验图片拍摄于济南市植物园的图片，即附图（三）的第 2、3 幅图。这一组图片中不仅含有运动的船只，而且含有荡漾的水波纹、飘动的树叶这些纹理性物体，这使得特征匹配更加困难。

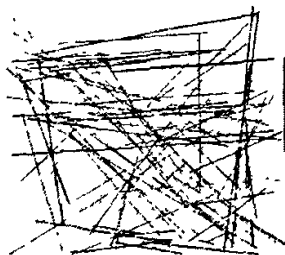


图 3-4 植物园照片的角点匹配连线图

在本例中, 其中图 1 检测出 393 个角点, 图 2 检测出 346 个角点, 最终得到 128 对匹配角点, 有效匹配率仅约为 30%。图 3-4 是匹配点连线图。可见错误的匹配几乎超过了正确匹配的数量, 主运动的方向被错配完全淹没。

### 3.3 改进的图像配准算法

下面总结一下角点匹配算法需要解决的问题:

- 1.) 待匹配图像中含有运动物体时会造成错配, 因为运动物体的存在使得图像像素已无法全部满足同一种变换关系, 即出现了主运动和从运动的差别。虽然使用后面介绍的 RANSAC 算法可以在一定程度上去除错配, 但 RANSAC 算法对输入匹配点外点的比率是有要求的, 过多外点的存在将导致算法稳健性和速度急剧降低。
- 2.) 待匹配图像中含有重复性纹理时, 由于角点匹配算法属于局部匹配方法, 它无法处理两个局部特征相近的物体, 比如大楼上的一排窗户都是一样的, 用该方法匹配就会发生混乱—虽然局部来看匹配了, 但却没有顾及到全局信息。如果使用飘动的树叶、水波纹这类纹理性运动物体的作为匹配特征, 则会发生大量的错配。
- 3.) 由于无法确定重叠区域, 角点匹配会在两幅图像的全部区域内进行遍历式搜索, 时间开销大, 而且无效匹配难以避免, 甚至重叠区域之外的点也可能被匹配。相关运算是整个算法中最为耗时的步骤之一。
- 4.) 算法对图像之间重叠区域的要求较高, 一般要求重叠区域大于 50% 以获得数目充足的匹配点, 这就很难做到实用性。

针对以上分析, 本文提出了导向角点匹配的思想, 即一种全局匹配指导局部匹配的方法。导向匹配指的是根据已知的主运动参数来指导角点匹配的局部搜索过程, 不仅很大程度上减少了无效的搜索, 也减少了局部独立运动导致的错配; 改进的 RANSAC 算法和两次导向匹配算法有效提高了正确匹配点的数量; 算法最后对透视变换模型参数继续进行非线性优化以进一步降低配准误差。

下面作者具体介绍一下这个算法。

### 3.3.1 相位相关法简介

信号在时域的时移对应于频域的线性相移；信号在时域的卷积对应于频域的相乘。这两个性质是基于频域的图像配准方法的重要理论依据。

相位相关法是一种基于傅氏功率谱的频域相关技术<sup>[39]</sup>，该方法利用了互功率谱的相位信息进行配准，对图像间的亮度变化不敏感，具有一定的抗干扰能力。

假设图像  $I_1$  和图像  $I_2$  的变换模型是平移运动模型（见表 2-1）：

$$I_2(x, y) = I_1(x - x_0, y - y_0) \quad (3-10)$$

$$\text{将其进行傅立叶变换：} \quad F_2(\xi, \eta) = e^{-j2\pi(\xi x_0 + \eta y_0)} \cdot F_1(\xi, \eta) \quad (3-11)$$

$$\text{互功率谱定义为：} \quad \frac{F_1^*(\xi, \eta) F_2(\xi, \eta)}{|F_1^*(\xi, \eta) F_2(\xi, \eta)|} = e^{-j2\pi(\xi x_0 + \eta y_0)} \quad (3-12)$$

其中  $F_1^*$  为  $F_1$  的复共轭。将式(3-12)进行傅立叶反变换得到：

$$\delta(x - x_0, y - y_0) = F^{-1}[e^{-j2\pi(\xi x_0 + \eta y_0)}] \quad (3-13)$$

寻找上式中冲激函数的峰值位置即可确定图像  $I_1$  和  $I_2$  之间的平移运动参数。

上述结论是基于两幅图像具备简单平移关系的假设。冲激函数的峰值高低反应了两幅图像的相关性大小。在窄基线条件下，两幅图像具备更为复杂的透视变换关系，同时被噪声干扰，甚至含有运动物体，那么冲激函数的能量将从单一峰值分布到其他小峰值，但其最大峰值的位置具备一定的稳定性。

运动物体和噪声虽然改变了图像的局部像素分配，但却没有对图像整体产生大的影响，而相位相关法对局部像素变化是不敏感的。

下面的实验中，作者将原始图像分别进行理想平移、旋转并加入大量噪声和运动物体后，使用相位相关法来估算图像的平移参数。



(a) 原始图像



(b) 经理想平移后



(c) 旋转并加入大量噪声



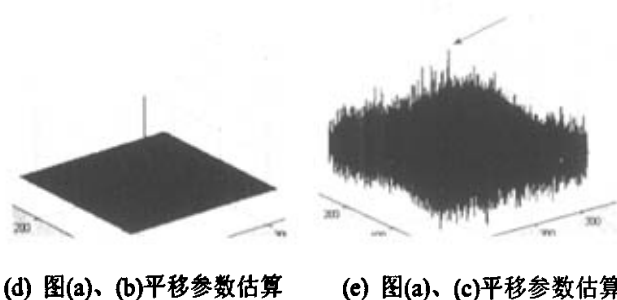


图 3-5 相位相关法估算平移参数

由图 3-5(d)可知,当两幅图像仅仅是平移关系时,能量全部集中在了冲激函数峰值位置处。而由图 3-5(e)可知,虽然图 3-5(c)被噪声和运动物体所严重干扰,也有一定的旋转角度,在这种情况下使用相位相关法估算平移量时,冲激函数的能量已经从单一的峰值分散开来,但其最大值的位置几乎没有变。作者还做过实验表明,当两幅图像之间重叠区域仅为 20%时,相位相关法也可以稳健的估计出图像间的平移关系。

相位相关法估算的平移参数不是很精确,但已经足以为角点匹配过程提供一个初始搜索范围。重要的是,相位相关法使用 FFT 实现,速度极快。

### 3.3.2 改进的角点匹配

虽然我们拍摄的实际图像之间是透视变换关系,但是在窄基线条件下仍然可以用相位相关法稳健的粗估算出图像间的平移参数,即图像的粗主运动方向。

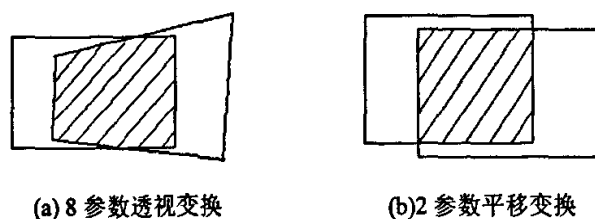


图 3-6 平移模型与透视变换模型

图 3-6(a)的 8 参数透视变换描述了两幅图像之间的真正变换关系。而图 3-6(b)是 2 参数平移变换。相比而言,平移模型的两个参数( $x_0, y_0$ )约等于透视变换模型的  $h_{13}$ 、 $h_{23}$ , 并且二者的重叠区域也相近。于是我们可以用图 3-6(b)的重叠区域近似当作两幅实际图像的重叠区域。这样我们仅仅在这个重叠区域内检测角点。图

3-6(b)的重叠区域范围可以简单的由相位相关法估算的平移参数和图像的宽、高来求得，实际计算中要留有约 10 个像素的富裕量，以抵消这两个模型的差别。

当确定图像的重叠区域后，仅仅在图像重叠区域内进行角点检测。

在改进的角点匹配过程中，对于图像  $I_1$  中的任意角点  $p_1$ ，根据相位相关法计算出的平移量为  $t$ ，则可以估算  $p_1$  对应点的位置  $p_2$  位于以  $(p_1+t)$  为中心的某个窗口内。这也是由于两个模型的差别造成的，实际计算中窗口大小一般选为  $25 \times 25$ （窗口大小参数的选择取决于图像的尺寸，本文所用的实验图像的分辨率平均约为 700 个像素）。角点匹配仅仅在这个窗口内进行搜索，可以大大的降低执行归一化相关的次数。这种思想相当于用已知的粗略的全局运动参数去指导更精确的局部特征匹配。由于此时的角点匹配搜索被限制在一个窗口内了，这不仅提高了速度，同时也使得成功匹配的命中率大大提高，也较好的避免了运动物体、重复性纹理的影响。比如说，在图 1 中拍摄到一辆车在图的右部，而图 2 中这辆车到了图的左部。如果使用遍历式角点匹配方法，则这辆车上的角点可能会成为匹配点。虽然对这辆车来说这些角点的匹配是正确的，但对于两幅图像的全局而言这无疑是错配。而本文提出的全局匹配指导局部匹配的思路，尽可能保证了主运动的匹配而忽略独立运动。在本文的图像配准系统中，主运动指的是大背景的运动，这是摄像机运动造成的；独立运动指的是前景物体的运动、噪声等。

### 3.3.3 变换估计与配准误差计算

经角点匹配后形成了图像 1 和图像 2 之间的一组映射点集。接下来的任务就是从这一组映射点集中估算透视变换模型的 8 个参数。

给定图像  $I_1$  和  $I_2$  及  $n$  组候选初始匹配点：  $(x_i, y_i) \Leftrightarrow (x'_i, y'_i)$ ，根据单应性矩阵公式(2-15)可知，每一对匹配点可以产生两个方程：

$$x'_i(h_{31}x_i + h_{32}y_i + h_{33}) = h_{11}x_i + h_{12}y_i + h_{13} \quad (3-14)$$

$$y'_i(h_{31}x_i + h_{32}y_i + h_{33}) = h_{21}x_i + h_{22}y_i + h_{23} \quad (3-15)$$

代入 4 对不共线的匹配点，得到 8 个方程：

$$\begin{pmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x'_1x_1 & -x'_1y_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y'_1x_1 & -y'_1y_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x'_2x_2 & -x'_2y_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -y'_2x_2 & -y'_2y_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x'_3x_3 & -x'_3y_3 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -y'_3x_3 & -y'_3y_3 \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x'_4x_4 & -x'_4y_4 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -y'_4x_4 & -y'_4y_4 \end{pmatrix} \begin{pmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{pmatrix} = \begin{pmatrix} x'_1 \\ y'_1 \\ x'_2 \\ y'_2 \\ x'_3 \\ y'_3 \\ x'_4 \\ y'_4 \end{pmatrix} \quad (3-16)$$

理想情况下，利用求解线性方程组的方法求解(3-16)即得到  $H$  的 8 个独立的未知参数。但在实际计算过程中由于噪声、角点定位误差、模型误差、错配等因素的存在，往往需要代入 10 对以上匹配点坐标，然后利用最小二乘法求解<sup>[40]</sup>。

平均几何配准误差是衡量配准算法精度的一个重要指标<sup>[31]</sup>。其公式如下：

$$E_n = \frac{1}{2n} \sum_{i \in n} (dist(x'_i, Hx_i) + dist(x_i, H^{-1}x'_i)) \quad (3-17)$$

$dist$  表示两点之间的距离。该式所计算的误差的单位是像素。如果配准误差过大，则后续的图像镶嵌等步骤将无法进行。配准误差计算示意图如图 3-7 所示。

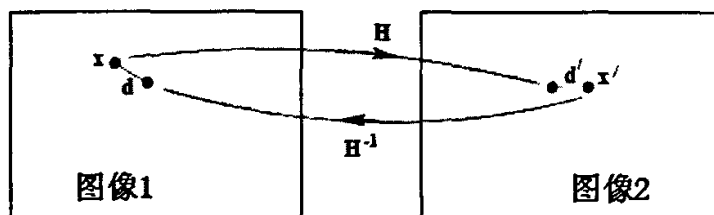


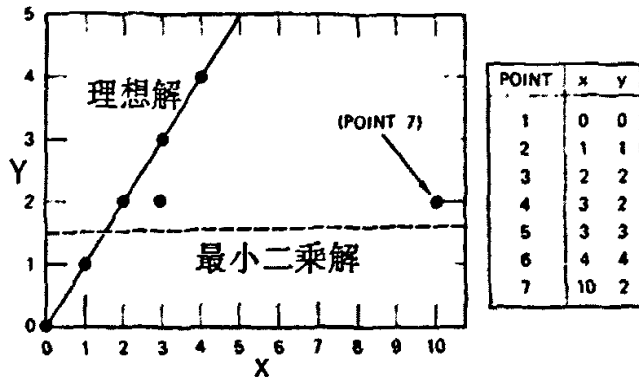
图 3-7 几何配准误差的计算

### 3.3.4 改进的鲁棒变换估计算法

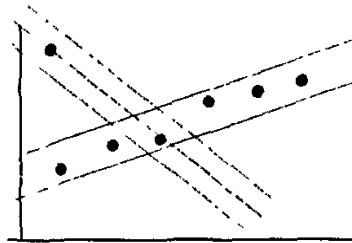
上一小节介绍了使用匹配角点估算变换模型参数的方法。在后面实验章节可以看到，虽然导向匹配可以去除大量错配，但不可能全部去除。如果直接用这些匹配点使用最小二乘法来估算模型参数，则会带来较大的配准误差（配准误差对图像镶嵌的影响请见第四章图 4-1 的对比）。

本文作者采用了改进的鲁棒变换估计算法—RANSAC<sup>[41]</sup>算法（RANDOM Sample Consensus，即随机抽样一致性算法）进一步去除错误的匹配。

下面先以简单的直线拟和的例子来介绍 RANSAC 算法的核心思想。



(a)



(b)

图 3-8 RANSAC 算法

图 3-8(a)中总共有 7 个点，其中有 6 个点属于内点，而第 7 个点属于外点。此时用最小二乘法拟和的直线如图中虚线所示，显然这是错误的，此时实际的最佳拟和直线应该是那条穿越了前 6 个点的黑色实线。

RANSAC 算法的思想简单而巧妙：首先随机地选择两个点，这两个点确定了一条直线。我们称这条直线的支撑为在这条直线的一定距离范围内的点。这样的随机选择重复数次，然后具有最大支撑集的直线被确认为是样本点集的拟和。在拟和的误差距离范围内的点被认为是内点，它们构成所谓的一致集，反之则为外点。根据算法描述，我们可以很快判断，如果只有少量外点，那么随机选取的包含外点的初始点集确定的直线不会获得很大的支撑，如图 3-8(b)所示。在图 3-8(a)的例子中，经过 RANSAC 算法后，点 7 就会被剔除掉。然后再使用剩余的 6 个点进行最小二乘拟和，就会得到接近理想解的直线。值得注意的是，过大的外点比例将导致 RANSAC 算法失败。

在直线拟和的例子中，由点集确定直线至少需要两个点；而对于透视变换，这样的最小集合要有 4 个点。本文采用了一种改进的 RANSAC 算法，步骤如下：

首先将当前最佳估计内点数目  $N_i$  设置为 0。

- 1.) 对图像按  $32 \times 32$  分割为一系列的子块, 然后在所有子块中随机抽取 4 个子块, 并在每个子块中随机抽取一个点, 就得到了 4 个初始点。这样做可以在一定程度上降低取到共线点的风险, 同时也避免了取得的 4 个点过于临近, 提高了随机抽样的效率。
- 2.) 根据这 4 个点用最小二乘法计算  $H$  的 8 个参数, 然后根据该  $H$  统计在给定距离阈值  $T$  内的匹配点数目。
- 3.) 比较当前内点数目与  $N_i$ , 如果大于  $N_i$  则将  $H$  作为当前最佳估计, 并更新  $N_i$ , 同时动态估测剩余所需迭代次数  $S$  ( $S$  正比于  $1/\log(p)$ , 其中  $p$  为  $N_i$  的比例)。如果当前迭代次数达到  $S$ , 则保留  $H$  并进行下一步; 否则循环执行 1~3。
- 4.) 根据此  $H$  和图像 1 中所有角点的位置信息, 估算这些角点在图像 2 中的位置区域, 并在这个区域内 (一般取  $3 \times 3$  的窗口, 而在第一次导向匹配时取  $25 \times 25$  的窗口) 重新进行角点匹配。这跟前面第一次用相位相关法估算的平移参数来导向匹配的思想是一致的, 但此时的  $H$  已经接近实际真值, 故可以更加精确、有效的指导角点匹配, 所以会有产生更多的有效匹配角点。
- 5.) 再次执行一遍 RANSAC, 然后退出。

本文作者提出的改进的 RANSAC 算法可以有效的提高随机抽样的效率, 同时导向匹配的思想被再次使用。这不仅增加了有效匹配角点的数目, 也更进一步的确保了最大程度上去除错配的角点。另外值得注意的是, 根据本文所使采用的迭代次数动态预测机制, 外点的比例将直接影响 RANSAC 算法迭代的次数, 外点比例越大, RANSAC 算法迭代次数也就越多。

为了更进一步降低配准误差提高配准精度, 本文使用 Levenberg-Marquardt 非线性优化算法进行优化, 其优化误差函数是式(3-17), 优化进行的起始值为第二次 RANSAC 算法得到的  $H_2$ 。因为此时 LM 优化是在较好的初始条件下进行的, 所以可以在较短的时间内完成, 并且不会陷入局部极小值问题。

当有效角点越多时, LM 优化的效果越好, 误差越低, 配准精度也就越高。

### 3.3.5 系统框架

总结一下这个算法的完整框架：

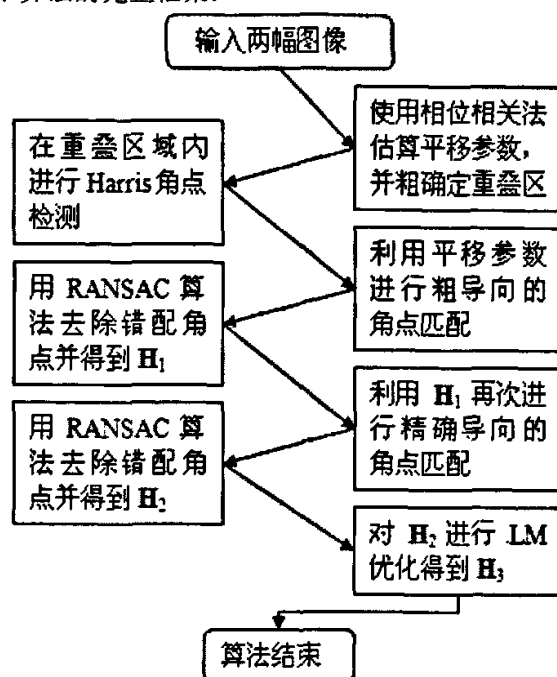


图 3-9 本文提出的图像配准算法流程图

### 3.4 实验结果与分析

虽然图 3-9 所示的本文算法步骤较多，但实际算法速度并不慢。这一节将用具体的实验验证本文算法的效果，并与文献<sup>[8]</sup>赵向阳提出的配准算法（简称原算法）进行对比。原算法使用了角点匹配和经典 RANSAC 算法。

本文第三、四章实验算法开发工具为 Matlab7.0 和 VC7.1，其中 Harris 角点检测算法使用 VC 编写并输出保存为文件，由于这一部分所耗时间相差不大，故不计入算法消耗时间；其他算法包括 NCC、相位相关法、RANSAC 算法等使用 Matlab 开发，利用了 Matlab 方便的矩阵运算库和强大的数据可视化能力。Matlab 属于解释执行语言，其速度较慢，但正可以便于我们分析算法的时间复杂度。

为了便于比较，作者对参数进行了统一设置：Harris 角点检测阈值为 2000（大约相当于  $600 \times 600$  的普通图像可以提取出 1000 个左右的角点）；归一化相关匹配窗口大小设定为  $7 \times 7$ ；RANSAC 算法距离阈值设定为 0.01，最大迭代次数设置

为 1000 次。另外作者规定配准误差大于 1.5 即认为配准算法失败。

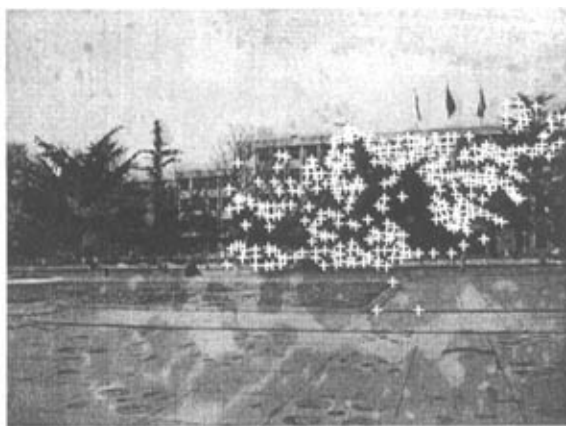


(a) 冲激函数的位置

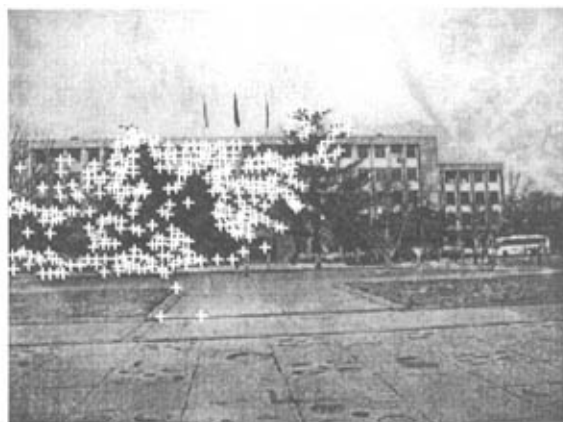


(b) 相位相关法导向角点匹配

(c) RANSAC 去错配



(d) 图像 1 中的匹配角点



(e) 图像 2 中的匹配角点

图 3-10 校园图片实验结果

第一组图片为校园图片。图 3-10(a)显示了相位相关法计算过程中冲激函数的

位置,即平移参数的大小,图 3-10(b)描述了使用相位相关法估算的平移参数进行导向角点匹配的结果,只有少量错配,跟图 3-2(c)相比,一次匹配有效率由 70% 上升到了 87%,继续使用 RANSAC 算法后匹配点有效率达到了 99%,如图 3-10(c)所示。图 3-10(d)和(e)为使用本文算法得到的最终匹配角点。

在本组实验中,原算法时间消耗总共为 13.8 秒, RANSAC 算法迭代 19 次完成,最终有效匹配角点为 338 对。而本文的算法中 RANSAC 仅迭代 5 次就完成,总耗时 2.8 秒,最终匹配角点为 479 对。原算法和本文算法估算的  $H$  分别为:

[ 1.1105 -0.0076 -277.0568 0.0496 1.1019 -20.3661 0.0002 0.0001 1.0000]  
[ 1.1159 -0.0101 -277.5927 0.0458 1.1117 -18.4349 0.0001 0.0000 1.0000]

经计算,在本组实验中本文算法的配准误差比原算法下降了 37.2%。本文的两次导向角点匹配可以产生更多的有效匹配角点,从而有效提高了配准精度。

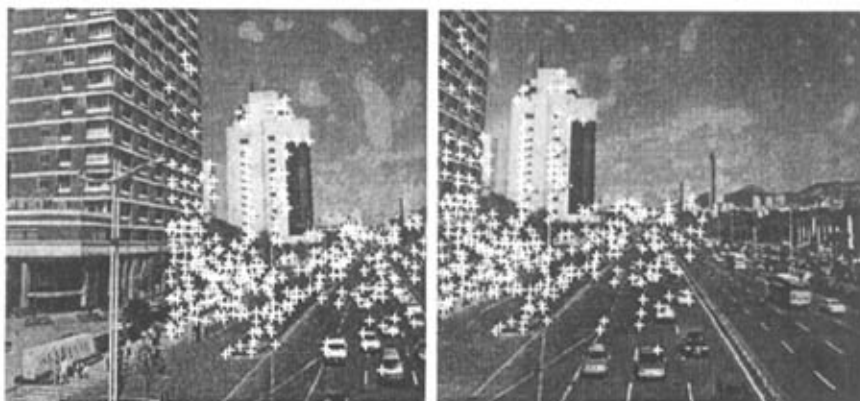
第二组实验图片为车流图片。



(a) 冲激函数的位置

(b) 导向角点匹配

(c) RANSAC 去错配



(d) 图像 1 中的匹配角点

(e) 图像 2 中的匹配角点

图 3-11 车流图片实验结果



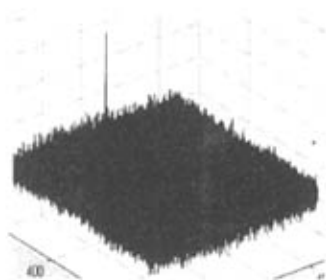
在这一组图片中，原算法 RANSAC 迭代 649 次完成，最终得到 31 对匹配角点，其中正确匹配 19 对，匹配角点有效率约为 61%，算法耗时 17 秒。

而在本文算法中，相位相关法依然稳健估算出平移参数，如图 3-11(a)所示。导向匹配后一次匹配有效率由 43% 提高到 78%，如图 3-11(b)所示，这大大的减轻了 RANSAC 算法的压力。RANSAC 算法在迭代了 24 次后结束，总时间消耗为 3.1 秒，最终得到的有效匹配角点数目为 302 对，最终角点匹配有效率约为 99%。

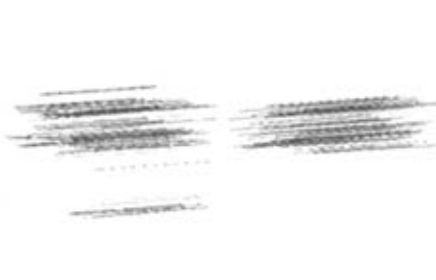
图 3-11(d)和(e)是使用本文算法最终得到的匹配角点。本文算法估算的  $H$  如下：

[1.0330 -0.0765 -167.2260 0.0812 0.9461 -33.7917 0.0002 -0.0002 1.0000]

第三组实验图片为植物园图片。



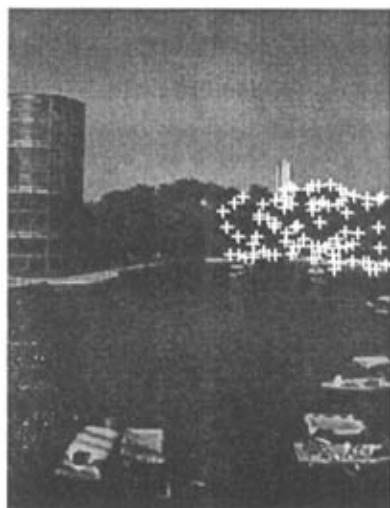
(a) 冲激函数的位置



(b) 导向角点匹配



(c) RANSAC 去错配



(d) 图像 1 中的匹配角点



(e) 图像 2 中的匹配角点

图 3-12 植物园图片实验结果

在本例中，因为存在水波纹、飘动的柳树、运动的船只等难以进行角点检测和匹配的物体，原算法 RANSAC 迭代 1000 次后失败，总共耗时间为 25 秒；本文算法采用了导向匹配后一次角点匹配有效率由 32% 提高到 86%，使得 RANSAC 仅迭代 11 次就结束，总耗时 2.6 秒，角点有效匹配率为 100%。可见本文算法成功的减少了荡漾的水波纹和运动船只上的错误匹配点。本文算法估算的  $H$  如下：

[0.9856 0.003895 -269.293 -0.0071 1.0101 8.0953 -0.0001 0.0001 1.0000]

在第四组含有众多行人的和树木的图片中（见附图四），使用本文算法也可以快速、正确的完成匹配，其匹配角点如图 3-13 所示。



(a) 图像 1 中的匹配角点



(b) 图像 2 中的匹配角点

图 3-13 行人图片角点匹配结果

除了以上四组实验图片外,本文作者总共采用了 25 组有代表性、较难匹配的图片以更加全面的验证算法。下表是本文算法与文献<sup>[8]</sup>算法的统计比较。

作者实验用机配置为 P4 1.5G CPU、256MB 内存、40G 硬盘。若使用效率更高的 C 语言重写算法,则速度还会有 3~5 倍的提升,可以达到更好的实用性。

	原算法	本文算法
平均匹配搜索次数	180.5 万	3.9 万
平均 RANSAC 迭代次数	190.2	19.7
平均有效匹配点数目	89.2	250.1
平均匹配点有效率	82.6%	96.8%
平均时间消耗	16.1 秒	2.9 秒
平均配准误差	0.58	0.37
算法成功率	84%	96%

表 3-1 本文算法与原算法的对比

可见,本文算法在速度上平均提高了 4 倍,匹配点有效率达 97%。在对所有的 25 组图片的实验中,本文算法平均可以产生 250 对有效匹配角点,并且配准误差更低。算法成功率为 96%,稳健性较好。

### 3.5 小结

在基于角点匹配的图像配准算法中,最为耗时的两个运算步骤是基于归一化互相关的遍历搜索过程和 RANSAC 迭代算法,这二者大约占据 90%的整体时间消耗。本文提出两次导向匹配的思想,使用速度极快的基于 FFT 的相位相关法估算的粗平移量来首次导向角点匹配,利用了全局运动参数来限制局部角点匹配,从而可以获得较好的初始有效匹配率,降低了 RANSAC 算法的压力;将改进的 RANSAC 算法估算的  $H$  再次用于导向角点匹配,从而可以获得更多的有效匹配角点,经 LM 优化后配准误差再次被降低。第二次 RANSAC 算法一般迭代 1~2 次就可完成,消耗时间可以忽略不计。

在本文作者选用的参数中,RANSAC 迭代算法距离阈值设定为 0.01,若设定为 0.0001 甚至更小,则可获得更精确、限制更为严格的内点集,但算法时间复杂

度会增加较多,从而失去了实用性。同时若增大归一化相关模板的窗口大小,相当于利用更多的角点邻域信息进行相似性比较,则角点匹配的稳定性也会加强,但会增加系统的时间开销;而若缩小窗口大小,则相当于牺牲了稳定性换取速度。本文算法折衷了速度和稳定性的考虑,最终确定选择了 $7\times 7$ 或 $9\times 9$ 的窗口尺寸。

另外在本章实验中,第一次导向匹配时使用相位相关法估算的粗平移参数,为了保证算法的稳定性而选择了较大的导向匹配搜索区域富裕量,为 $25\times 25$ 的窗口;第二次导向匹配时使用 RANSAC 算法估算的透视变换参数,考虑到此时已经是精确全局运动导向的局部搜索,所以设定导向匹配搜索区域的富裕量为 $3\times 3$ 的窗口既可。

本章最后的实验结果和数据表明,本文作者提出的这种由粗及精的导向匹配方式在速度、稳健性和配准精度上与现有算法相比都有不错的改进,拥有较好的实用价值。

## 第四章 图像变换与镶嵌技术

本章作者将介绍图像插值、图像变换、图像融合技术，并分析影响图像合成清晰度的几个因素，最后介绍作者提出的算法和实验结果。

### 4.1 理论基础介绍

#### 4.1.1 图像插值

在进行图像镶嵌的过程中，往往需要将一幅图像向另一幅图像变换（称为图像变形技术，即 Image Warping<sup>[42]</sup>），存在对图像进行缩放、旋转、透视变换等操作，这时就需要利用插值方法以获取非整数坐标点的像素值。下面作者介绍两种常用的插值方法：最近邻插值和双线性插值。

最近邻插值法(Nearest Neighbor Interpolation)<sup>[43]</sup>是最简单的一种插值方法。插值点的灰度值与其最临近点的灰度值相同。最近邻插值法的本质就是放大像素，计算速度很快。其缺点是会产生锯齿。

与最近邻插值法不同，双线性插值(Bilinear Interpolation)利用四个临近点来得到插值点的灰度。双线性函数定义如下：

$$z = f(x, y) = q_1 + q_2x + q_3y + q_4xy \quad (4-1)$$

双线性插值法假定插值点邻域内的点均满足双线性函数，则有：

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 & x_1y_1 \\ 1 & x_2 & y_1 & x_2y_1 \\ 1 & x_1 & y_2 & x_1y_2 \\ 1 & x_2 & y_2 & x_2y_2 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} \quad (4-2)$$

由式(4-2)解出系数：

$$q_1 = \frac{x_2y_2z_1 - x_2y_1z_2 - x_1y_2z_3 + x_1y_1z_4}{(x_2 - x_1)(y_2 - y_1)} \quad (4-3)$$

$$q_2 = \frac{-y_2z_1 + y_1z_2 + y_2z_3 - y_1z_4}{(x_2 - x_1)(y_2 - y_1)} \quad (4-4)$$

$$q_3 = \frac{-x_2z_1 + x_1z_2 + x_2z_3 - x_1z_4}{(x_2 - x_1)(y_2 - y_1)} \quad (4-5)$$

$$q_4 = \frac{z_1 - z_2 - z_3 + z_4}{(x_2 - x_1)(y_2 - y_1)} \quad (4-6)$$

对于数字图像，一般有  $x_2 - x_1 = y_2 - y_1 = 1$ 。将式(4-3)~(4-6)代入双线性函数，即可以计算出插值点的灰度值：

$$z = z_1 + (z_2 - z_1)(x - x_1) + (z_3 - z_1)(y - y_1) + (z_4 - z_3 - z_2 + z_1)(x - x_1)(y - y_1) \quad (4-7)$$

由式(4-7)就可以确定插值点的灰度。双线性插值法具有一定的抗锯齿能力。

此外，目前还存在很多其它的插值方法，如双三次方插值<sup>[44]</sup>、B 样条插值<sup>[45]</sup>。这些插值方法在获得更高的精度的同时，计算复杂性也随之提高。对于本文的图像镶嵌应用来说，双线性插值法在计算效率和精度方面可以达到一个很好的平衡，故本文在图像变换、重投影等过程中涉及到图像插值时选用了双线性插值。

#### 4.1.2 图像变形与流形

当图像序列帧与帧之间的变换模型参数估计出来之后，选定参考帧图像，将图像序列的其他帧重投影到统一的二维曲面上进行图像的镶嵌。

图像重投影用到的关键技术就是图像变形技术<sup>[42]</sup>。图像变形是实现图像之间插补变换的一类被广泛使用的技术，它通过图像像素位置的内插来产生新的图像。对于图像镶嵌来说，图像变形是指根据计算出的图像变换模型参数将源图像变换到目标图像所处的坐标空间上。

本文所用的基于 8 参数透视变换矩阵的图像变形过程包括三个步骤：确定源图像的四个顶点变换后的坐标；确定源图像变换后所占区域的边界；通过坐标变换及插值技术获取变换后的新图像。

下面以附图（一）校园图片的前两张图为例具体介绍一下图像变形。

根据第三章的图像配准算法，可以得到图像 1 和图像 2 之间的坐标变换公式，即透视变换矩阵  $H$ 。我们以图像 1 所在坐标系作为参考坐标空间，将图像 2 根据

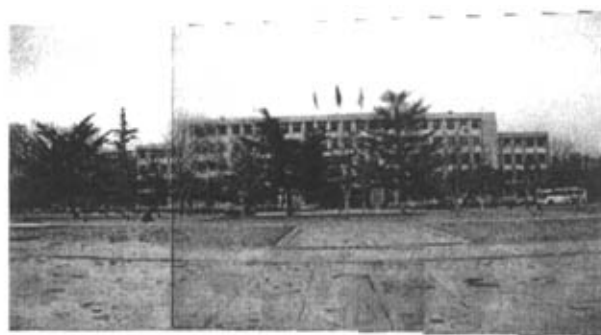
H 向参考坐标空间进行映射, 如图 4-1(a)所示。然后将图像 1 和经映射后的图像 2 进行简单的混合 (即像素灰度值取平均), 就得到图 4-1(b)。



(a) 图像 2 向参考坐标系变形后



(b) 配准误差为 0.25 时图像 1、2 混合的结果



(c) 配准误差为 0.6 时图像 1、2 混合的结果



(d)配准误差为 3.5 时图像 1、2 混合的结果

图 4-1 图像变形与混合

图 4-1(c)是当配准误差为 0.6 时,经变形后的图像 2 和图像 1 简单混合的结果,可见,混合后图像清晰度降低,大楼左上角存在重影,接缝处的地线也不连续。图 4-1(d)为当配准误差为 3.5 时的混合结果。此时变换矩阵  $H$  已几乎完全估算失误,图像 2 和图像 1 几乎没有成功融合起来。

在图 4-1 中,其中图 4-1(b)为使用本文配准算法估算的  $H$  进行简单混合的结果;图 4-1(c)为使用文献<sup>[8]</sup>的算法估算的  $H$  进行混合的结果;图 4-2(d)为不使用 RANSAC 算法直接由包含错配的点使用最小二乘法估算的  $H$  进行混合的结果。

下面来介绍一下图像重投影(reprojection)和流形(manifold)的概念。

我们将两两匹配的图像形成全景图时,首先要将这些关联的图像重投影到统一的表面,然后在这个表面上进行融合以形成新的图像镶嵌—全景图。我们称图像镶嵌形成的二维曲面为流形<sup>[46]</sup>,现在使用的有平面流形、柱面流形、球面流形和立方体流形等。其中,平面流形和柱面流形较为常用,而球面流形和立方体流形由于其构造比较复杂,因而较少采用。

本文作者在这里仅介绍平面流形和柱面流形。

#### (A) 平面流形

平面流形是最简单也是最常用的一种流形。如图 4-2 所示,当图像序列投影到平面流形进行镶嵌时,会形成一个蝴蝶结似的形状。使用平面流形进行图像重投影时,往往选定图像序列当中的一幅图像的左上角作为参考坐标系原点,而将



其他所有图像向参考坐标系进行变形，然后将它们镶嵌起来。

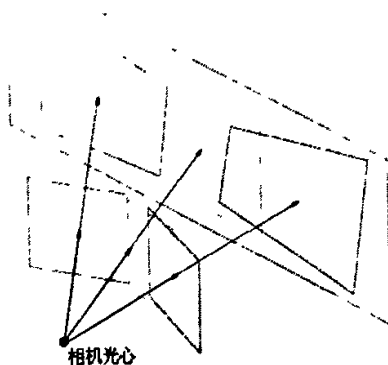


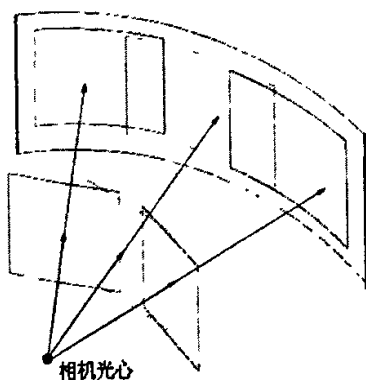
图 4-2 平面流形

对于任意给定的图像  $I_k$  与参考图像  $I_{ref}$  之间的变形公式由一系列透视变换矩阵连乘得到，比如选择  $I_1$  为  $I_{ref}$ ，根据  $I_2 = H_{21}I_1$  和  $I_3 = H_{32}I_2$  就可以计算出  $I_3$  到  $I_{ref}$  的变形公式： $I_3 = H_{32}H_{21}I_1 = H_{31}I_1 = H_{31}I_{ref}$ 。

前面的图 4-1(b)就是一个最简单的、只含有两幅图片的平面流形的例子。

平面流形的优点之一是可以适用于所有具备 8 参数透视变换关系的图像的镶嵌，优点之二是图像重投影后直线依然会保持为直线，不会产生弯曲，真实感强。但平面流形拼接后的图像视角跨度不能超过 90 度，否则图像镶嵌的两边会趋于无穷大从而无法满足视觉一致性。实际应用中，平面流形一般用来处理视角跨度不大的全景图拼接，或者用于需要快速切换视角的实时渲染系统<sup>[29]</sup>。

## (B) 柱面流形



(a) 柱面流形



(b) 经柱面投影后的图像

图 4-3 柱面流形与柱面投影

柱面全景图就是将图像投影到柱面流形后再镶嵌的结果。如图 4-3(a)所示,柱面流形可以支持 360 度视角的图像镶嵌,但经柱面投影后直线会被弯曲,从而造成一定的失真。另外使用柱面流形时要求尽量将摄像机水平扫动进行拍摄,且拍摄时不能变焦,这对满足 8 参数透视变换的两个近似条件又增加了限制。

柱面投影的坐标变换公式<sup>[47]</sup>如下:

$$\theta = \tan^{-1}(x/f), \quad v = y/\sqrt{x^2 + f^2} \quad (4-8)$$

其中  $(x, y)$  为平面坐标,  $f$  为焦距,  $(\theta, v)$  为柱面坐标。本文使用文献<sup>[47]</sup>的方法估测焦距:

$$f = \left( \frac{h_{23}^2 - h_{13}^2}{h_{11}^2 + h_{12}^2 - h_{21}^2 - h_{22}^2} \cdot \frac{h_{13}h_{23}}{-h_{11}h_{21} - h_{12}h_{22}} \right)^{\frac{1}{4}} \quad (4-9)$$

图 4-3(b)为将校园图片的第一幅投影到柱面的效果。可见图像中下方的直线发生了弯曲。

#### 4.1.3 图像融合算法介绍

当把图像序列重投影到流形上之后,接下来使用某种算法对重叠区域的图像进行融合。图像融合算法的好坏将直接关系到合成图像的清晰度。

目前存在很多重叠区域图像融合的方法,作者将介绍最常用的三种。

##### (1) 直接平均法

该方法将图像重叠区域对应像素点的灰度值进行叠加再求平均。这种融合方法使用简单,但是会在接缝处产生较强的不连续感。

##### (2) 线性加权过渡法

该方法与直接平均法相似,但重叠区域的像素值不是简单的叠加平均,而是加权叠加再平均。对于每帧图像来说,图像中心区域的像素有较高的贡献权值,而图像边缘区域权值较低,从而降低了接缝处的不连贯感。

##### (3) 多分辨率法<sup>[48]</sup>

该方法首先将图像分解成一系列具有不同分辨率的子带图像,然后在各层子空间上进行拼合,最后使用重构算法合成出原分辨率下重叠区域的图像。该方法一方面可以有效的保持细节信息不在融合时被平滑掉,另一

方面使得图像低频信息可以很好的实现平滑过渡，其镶嵌后的图像背景较为连贯。

#### 4.1.4 影响图像合成清晰度几个因素

能否生成高质量的合成图像是判断配准与镶嵌算法好坏的重要依据。由于多种因素的存在，图像重叠区域内分别处于相邻两幅图像的两个对应点并不一定能很好的重合在真实的场景点。作者将影响图像合成清晰度的主要因素概括如下：

- 1.) 配准误差的存在。当配准误差较大时，两幅图像映射后的点不能够完全重合，这一点可以简单的由图 4-1(b)、(c)、(d)之间的差别看出来。
- 2.) 相机模型误差。作者在第三章介绍了相机成像模型，并将针孔相机的 8 参数透视变换模型作为本文算法所使用的图像变换模型。实际上这是不够精确的，因为镜头存在径向失真等畸变，而且会受到成像噪声的干扰。尤其是当使用短焦拍摄的时候，镜头畸变的影响会更为明显。
- 3.) 视差(parallax)。视差存在的原因是相机在拍摄过程中发生了移动，或者所拍摄的景物较近，这破坏了透视变换矩阵成立的近似条件。
- 4.) 白平衡与曝光差异。当使用相机拍摄时，相机会根据景物和光线去测光，然后自动调整曝光、白平衡参数。如果待拼接的两幅图像存在较大的白平衡和曝光差异，其色彩和亮度会相差较大。此时无论使用何种方法融合，都会产生明显的色彩不连贯感。4.2 节介绍了作者提出的一种自动调整白平衡与亮度的图像拼接算法可以较好的解决这个问题。
- 5.) 运动鬼影(Motion Ghost)的存在。当我们要拼合的图像重叠区域含有运动物体时，即使此时图像大背景、主运动已被良好的配准，这些独立运动物体的存在将会使得最终融合的图像产生重影，甚至同一个物体在拼接后的图像中会出现两次。Echigo 提出了一种基于中值滤波的方法可以较好的解决这个问题<sup>[49]</sup>。

## 4.2 白平衡及亮度差异自动调整算法

传统的图像自动拼接算法要求拍照时相机参数严格一致，而当使用普通相机

拍摄时，由于曝光参数和白平衡都是相机根据景物和光线自动调整的，这样所拍摄的图像之间往往存在背景亮度和色彩的差异。如果不去除这些差异而直接进行拼接，合成的图像将没有连贯感。本文作者提出了一种简单实用的解决方法。

假设两幅图像已经使用第三章的方法几何配准并且将图像 2 向图像 1 的坐标空间进行了变形，使得图像 1 与图像 2 的对应点重合起来。接下来将图像的光度进行配准，以减少两幅图像色彩和亮度上的差异。

将红、绿、蓝三个色彩通道分开考虑，这样在每个颜色通道中，两幅图像重叠区域对应点的光度变换可以简单的用线性变换来近似<sup>[29]</sup>，如式 4-10 所示。

$$\begin{pmatrix} R_2 \\ G_2 \\ B_2 \end{pmatrix} = \begin{pmatrix} c_r & 0 & 0 \\ 0 & c_g & 0 \\ 0 & 0 & c_b \end{pmatrix} \begin{pmatrix} R_1 \\ G_1 \\ B_1 \end{pmatrix} + \begin{pmatrix} d_r \\ d_g \\ d_b \end{pmatrix} \quad (4-10)$$

其中  $(c, d)$  为线性变换参数，称为光度变换参数，是我们要估算的参数。计算时每个像素点就可以得到两组 RGB 数值：图像 2 中的  $(R_2, G_2, B_2)$  与图像 1 中的  $(R_1, G_1, B_1)$ 。采样的像素点的选取为匹配角点。

在理想情况下，直接使用最小二乘法就可以求出三组总共 6 个光度变换参数。但是在实际计算中，由于配准误差、运动物体和噪声等因素的存在，部分匹配点与式(4-10)的约束差异较大，成为模型估计中的外点。作者下面对植物园图片中的第 2、3 张进行分析，如图 4-4 所示。从附录（三）可见，第 2 张图片在拍摄的时候受太阳光影响而偏红，而且有些曝光过度，比其他图片稍亮。

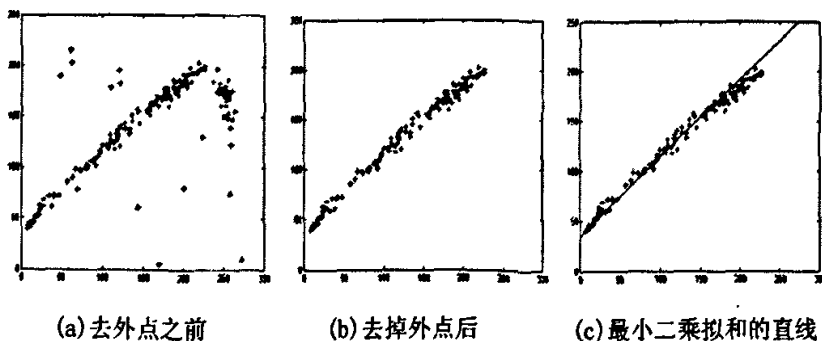


图 4-4 红色通道线性变换参数的估算

以红色通道为例，对该通道的对应点数据进行分析，如图 4-4(a) 所示，红色通道的数据存在部分外点。如果不去除这些外点而直接使用最小二乘法将会存在

较大误差甚至失败。为此这里使用 RANSAC 算法<sup>[41]</sup>去除掉这些外点,如图 4-4(b),然后使用最小二乘法拟和出每个色彩通道的色度变换参数,如图 4-4(c)所示。

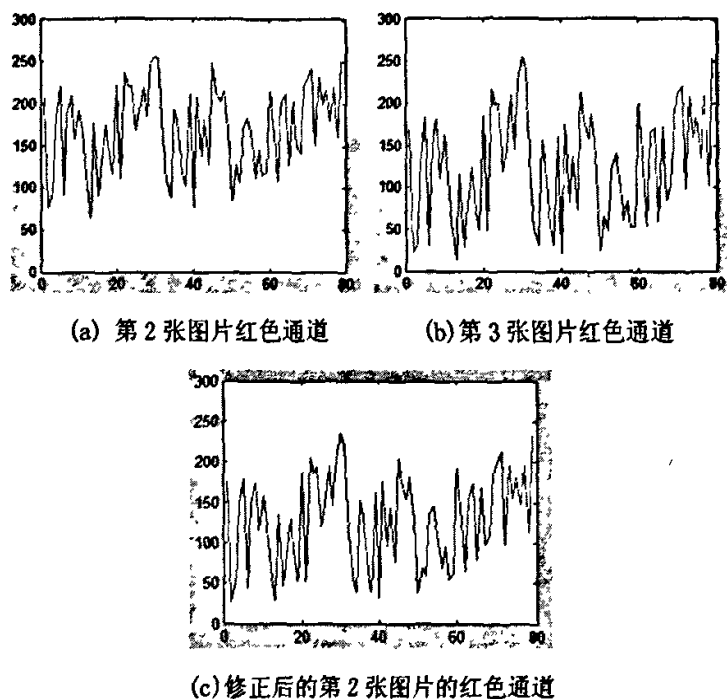


图 4-5 修正前后的红色通道

图 4-5 为对应点的红色通道采样图。其中图 4-5(a)为存在色偏和轻微过曝的第 2 张图片,图 4-5(b)为无色偏的第 3 张图片。使用本文算法对第 2 张图片进行校正后的结果如图 4-5(c)所示。可见对应点处的红色通道整体曲线已经较为接近无色偏的图 4-5(b),色彩和亮度差异得到了较好的自动调整。

将三个颜色通道分别进行调整后再进行图像融合,就可以有效的削弱白平衡与曝光差异的影响。

### 4.3 实验结果与小结

本节实验将对校园图片和植物园图片进行自动配准与镶嵌。

图 4-6 是附图一的 6 张校园图片拼接而成的柱面全景图,图 4-7 是附图三的 4 张植物园图片拼接而成的柱面全景图。两幅全景图都经过了裁减,去掉了不完整的边界区域。在拼接植物园图像时,使用了本文的白平衡与亮度自动调整方法,

可见从拼接图中几乎已看不出第二幅图色偏的影响。

(作者注:受论文打印条件限制,本文多数图像效果可能无法原貌展现。)



图 4-6 校园全景图



图 4-7 植物园全景图

从这两组实验图片的自动镶嵌结果可以看出,本文提出的图像配准算法配准精度较高,本文采用的图像镶嵌算法可以较好实现图像之间的连贯光滑过渡。

图 4-8(a)、(b)分别为使用线性加权融合算法和多分辨率融合算法镶嵌的图像在局部细节的对比。



(a) 线性加权融合算法

(b) 多分辨率融合算法

图 4-8 两种融合算法效果对比

## 第五章 SIFT 特征匹配技术

本文第二章最后简要介绍了基线的概念，第三章介绍了窄基线条件下的角点匹配技术。窄基线条件下是根据特征点邻域灰度信息的相似性进行匹配的；而在宽基线条件下，两幅图像成像条件相差较大，此时使用邻域相关信息进行匹配已不能满足要求。

本章研究了宽基线条件下的点特征匹配问题，并研究了先进的 SIFT 特征匹配算法及其应用。

### 5.1 宽基线特征匹配概述

宽基线条件下点特征匹配的首要任务就是提取稳定的特征，并进行描述。这里稳定一词的含义指的是希望该特征能对旋转、尺度缩放、仿射变换、视角变化、光照变化等图像变化因素保持一定的不变性，而对物体运动、遮挡、噪声等因素也保持较好的可匹配性，从而可以实现差异较大的两幅图像之间特征的匹配。

对图像变化保持稳定的特征描述符称为不变量，比如对图像的旋转保持稳定的不变量称为旋转不变量(Rotation Invariant)，对尺度缩放保持稳定的不变量则称为尺度不变量(Scale Invariant)。

特征描述符(Feature Descriptors)指的是检测图像的局部特征(比如边缘、角点、轮廓等)，然后根据匹配目标的需要进行特征的组合、变换，以形成易于匹配、稳定性好的特征向量，从而把图像匹配问题转化为特征的匹配问题，进而将特征的匹配问题转化为特征空间特征向量的聚类问题<sup>[51]</sup>。

宽基线条件下的点特征匹配一般包括下面四个步骤<sup>[30]</sup>：

- 1.) 特征点检测。这些特征点一般是灰度变化的局部极值点，含有显著的结构信息，甚至这些点也可以没有实际的直观视觉意义，但却在某种角度、某个尺度上含有丰富的易于匹配的信息。
- 2.) 特征点描述，即建立特征向量。这是各匹配算法主要的不同所在。特征空间的选择决定了图像的哪些特性参与匹配，哪些特性将被忽略。特征点的

特征描述符应是不变量，以确保最低限度的受摄像机的运动或光照变化等因素的影响。选择合理的特征空间可以降低各类图像变化因素对匹配算法速度、稳健性的影响。

- 3.) 进行特征匹配以获得候选匹配点。这一步根据特征向量的相似性来进行匹配，一般采用各种距离函数作为特征的相似性度量，如欧氏距离、街区距离、马氏距离等<sup>[51]</sup>。
- 4.) 消除错配。无论采用何种特征描述符和相似性判定度量，错配难以避免。这一步主要做的就是根据几何或光度的约束信息去除候选匹配点中的错配。常用的去外点方法是 RANSAC 随机抽样一致性算法<sup>[41]</sup>，常用的几何约束是极线约束关系(Epipolar Line)<sup>[24]</sup>。

## 5.2 SIFT 特征匹配

### 5.2.1 图像多尺度表示

尺度空间理论最早出现于计算机视觉领域时其目的是模拟图像数据的多尺度特征。Koendetink 在文献<sup>[52]</sup>中证明高斯卷积核是实现尺度变换的唯一变换核，而 Lindeberg<sup>[53]</sup>等人则进一步证明高斯核是唯一的线性核。

二维高斯函数定义如下：

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (5-1)$$

$\sigma$  代表了高斯正态分布的方差。

一幅二维图像，在不同尺度下的尺度空间表示可由图像与高斯核卷积得到：

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (5-2)$$

式(5-2)中， $(x, y)$ 代表图像的像素位置， $\sigma$ 称为尺度空间因子，其值越小则表征该图像被平滑的越少，相应的尺度也就越小。大尺度对应于图像的概貌特征，小尺度对应于图像的细节特征。 $L$ 代表了图像的尺度空间。



## 5.2.2 SIFT 特征匹配算法

David G.Lowe 在 2004 年总结了现有的基于不变量技术的特征检测方法,并正式提出了一种基于尺度空间的、对图像缩放、旋转甚至仿射变换保持不变性的图像局部特征描述算子—SIFT 算子<sup>[53]</sup>,其全称是 Scale Invariant Feature Transform,即尺度不变特征变换。

SIFT 算法首先在尺度空间进行特征检测,并确定关键点(Keypoints)的位置和关键点所处的尺度,然后使用关键点邻域梯度的主方向作为该点的方向特征,以实现算子对尺度和方向的无关性。

SIFT 算法提取的 SIFT 特征向量具有如下特性:

- a) SIFT 特征是图像的局部特征,其对旋转、尺度缩放、亮度变化保持不变性,对视角变化、仿射变换、噪声也保持一定程度的稳定性。
- b) 独特性(Distinctiveness)好,信息量丰富,适用于在海量特征数据库中进行快速、准确的匹配<sup>[23]</sup>。
- c) 多量性,即使少数的几个物体也可以产生大量 SIFT 特征向量。
- d) 高速性,经优化的 SIFT 匹配算法甚至可以达到实时的要求。
- e) 可扩展性,可以很方便的与其他形式的特征向量进行联合。

Lowe 在图像二维平面空间和 DoG(Difference -of-Gaussian)尺度空间中同时检测局部极值以作为特征点,以使特征具备良好的独特性和稳定性。DoG 算子定义为两个不同尺度的高斯核的差分,其具有计算简单的特点,是归一化 LoG (Laplacian-of-Gaussian)算子的近似。DoG 算子如式(5-3)所示:

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (5-3)$$

对于图像上的点,计算其在每一尺度下 DoG 算子的响应值,这些值连起来得到特征尺度轨迹曲线。特征尺度曲线的局部极值点即为该特征的尺度。尺度轨迹曲线上完全可能存在多个局部极值点,这时可认为该点有多个特征尺度。

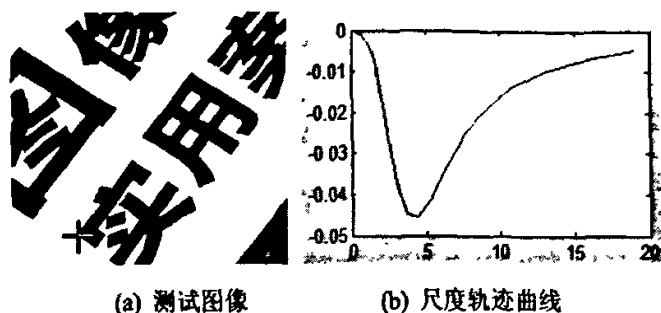


图 5-1 尺度轨迹

图 5-1(b)为图 5-1(a)中十字花点处的尺度轨迹曲线。可见该图中尺度轨迹曲线在大约  $\sigma=5$  处取得局部极小值。

SIFT 特征匹配算法包括两个阶段，第一阶段是 SIFT 特征的生成，即从多幅待匹配图像中提取出对尺度缩放、旋转、亮度变化无关的特征向量；第二阶段是 SIFT 特征向量的匹配。

下面本文来具体介绍一下 SIFT 特征匹配算法。

首先，一幅图像 SIFT 特征向量的生成算法总共包括 4 步：

(1) 尺度空间极值检测，以初步确定关键点位置和所在尺度。

图 5-2 为 DoG 尺度空间的三个相邻尺度。

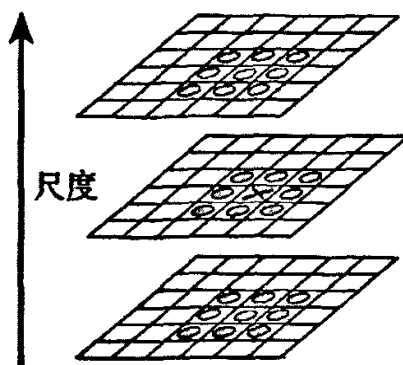


图 5-2 DoG 尺度空间局部极值检测

在检测尺度空间极值时，图中标记为叉号的像素需要跟包括同一尺度的周围邻域 8 个像素和相邻尺度对应位置的周围邻域  $9 \times 2$  个像素总共 26 个像素进行比较，以确保在尺度空间和二维图像空间都检测到局部极值。

(2) 通过拟和三维二次函数以精确确定关键点的位置和尺度，同时去除低对比度的关键点和不稳定的边缘响应点(因为 DoG 算子会产生较强的边缘响应)，以

增强匹配稳定性、提高抗噪声能力<sup>[53]</sup>。

(3) 利用关键点邻域像素的梯度方向分布特性为每个关键点指定方向参数，使算子具备旋转不变性。

$$\begin{aligned} m(x, y) &= \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \\ \theta(x, y) &= a \tan 2((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \end{aligned} \quad (5-4)$$

式(5-4)为 $(x, y)$ 处梯度的模值和方向公式。其中  $L$  所用的尺度为每个关键点各自所在的尺度。

在实际计算时，我们在以关键点为中心的邻域窗口内采样，并用直方图统计邻域像素的梯度方向。梯度直方图的范围是  $0 \sim 360$  度，其中每  $10$  度一个柱，总共  $36$  个柱。直方图的峰值则代表了该关键点处邻域梯度的主方向，即作为该关键点的方向。图 5-3 是采用  $7$  个柱时使用梯度直方图 of 关键点确定主方向的示例。

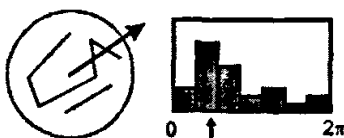


图 5-3 由梯度方向直方图确定主梯度方向

在梯度方向直方图中，当存在另一个相当于主峰值  $80\%$  能量的峰值时，则将这个方向认为是该关键点的辅方向。一个关键点可能会被指定具有多个方向（一个主方向，一个以上辅方向），这可以增强匹配的鲁棒性<sup>[53]</sup>。

至此，图像的关键点已检测完毕，每个关键点有三个信息：位置、所处尺度、方向。由此可以确定一个 SIFT 特征区域。

(4) 生成 SIFT 特征向量。

首先将坐标轴旋转为关键点的方向，以确保旋转不变性。

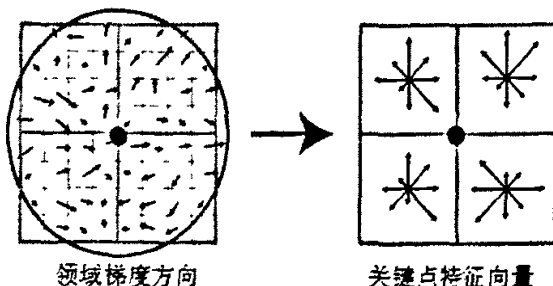


图 5-4 由关键点邻域梯度信息生成特征向量

接下来以关键点为中心取  $8 \times 8$  的窗口。图 5-4 左部分的中央黑点为当前关键点的位置，每个小格代表关键点邻域所在尺度空间的一个像素，箭头方向代表该像素的梯度方向，箭头长度代表梯度模值，图中蓝色的圈代表高斯加权的范围（越靠近关键点的像素梯度方向信息贡献越大）。然后在每  $4 \times 4$  的小块上计算 8 个方向的梯度方向直方图，绘制每个梯度方向的累加值，即可形成一个种子点，如图 5-4 右部分所示。此图中一个关键点由  $2 \times 2$  共 4 个种子点组成，每个种子点有 8 个方向向量信息。这种邻域方向性信息联合的思想增强了算法抗噪声的能力，同时对于含有定位误差的特征匹配也提供了较好的容错性。

实际计算过程中，为了增强匹配的稳健性，Lowe 建议对每个关键点使用  $4 \times 4$  共 16 个种子点来描述，这样对于一个关键点就可以产生 128 个数据，即最终形成 128 维的 SIFT 特征向量。此时 SIFT 特征向量已经去除了尺度变化、旋转等几何变形因素的影响，再继续将特征向量的长度归一化，则可以进一步去除光照变化的影响。

当两幅图像的 SIFT 特征向量生成后，下一步我们采用关键点特征向量的欧式距离来作为两幅图像中关键点的相似性判定度量。取图像 1 中的某个关键点，并找出其与图像 2 中欧式距离最近的前两个关键点，在这两个关键点中，如果最近的距离除以次近的距离少于某个比例阈值，则接受这一对匹配点。降低这个比例阈值，SIFT 匹配点数目会减少，但更加稳定。

### 5.2.3 实验结果

上一节具体介绍了 SIFT 特征向量的生成算法和利用 SIFT 特征向量进行特征匹配的方法。本节来介绍 SIFT 算法用于窄基线和宽基线图像点特征匹配的实验。实验算法采用 VC7.1 开发，程序界面如图 5-5 所示。

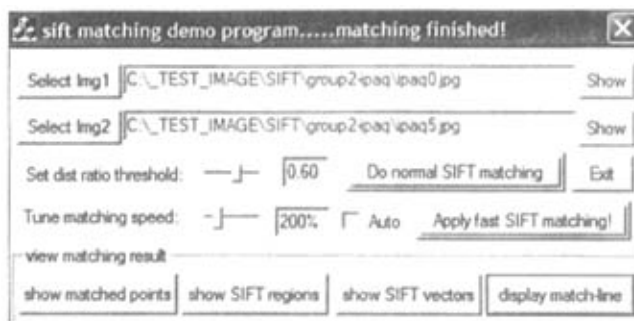
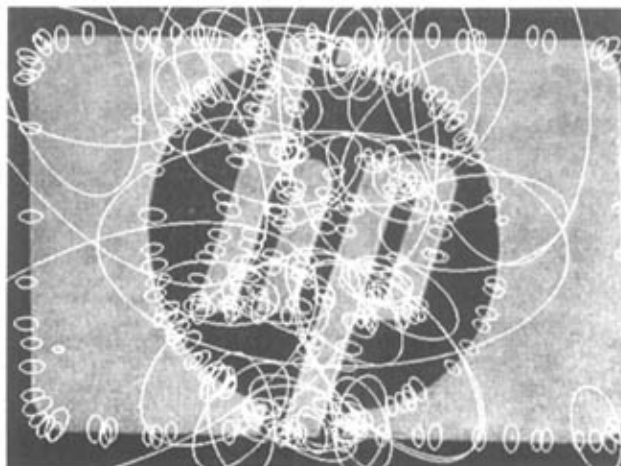


图 5-5 SIFT 特征匹配程序界面

第一组实验图片为包装盒图片(附图五)。该组图片在拍摄时进行了相机平移、转动和变焦操作。图 5-6(a)为第一幅图中检测出的 SIFT 特征。本文用不同颜色的椭圆来表示 SIFT 特征,椭圆的中心位置代表了关键点在图像中的二维坐标位置,椭圆的长轴代表了关键点的尺度,椭圆的方向代表了该关键点方向。

本例中取比例阈值为 0.4。图 5-6(b)是特征匹配的结果。可见,两幅图像中匹配的 SIFT 特征较为稳定,其位置信息、方向性和尺度信息准确,椭圆的长轴垂直于某个尺度下的灰度变化取得最大值的方向,与人眼的直观判断相符。SIFT 特征描述了图像的局部结构性信息,虽然不如角点和边缘特征那样有十分明显的视觉意义,但其因估算了特征的尺度和方向,从而对几何形变较大的宽基线图像具有较强的匹配能力。



(a)图 1 的 SIFT 特征



(b)匹配的 SIFT 特征

图 5-6 包装盒图片的 SIFT 特征检测与匹配



图 5-7 遮挡图书的 SIFT 特征匹配

第二组实验测试了算法对遮挡和视角变化的适应能力(附图六)。其中第一幅图为目标图书的正面拍摄图,第二幅图则含有4本图书且摆放顺序杂乱、互相有堆叠遮挡,另外第二幅图为侧面拍摄图,与第一幅图视角差距较大。图5-7为使用SIFT算法进行特征点匹配的结果,两幅图像中的对应SIFT特征点被成功匹配出,并且在较大视角差异下,SIFT特征依然保持了较好的方向性。

第三组实验图片(附图七),使用SIFT算法将不同方位、不同焦距拍摄的、具有一定倾斜角的、亮度不同的人体照片与复杂场景中的人体进行特征点匹配。实验结果如图5-8所示,可见本例中SIFT算法表现出较好的目标定位能力,且能够在一定程度上排除亮度差异、视角变化和复杂场景中其他物体的影响。



图 5-8 场景中人体的 SIFT 特征匹配

上面使用了三组宽基线图片作了实验,下面是以附图二的车流图片为例分析SIFT特征匹配算法对窄基线情况的图像进行点特征匹配的实验。

第三章的图3-11是使用本文提出的导向角点匹配算法进行特征点匹配的结果,图5-9、图5-10为使用SIFT算法进行匹配的结果。

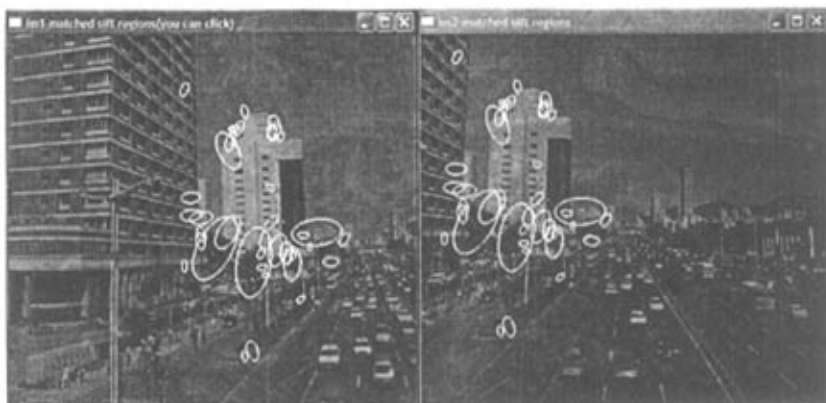


图 5-9 车辆图片的 SIFT 特征匹配

图 5-9 为 SIFT 特征区域的椭圆表示形式，图 5-10 为 SIFT 特征的箭头表示形式，箭头的起点代表该关键点的位置，箭头的长度代表该关键点所处的尺度，箭头的方向代表该尺度下关键点所处邻域的主梯度方向。

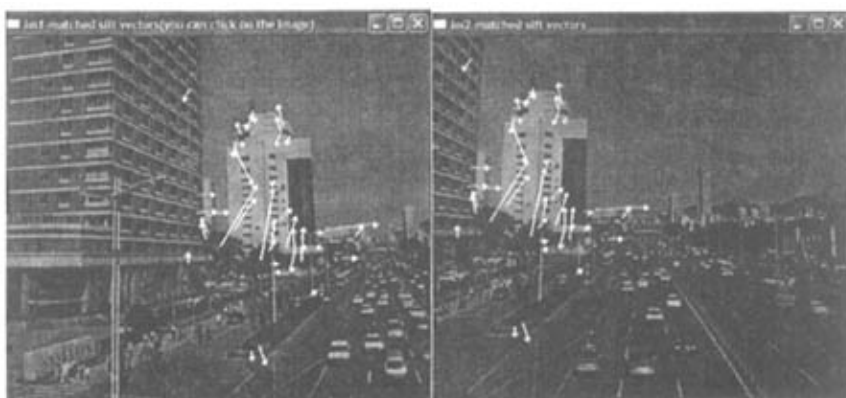


图 5-10 车辆图片的 SIFT 特征匹配

可见，局部匹配能力较强的 SIFT 特征匹配算法也可以较好的适用于窄基线图像特征匹配的情况，SIFT 算子可以较稳健的对发生几何形变、退化、受噪声干扰的图像局部特征进行准确的匹配。而且由于 SIFT 算法在计算关键点方向时充分利用了邻域信息，这样在一定程度上可以避免在小运动物体上匹配特征点，因为小运动物体的邻域信息即使去除了尺度和旋转的因素后也仅是具备较少的梯度方向相似性；同时 SIFT 算法在计算关键点处的梯度方向时使用了直方图统计和高



斯加权的思想,这就对存在定位偏差的特征点匹配提供了更好的适应性。

由于 SIFT 算法需要在各个尺度上进行计算,其时间复杂度相对较高,而且经 RANSAC 算法后得到的有效匹配点数目往往不是很理想。如果在窄基线条件下应用 SIFT 算子进行特征匹配,考虑到窄基线图像的自身特点,笔者建议仅选择相近的几个尺度进行计算,同时辅以图像的主运动信息和有效的几何限制,加速 SIFT 算法的匹配过程、提高匹配点数量,同时更进一步的提高匹配算法的稳健性,以更好的发挥 SIFT 算法的优势。

#### 5.2.4 小结

本文作者设计拍摄了 20 种宽基线情况下有代表性的实验图片,这些图片在拍摄时相机光心位置偏移较大,且均伴有 2~4 倍光学变焦操作,并选用不同的曝光参数(1/200 s ~1/300 s)或光圈大小(F2.8 ~ F5.6)以模拟亮度差异。选择合适的比例阈值(0.4~0.6 之间),在仅使用欧式距离进行 SIFT 特征向量匹配的前提下, SIFT 算法匹配的特征点有效率平均可以达到 80%左右;如果根据拍摄条件而选用透视变换矩阵、基础矩阵等<sup>[24]</sup>对其进行几何约束,采用 RANSAC 算法<sup>[41]</sup>、Hough 聚类法<sup>[54]</sup>进一步去除错配,则 SIFT 算法可以达到更高的正确匹配率。

另外本文也做了窄基线图像的点特征匹配的实验,分析了 SIFT 算法的优势,并给出了窄基线条件下应用 SIFT 算法进行特征匹配的建议。

在点特征匹配基础之上,可以进一步将 SIFT 算子应用于目标辨识、图像数据库检索等技术中<sup>[22,23,53,55]</sup>,其中 Andrew Zisserman 在 ICCV2003 会议上介绍了使用 SIFT 思想的一个基于物体识别的视频搜索系统—Video Google<sup>[55]</sup>,成功的演示了海量视频数据库检索中的快速图像检索,这些都在预示着基于不变量理论的特征匹配技术将拥有美好的发展前景和更加广阔的研究空间。

## 第六章 总结与展望

图像配准是一门实用性很强、理论与算法并重的技术，是很多计算机视觉算法研究的关键和前提。图像镶嵌则是根据图像配准的结果作进一步的图像融合，还可以使用融合信息进一步进行运动检测、视频压缩等应用。目前图像配准、镶嵌技术广泛应用于虚拟现实、视频压缩、图像复原、数据检索等技术中。

本文主要研究了窄基线条件下基于点特征的图像配准算法，同时也针对宽基线条件下的点特征匹配作了初步研究。主要工作成果总结如下：

- 1.) 系统介绍了摄像机几何成像模型的重要知识，阐述了摄像机内外参数的含义，论证了单应性矩阵成立的两个近似条件，并概述了估算单应性矩阵参数的几类方法，介绍了单应性矩阵的线性求解及非线性优化方法。
- 2.) 系统研究并实现了较常用的边缘、角点特征检测算法，包括 Canny 算法、SUSAN 算法和 Harris 算法；研究了窄基线条件下基于邻域灰度互相关的匹配。
- 3.) 系统的介绍并研究了三类图像配准方法，包括基于特征的方法、直接像素亮度差优化的方法和基于变换域信息的方法。介绍了各类方法的主流研究成果。并分析了彼此优势。
- 4.) 分析了窄基线条件下图像配准的难点所在，提出并实现了一种两次导向匹配思想的基于角点特征的图像配准算法，首先使用相位相关法估算平移运动参数并指导第一次角点匹配（粗），然后使用 RANSAC 算法稳健估计透视变换矩阵并指导第二次角点匹配（精），改进了现有角点匹配算法的速度、稳健性和精度，并用大量实验进行了分析对比。
- 5.) 介绍了插值技术、图像变形技术，介绍了平面流形和柱面流形，概述了图像融合的方法，并系统分析了影响图像合成清晰度的几大因素。
- 6.) 提出并实现了一种基于线性模型的色偏自动校正算法。首先根据本文的算法对图像进行几何配准，再利用匹配点在三个彩色通道的信息估算出各个色彩通道的线性变换参数，并将该算法用于全景图自动拼接系统中。
- 7.) 介绍了宽基线和窄基线的在特征匹配上的重要区别和不变量技术的概念，

研究了 LoG、DoG 算子及图像的多尺度分析技术，并研究了基于尺度空间的 SIFT 特征匹配算法。同时用丰富有趣的实验结果介绍了 SIFT 算法在遮挡物体匹配和复杂场景人体特征定位中的应用。

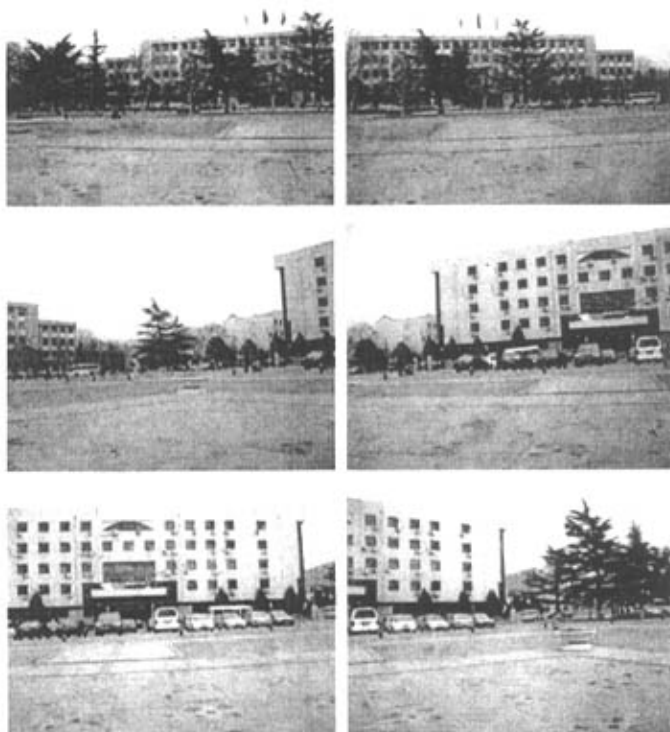
论文的不足之处和值得进一步研究的问题：

1. 论文提出了一种两次导向匹配的图像配准算法，如果在最开始的时候使用相位相关法失败，则整个算法就失败了。虽然笔者采用了 25 组实验图片中只有一例失败，但却没有深入分析失败的原因，也没有进一步的研究处理该问题的对策。
2. 在给定匹配点的情况下，利用线性解法和非线性解法求解变换矩阵有时会进入死锁，或者求解错误，尤其当匹配点数目较少时这将变得极不稳定。这需要更进一步的研究变换矩阵各参数和求解过程的互动关系。
3. 文章虽然分析了影响图像合成清晰度的几个因素，但却未编写实际算法。尤其拍摄实际图像时，运动物体往往无法避免，这就必须编写实际的去运动鬼影算法。这是一个很实用的算法。
4. 本文仅研究了静态镶嵌，对当今的研究热点—动态镶嵌和基于动态镶嵌技术的视频压缩还未作研究。
5. 本文针对 SIFT 特征匹配算法仅做了初步研究，而没有深入研究其在目标识别、图像检索中的应用。这些都是极富有吸引力的研究课题。

图像配准是个极富挑战性的研究领域，随着社会经济发展的不断进步，智能监控检测系统、公安追踪系统、数字娱乐系统等等，一切的进步都离不开图像配准这一核心问题上所取得的一个又一个突破。图像配准技术必将有更广阔的发展空间和美好的前途。

## 附录 本文实验图片

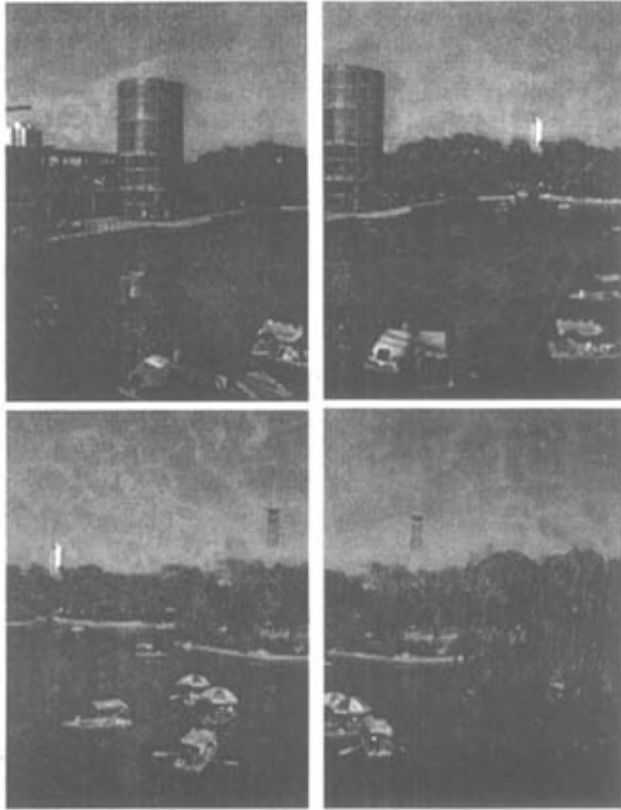
### 附图（一） 校园建筑图片



### 附图（二） 车流图片



附图（三） 植物园图片



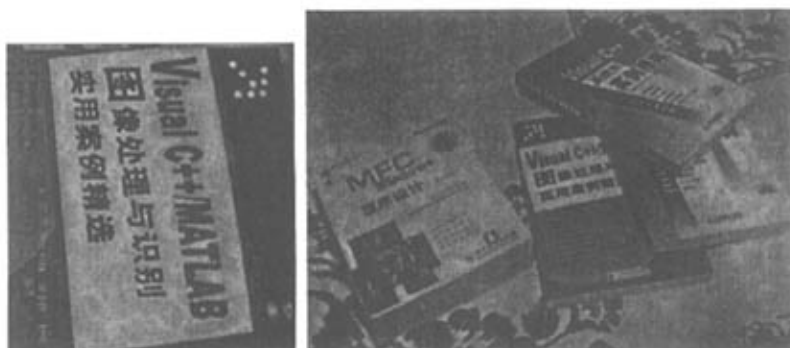
附图（四） 行人图片



附图（五） 惠普包装盒



附图（六） 遮挡的图书



附图（七） 场景中的人体匹配



## 参考文献

- [1] Barbara Zitova, Jan Flusser. Image registration methods: a survey[J]. Image and Vision Computing 21(2003), p977-1000
- [2] Cordelia Schmid, Roger Mohr and Christian Bauckhage. Comparing and Evaluating Interest Points[C]. In Proc. of the 6th International Conference on Computer Vision, pages 230-235, Bombay, 1998.
- [3] Lisa Gottesfeld BROWN. A survey of image registration techniques[J]. ACM Computing Surveys. Vol.24, No.4, December 1992: p325-376
- [4] Martin Jagersand. Recent Methods for Image-Based Modeling and Rendering [R]. IEEE Virtual Reality 2003 tutorial.
- [5] Richard Szeliski and Heung-Yeung Shum. Creating Full View Panoramic Image Mosaics and Environment Maps[C]. In Proc. of SIGGRAPH 97, pages 251-258
- [6] Shenchang Eric Chen. QuickTime VR - An Image-Based Approach to Virtual Environment[C]. In Proc. SIGGRAPH 95, pages 29-38.
- [7] Leonard McMillan. Plenoptic Modeling: An Image-Based Rendering System [C]. In Proceedings of SIGGRAPH 95, pages 39-46
- [8] 赵向阳,杜立民. 一种全自动稳健的图像自动拼接融合算法[J]. 中国图象图形学报, 2004, 9(4): p417-422
- [9] C.Harris, M.Stephens. A combined corner and edge detector[C]. The 4th Vision Conference, 147-151, 1988
- [10] M. Brown and DG Lowe. Recognising Panoramas[C]. Proceedings of IEEE International Conference on Computer Vision 2003. page(s): 1218- 1225 vol.2
- [11] M. Brown and D. Lowe. Invariant Features from Interest Point Groups[C]. In Proc of the 13th British Machine Vision Conference, p 253-262, Cardiff, 2002.
- [12] K. Mikolajczyk, C. Schmid. Scale & Affine invariant interest point detectors[C]. In IJCV 1(60):63-86, 2004.
- [13] 蒋晓悦,赵荣椿. 几种块匹配运动估计算法的比较[J]. 计算机应用研究 2004

Vol.21 No.7 P.1-3

- [14] Dae-Hyun Kim, Yong-In Yoon, Jong-Soo Choi. An efficient method to build panoramic image mosaics[J]. Pattern Recognition Letters 24 (2003) 2421–2429
- [15] B. Reddy and B. Chatterji. A FFT-Based Technique for Translation, Rotation, and Scale Invariant Image Registration[J]. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 5(8), pp. 1266-1271, 1996.
- [16] 李忠新,茅耀斌,王执铨. 一种基于频域相关技术的柱面全景图生成方法[J]. 计算机工程与应用. 2004 Vol.40 No.11 pp81-82,145
- [17] 解凯等. 图像 Mosaics 技术综述[J]. 电子学报 2004 Vol.32 No.4.p630-p334
- [18] Michal Irani, Steve Hsu. Video compression using mosaic representations[J]. Signal Processing:Image Communications(7), No. 4, Nov 1995, pp. 529-552
- [19] Michal Irani , P. Anandan , Steve. Hsu . Mosaic based representations of video sequences and their applications[C]. Procs of the Fifth International Conference on Computer Vision, p.605, June 20-23, 1995. 8
- [20] M. Irani and S. Peleg. Improving Resolution by Image Registration[C]. CVGIP, vol. 53, pp. 231--239, May 1991
- [21] 陈小蕾等.多帧超分辨率重建中的图像插值[J].信号处理 vol20(2004) p268-271
- [22] C. Schmid, R. Mohr. Local Grayvalue Invariants for Image Retrieval[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 19(5):530–535, May 1997.
- [23] D. Lowe. Object Recognition from Local Scale-Invariant Features[C]. In Procs of the International Conference on Computer Vision, pages 1150–1157,Corfu, Greece, September 1999.
- [24] 马颂德, 张正友. 计算机视觉: 计算理论与算法基础[M]. 科学出版社 2004
- [25] Richard Hartley and Andrew Zisserman. Mutliple View Geometry[R]. In CVRP June 1999
- [26] 于泓,陈辉,赵辉. 基于交叉垂直线的摄像机标定新算法[J]. 计算机应用 Vol.26 No.1 Jan.2006, p163-164,p168



- [27] Dublin, Ireland. Tutorial on 3D Modeling from Images[R]. ECCV 2000, 26 June
- [28] Faugeras, O. Three-Dimensional Computer Vision: A Geometric Viewpoint. MIT Press[M], Cambridge, MA. 1993
- [29] David Peter Capel. Image Mosaicing and Super-resolution[D]. PhD Thesis, Robotics Research Group Department, University of Oxford, 2001
- [30] 胡志萍. 图像特征提取、匹配和新视点图像生成技术研究[D]. 博士学位论文. 大连理工大学 2005 年
- [31] Richard Hartley. Geometric Optimization Problems in Computer Vision[R]. Systems Eng, Australian National University
- [32] Krystian Mikolajczyk. Detection of local features invariant to affine transformations: Application to matching and recognition[D]. PhD Thesis. Imagieric Vision Robotique Institut National Polytechnique de Grenoble.
- [33] Steve Seitz. Computer Vision : Edge Detection[R]. University of Washington 2002
- [34] CANNY J. A Computational approach to edge detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8(6):679-698
- [35] Haralick, R.M. Edge and Region Analysis for Digital Image Data[C]. Computer Graphic and Image Processing (T8 , Vol. 12, No. 1, pp60 January. 1980)
- [36] David A. Forsyth, Jean Ponce. Computer Vision: A Modern Approach[M]. 2002
- [37] SM Smith. SUSAN- A New Approach to Low Level Image Processing[R]. TR95SMS1c
- [38] Dmitry Chetverikov. IPAN99 method[P]. Image and Pattern Analysis Group 1999
- [39] Reddy, B.S., Chatterji, B.N. A FFT-Based Technique for Translation, Rotation, and Scale Invariant Image Registration[J]. IEEE Trans on Image Processing, 5:8, August 1996
- [40] Gilbert Strang. Linear Algebra and its Applications[M]. 1997
- [41] Fischler M A: Random Sample Consensus: a paradigm for model fitting with

- p application to image analysis and automated cartography[J].Communication Association Machine,1981,24(6):381-395
- [42] George Wolberg. Digital Image Warping[M]. IEEE Computer Society Press 1990.
  - [43] Lertrattanapanich S .Image registration for video mosaic[D]. Master thesis, The Pennsylvania State University,1 999
  - [44] Hou H S, Andrews H C. Cubic splines for image interpolation and digital filtering[J].IEEE Trans ASSP.1978,26 (6):508-517
  - [45] Unser M,Aldroubi A,Eden M .Fast B-spline transforms for continuous image representation and interpolation[J].IEEE Trans on PAMI,1991,13(3):277-285
  - [46] Shmuel Peleg, Joshua Herman. Panoramic Mosaics by Manifold Projection[C]. In CVPR97 Page: 338
  - [47] Hui CHEN. Building Panoramas from Photographs Taken with a Hand-held Camera[D]. PhD Thesis. Hongkong University 2002
  - [48] P. J. Burt and E. H. Adelson. A multiresolution spline with application to image mosaics[J]. ACM Transactions on Graphics, 2(4):217-236, 1983.
  - [49] T. Echigo, R. Radke . Ghost Error Elimination and Superimposition of Moving Objects in Video Mosaicing[C], ICIP 1999
  - [50] Richard O.Duda Peter E.Hart David G.Stork. Pattern Classification[M]. 2004.2
  - [51] Koenderink. The structure of images[J]. Biological Cybernetics,1984,p363-396.
  - [52] Lindeberg .Scale-Space for discrete Signals[J]. IEEE Trans.PAMI,1980,20:7-18
  - [53] David G.Lowe. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision,2004
  - [54] Hough,P.V.C. Method and means for recognizing complex patterns[P]. U.S. Patent No.3069654
  - [55] Josef Sivic and Andrew Zisserman.Video Google: A Text Retrieval Approach to Object Matching in Videos. ICCV2003

## 致 谢

本论文是在导师陈辉教授的细心指导下完成的。研究生三年，陈老师为我提供了优良的学习、科研和实践的平台，无论是搞项目开发、做计算机视觉课题研究，还是培养克服难题的自信、提高独立解决问题的能力，陈老师孜孜不倦的指导使我受益匪浅。在此学生特向陈老师表达心中深深的谢意。

感谢王欣老师的教诲，王老师淡薄名利、专注事业，乃学生一生的楷模；感谢王汇源老师对我的耐心指导；还要感谢周晓雁老师、付永生老师无微不至的帮助；尤其要感谢赵辉、宋刚和王洪君老师多年来对学生不变的厚爱和支持。

感谢我的同学：张权、于泓、郭恺、侯健、曲怀敬、赵振、王增锋、冷严、赵彩华、田宏阳、马文静、张晶等，是你们陪伴我度过了三年充实难忘的时光。

感谢上海杰图软件公司技术总监王帅对我的技术指导和帮助；感谢山东鲁光公司董事长黄华光为我提供项目实践的机会；感谢四维工作室负责人汪晓平、人民邮电出版社责任总编汤倩在我最困难的时候给予我莫大的信任和支持；感谢 Inria 的 David 博士提供给我一流的跨平台 C++ 图形类库；感谢 CSDN、Image2003、China-Vision、Matwav 的朋友和我一起交流并解决了许多技术难题。

感谢我的爸爸和妈妈，他们在精神和物质上给予我的支持和帮助，在生活上给予我无处不在的呵护，是我最坚强的后盾。另外感谢女友给我的支持和帮助。

最后，感谢所有帮助和关心过我的人！

**I would like to thank**

**my parents**

**my teachers**

**my mates and friends**

**and everyone who helped me !**

## 攻读学位期间发表的论文

- 1 赵辉, 陈辉, 于泓, “一种改进的全景图自动拼接算法”, 《中国图象图形学报》, 已录用, 将于近期刊出
- 2 赵辉, 陈辉, 于泓, “基于亮度与白平衡自动调整的图像拼接算法”, 《计算机应用》, 已录用
- 3 赵辉, 陈辉, 于泓, “一种快速稳健的图像配准方法及其应用”, 《中国电子学会第十一届青年学术年会》会议论文集
- 4 于泓, 陈辉, 赵辉, “基于交叉垂直线的摄像机标定新算法”, 《计算机应用》, 2006 Vol.26(1) p.163-168
- 5 于泓, 陈辉, 赵辉, “基于形状模板匹配的图像拼接算法”, 《计算机工程与应用》, 已录用