

武汉科技大学

硕士学位论文

基于SIFT特征的全景图像拼接算法研究

姓名：郑辉

申请学位级别：硕士

专业：模式识别与智能系统

指导教师：吴谨

20100427

摘 要

图像拼接是计算机视觉与图像处理领域的一个研究热点，是指将同一场景，相互之间存在部分重叠的图像序列进行空间匹配对准，再经图像融合形成一幅包含各图像序列信息的、较高分辨率或宽视角的新图像。图像拼接技术解决了视野与分辨率之间的矛盾问题，在遥感图像处理、医学图像分析、数字视频压缩、虚拟现实技术等领域都有广泛的应用。

本文首先概括了图像拼接技术的研究意义、现状以及应用领域，并系统地介绍了图像拼接的基本流程及其关键技术，对图像配准和图像融合的主要方法、性能以及存在的主要问题进行了分析和总结。在前人研究的基础上，本文提出一种改进的基于 SIFT 特征的全景图像拼接算法。首先提取每幅图像的 SIFT 特征并进行提纯匹配，针对多幅图像的配准，引入概率接收模型确认图像间的相互匹配关系，去除干扰图像，再采用柱面投影模型连接起所有的图像，最后进行加权平均融合处理,得到无缝的全景拼接图。

本文的创新之处在于针对传统 SIFT 特征匹配算法匹配时间较长和存在错误匹配的不足提出了改进。在 SIFT 特征提取方面，构建高斯金字塔时减少图像采样形成的组数，选择 3 个采样组；在 SIFT 特征匹配方面，在 Kd-Tree 搜索算法基础上，采用 BBF 搜索算法提高特征点的搜索效率；在 SIFT 特征提纯方面，采用改进的 RANSAC 算法过滤外点，去除冗余和误匹配对。实验证明本文算法在继承了传统 SIFT 算法较强鲁棒性的同时，进一步提升了拼接的速度和精度,而且降低了对图像序列获取的要求，增强了算法的自适应性。

关键词：图像拼接，图像配准，尺度不变特征变换，图像融合，随机采样一致性算法

Abstract

Image stitching is one of the most important research areas among computer vision and image processing. It's a task used to match and register a group of images having spatial overlap into a reference surface, and then use blending method to form a seamless panoramic image, which contains the image sequence information and resolves the problem of high resolution and wide viewing angles. Because the image mosaics technology has solved the contradictory problems between the field of vision and the resolution, it has many applications, such as remote sensing image processing, medical image processing, digital video compression, motion analysis, virtual reality and so on.

First, the present research status, application domain and the broad application prospect of mosaics are introduced in this paper. Later, the basis steps and key techniques of image mosaic are discussed. The common image matching and image blending methods are introduced, and their performance as well as the existing problems is analyzed. Based on the former research, this paper describes an automatic panorama generation method using SIFT. Firstly, it extracts and purifies SIFT feature from each image. In addition to finishing the Image stitching of several disorder images even including noise images, it supplies a probabilistic model to verify the image matches, and meanwhile process the noise images. Later, this paper links all the images adopting cylinder projection model. Lastly, the paper uses weighted average method to realize image blending in order to have a more satisfied panorama image.

In this paper, an improved SIFT feature-based approach is proposed whose goal is to get longer match time and overcome the setback of wrong match points in the traditional SIFT feature-based algorithm. In the part of SIFT feature extraction, it chooses three sample groups while constructing Gaussian pyramid. In the part of SIFT feature matching, in order to improve search efficiency it uses BBF search algorithm based on Kd-Tree search algorithm. In the part of SIFT feature purification, this paper uses modified RANSAC algorithm to eliminate redundancies and wrong match parts. Experimental results indicate that this algorithm is as robust as SIFT, meanwhile, the efficiency and precision are improved. Furthermore, this method reduces image collection condition and enhances adaptability in image mosaic.

Keywords: Image stitching, Image matching, SIFT, Image blending, RANSAC algorithm

第一章 绪论

1.1 引言

图像拼接技术^[1] (Image Mosaic), 也称为图像镶嵌技术, 是数字图像处理与计算机视觉领域的一个研究重点。图像拼接是指在相同的场景下, 将一组存在两两重叠区域的图像经图像配准和融合处理后生成一幅新图像, 处理后的图像包含了原图像的所有信息, 并且具有较高的分辨率及较大视野。

图像拼接技术是要解决如何将小视域的图片拼接成一张大场景图片的问题, 以满足用户全景摄影的要求。图像拼接技术最早出现在遥感技术领域^[2], 由于距离、角度、光线亮度和摄像设备等各种因素的限制, 需要对大范围领域进行监控或研究, 仅用一张照片不可能完成任务, 需要通过将两幅或多幅的具有相互重叠区域的组合成一幅完整的场景遥感图。

人们在日常生活中, 通常采用普通照相机采集图像信息, 希望采集到大视域全景图, 必须调整焦距, 缩放镜头才能得到, 可是采用此办法采集到的大视域场景图像分辨率较低。某一相机的分辨率是不变的, 采集图像的场景面积与分辨率成反比, 所以人们只能牺牲全景视野来换取高分辨的场景照片或者相反。此外, 对于某些特殊领域, 比如航天图像的拍摄中, 由于距离和角度等条件的限制, 即使在传感器的分辨率增加的情况下, 一些超大规模的物体或场景不可能用一张照片采集到。

为了更好的解决图像分辨率与图像场景大小之间的矛盾, 相关研究人员提出了图像拼接技术, 将普通图像或视频图像进行无缝拼接得到超宽视角甚至 360° 全景图, 实际上是在不增加硬件成本的条件下, 利用软件处理技术实现普通相机对超宽视角景物的拍摄, 可以在保持图像高分辨率的同时, 扩大所得图像的视野范围。

1.2 图像拼接技术国内外研究现状

至目前为止, 国内外学者们提出了各种各样的图像拼接方法, 各种方法都是面向一定的应用领域, 具有其自身的特点。

在全景图像拼接领域, R .Szeliski 和J .Coughlan^[3]于1996年提出了一种2D空间上八参数投影变换模型来进行图像的配准, 用搜索参数空间的办法来获得变换参数, 该方法效果较好, 可以处理具有平移、旋转、仿射等多种变换的待拼接图像, 但是拼接速度较慢。2000年, S.Peleg等人在R.Szeliski理论的基础上提出了自适应图像拼接模型^[4], 根据相机的不同运动, 自适应选择拼接模型, 通过把图像分成狭条后进行多重投影来完成图像的拼接, 自此, 自适应问题成为图像拼接领域的新的研究热点。Sevket Gumustekin^[5-6]对生成360度全景图像时产生的投影变形进行消除, 通过标定摄像机来建立成像模型, 再利用成像模型将

待拼接图像投影到统一的高斯球面上,来进行图像拼接。2003年, M.Brown^[7]使用基于不变量技术的SIFT算法进行图像拼接,算法完全自动完成且效果较好, M.Brown的研究成果再次掀起全景图像拼接技术研究的热潮。

在国内,图像拼接技术近年来也取得较大的发展。在基于特征点的图像拼接方面,西安电子科技大学^[8-9]在基于特征点及其方向的图像拼接技术取得了较大进展,中科院自动化所模式识别国家重点实验室利用基于特征点实现了集成电路图像的拼接。在全景图像拼接领域,国防科技大学的孙立峰、张茂军等人构建了基于柱面模型全景图的虚拟现实系统^[10-11]。

图像拼接技术具有广阔的应用前景,其方法的研究越来越受到了国内外的广泛关注,图像拼接领域的最新研究动向主要有:实时图像拼接、图像自动拼接、彩色图像拼接、3维立体图像拼接。

1.3 图像拼接的应用领域

图像拼接技术关系到图像处理(Image Processing)、计算机视觉(Computer Vision)、计算机图形学(Computer Graphics)和虚拟现实(Virtual Reality)等领域的重要课题研究,在遥感图像处理、近海矿藏勘探、虚拟现实技术、数字视频、军事侦察、公安取证、生物医学、机器人视觉和3-D图像重建等领域有着广泛的应用^[12-15],主要表现在以下几个方面:

1.遥感图像处理

一般来说,遥感图像的拍摄对象具有尺寸大范围广的特点,并且由于受到拍摄高度的限制,要得到拍摄目标概貌的全景图通常要进行各个角度、多个场景的多次拍摄,再将所拍摄的局部区域的遥感图像通过图像拼接处理技术得到一幅完整的场景遥感图像,进一步拓宽场景视野,提高图像分辨率。

遥感技术在国土资源勘探、土地、植被、海洋生物资源调查、军事侦察、环境监测、规划、管理和考古等领域具有广泛的应用。航空遥感以其独特的机动灵活和快速响应特性,在应对各种紧急情况和实时动态监测灾害时,发挥着极其重要作用。通过将多幅遥感图像拼接成大视野图像可实现对某一地区某一情况的全面监测,如对某地区的农作物病虫害情况进行监控,可以使农业部门更全面的了解整体情况,并为采取措施提供实时依据。对地震灾害情况的监控,可以使救援部门迅速发现情况,及时采取应对策略,使灾后重建部门对灾区整体进行全面了解和规划。

2.医学图像分析

随着可视化技术的不断发展,现代医学已越来越离不开医学图像的信息。医学图像在临床诊断、教学医学影像、科研等方面发挥着重要作用,从CT图像、X光照片到人体的细胞图片,大幅度的全景图像能帮助医生对病灶及其周围部位的情况进行全面、直观地视觉观察。另外,在利用计算机网络实现医学资源共享的远程医疗会诊中,处理异地远端传输的序列病理图片时,图像拼接技术也是不可缺少的技术。利用图像拼接技术,不但可以使

当地医院和远程终端用户能够随时随地看到医疗会诊的完整“现场”，还方便医院对这些资料进行保存存档，便于以后的对比与研究。

3. 虚拟现实技术

虚拟现实技术是利用计算机构造一个趋于真实的虚拟空间，人们通过某些一定的专用工具，能像对虚拟空间中的实物进行观察与控制，跟在自然现实环境中没有很大差别。图像拼接技术是虚拟现实场景绘制方法中的一项最基本的技术，利用图像拼接可以生成全景图像，通过合适的空间模型把多幅全景图像组织成虚拟全方位空间，从而可以使用全景图代替场景建模和绘制。目前，互联网的普及极大地促进了虚拟现实技术的推广与应用，这在很大程度上也促进了全景图像拼接技术的发展。

4. 数字视频

视频监控系统中，图像拼接技术将多路摄像头获取的视频监控图像拼接成一幅无差别、连续的大视场视频，兼顾了场景的完整性及画面的连续性，便于现场监控，无需分别监控每一路视频。

图像拼接不仅可以满足对广场、道路、剧场等场景的监控需求，还可以通过将视频图像序列构造成全景视图，满足人们的不同需求。例如在大型体育赛事过程中，通常都是若干个摄像机从不同地点、不同角度对赛场的情况进行跟踪拍摄，采用数字图像的拼接技术，将某时刻不同视频流的的视频图像进行拼接，观众可以在任何时间任何方位观看的赛场所有情况，2008北京奥运会期间许多比赛场景、图像，都是相关人员运用数字图像拼接技术得到的，这样不但方便了观众观看赛事，还增加了比赛的观赏性。

1.4 本文主要内容及各章节的安排

本文主要研究了基于特征点的图像拼接的关键技术，对经典的基于尺度空间不变特征(Scale-Invarious Feature Transform, 简称SIFT)图像配准算法进行了深入分析，针对该算法中存在的特征匹配时间较长和错误匹配的不足，提出了一种改进和优化方法，实现了图像的精确配准。完成配准后，采用渐入渐出加权平均法对拼接结果进行图像融合，最后实现了多幅景物图像的全景图自动拼接。

全文共分六章，各章内容安排分别如下：

第一章阐述了图像拼接的意义、应用领域和国内外研究现状，分析了当前技术发展存在的主要问题并提出了本文研究的主要内容；

第二章简要介绍了图像拼接基本理论、拼接的基本流程和各流程中的关键技术，重点研究了图像配准技术，为图像的平面拼接奠定了理论基础；

第三章首先介绍了研究了基于SIFT特征的经典算法，对于该算法中存在的特征匹配时间较长和错误匹配的不足，在SIFT特征点提取方面，综合考虑提取速度和精度，在构建高斯金字塔时选择3个图像采样组，在SIFT特征点匹配方面，在K-d树算法基础上，采用近似最近邻算法(Best-Bin-First, 简称BBF)搜索策略提高关键点的搜索效率，在SIFT特征点提纯部分，采样改进的随机采样一致性算法(Random Sample Consensus, 简称RANSAC)过滤外

点，去除误匹配对。实验证明本文算法可以较好的解决经典SIFT算法存在的上述问题；

第四章介绍了图像融合一些经典方法，分析其优缺点及实际应用场合，并利用渐入渐出法对本文所采集的景物图像进行了融合处理，实现了重叠区域的平滑过渡；

第五章在两幅图像拼接的基础上，实现了若干幅图像的基于SIFT特征的全景图像拼接，并用实验证明了该算法的可行性；

第六章对全文工作进行了总结，介绍了当前图像拼接领域研究与应用中存在的一些难点，并对本文实验中出现的具体问题分析其原因、探讨其解决办法，最后对图像拼接技术和进一步的研究工作提出展望。

第二章 图像拼接基本理论

图像拼接主要由图像采集、图像预处理、几何校正、图像配准和图像融合五步组成，虽然不同的算法在每步的处理上方法多种多样，但是总体的步骤还是基本一致的，基本流程如图2.1所示。

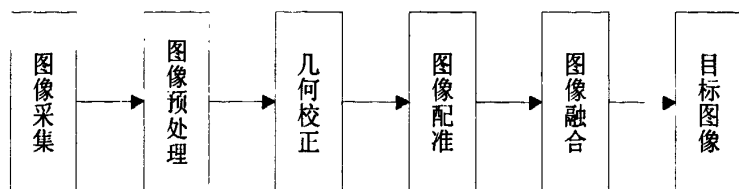


图2.1 图像拼接基本流程

首先通过图像采集得到了图像拼接的原始素材，再对采集到的图片进行一定的预处理和几何校正，以减少干扰图像拼接的条件；接着对经过处理的两幅图片分别提取各自的特征信息，并将特征信息进行分类、筛选、提炼；然后使用特征信息计算两幅图片之间的平移、旋转等拼接参数，进行特征信息匹配；将匹配得到的图像进行融合处理，以消除拼接缝隙，得到最终的拼接结果。

2.1 图像采集

图像采集通常是图像处理系统的首要步骤，又被称作图像信息获取，就是通过某种方式得到原始图像的过程。图像的采集方式多种多样，针对不同的应用场合选择不同的采集方式，如航空宇宙采用望远镜，医学上需要应用显微镜来采集微生物图片，生活中则使用双目摄像机，普通数码相机或手机等等。

使用摄像机捕捉到实验所需的图像，不同的获取方式得到相互区别的输入图像，使得需采用的拼接算法以及最终的拼接结果也相应变化。根据照相机拍摄时的运动状态，图像获取方式一般有以下三种情况^[16]：

1、旋转式照相机拍摄

获取图像过程中，不改变相机三脚架的落脚点，摄相机绕某一轴旋转，每旋转一定的角度，拍摄一张照片，旋转的角度不能过大，必须保证相邻图片存在重叠区域。

2、手持式照相机拍摄

这种方式又可分为以下两种：手持照相机在固定地点旋转拍摄、按照与采集目标相平行的路径拍摄。在忽略微小位移和相机抖动的情况下前者可视为旋转式拍摄，后者可视为下文所介绍的平移式拍摄。

3、平移式照相机拍摄

这种方式下，摄相机在拍摄过程中，一直保持与采集目标平面相平行。当焦距特定时，相机一般是放在滑轨上移动拍摄场景图片。最终的图像拼接质量取决于拍摄目标的大小变

化以及其到相机的距离不同。所以，平移式拍摄必须要确保相机平行移动的方向跟图像平行。

比较以上三种拍摄方式，可以知道：旋转照相机采集的优点是拍摄位置不变，所以需要的相机视图参数较少，易于实现，但是因为得到的图像在互异的平面上，还须将其投影到相同平面上才能进行后续的拼接过程，在生成柱面或球面全景图中使用较多。手持式拍摄不需要专业的设备且操作简便，故在图像信息获取中经常使用，本文就是用此方式来获取所需图像。但是实验证明，拼接手持相机获取的图片存在一些困难，因为在拍摄过程中，相机不是按既定的方式运动，不易估计。要使拼接效果最佳，必须减小相机旋转、平移的幅度，降低相邻图像间的不连续性，或者增加相邻图像间的重合区域，减少缝合困难。对平移照相机方式拍摄而言，由于获得的图像在相同平面上，采集到的图片在旋转、仿射等变化微弱，拼接处理过程较为简单，而且计算量也相对较小；但是获得的全景图的3维效果比旋转式拍摄的要差，此外，这种方法需要昂贵的采集设备，故实用性不强。

为了使采集到的图像满足要求，上述方式都规定了摄像机在一定范围内运动。但是不管采用何种方法，在图像采集过程中都会出现不确定、不稳定因素，例如由于光照变化、运动目标的位移、焦距的变化和旋转等因素带来的影响，得到的多幅图像之间会有光的强度的改变，并出现视觉差异，给图像镶嵌带来困难，因此，在图像镶嵌技术上，应该选取适应性较强的算法来弥补在拍摄上所带来的误差^[17]。

2.2 图像预处理

图像预处理的目的是确保图像匹配的准确度。在全景图拼接中，由于匹配的范围很大，因此带来的计算量也大，匹配的效率和精度较低。通过对待匹配图像进行预处理，包括对采集到的图像做相应的折叠、坐标变换，对图像进行初步定位，大致确定匹配位置，从而减小匹配范围、降低运算量、提高匹配速度。

由于采集到的图像的特点、噪声及应用场合各不相同，所以我们得根据实际来选择图像预处理方法。如对因拍摄、折叠变换和随机噪声引起相邻图像的重叠部分细节存在差异，需要先对图像进行平滑与边缘锐化处理。对图像只存在平移的情况，可以采用相位相关算法处理。对于水平方向上位移不大，垂直方向上位移能略去不计的情况，应该应用灰度图投影算法对相邻图像进行初步定位。对于因光照条件发生变化引起图像特征定位困难，可以对图片先进行中值滤波。对于待拼接的彩色图像在获取过程中，由于光照强度、相机感光元件的噪声、拍摄时相机视角的变化等因素的影响，以及图像可能来源于不同传感器，采集到的各个图像颜色不一致，必须利用特定的图像颜色校正算法对各个子图像颜色信息进行校正，从而使该全景图像的各个子图的颜色趋于一致。

2.3 图像几何校正

几何校正目的是消除和改正图像因几何畸变带来的几何误差^[18]，基本思路是通过图

像做相应的几何运算处理, 改变图像中目标的形状和空间位置。图像匹配是以图像的特征信息为基准, 但是特征信息会随待匹配图像类型、成像条件变化而变化, 因此会存在几何畸变误差, 影响匹配参数的确定, 导致拼接结果不理想。所以在人为因素和自然因素的影响下, 采集到的图像都需要进行几何校正, 甚至还需要照相机的标定, 以保证能够得到适于拼接的高质量图像。

2.3.1 摄像机运动模型

获取图像信息时, 图像的成像状态取决于摄像机的运动状态, 成像的结果取决于摄像机的运动方式。摄像机的运动有以下五种方式^[19]: 平移运动、垂直旋转、水平旋转、旋转旋转、变焦运动, 其具体运动方式以及与成像结果之间的关系如下:

- 1、平移运动是指摄像机移动方向始终都与目标平面xy平行, 引起图像的平移变换;
- 2、垂直旋转是指摄像机绕水平轴(x轴)旋转, 映射到图像变换上为投影变换;
- 3、水平旋转与垂直旋转类似, 指摄像机绕垂直轴(y轴)旋转, 映射到图像变换上也为投影变换;
- 4、饶轴旋转是指摄像机绕主光轴(z轴)旋转, 引起图像的旋转变换;
- 5、变焦运动是指摄像机的焦距发生了变化, 即改变了场景平面与成像平面间的距离, 从而导致图像的放缩变换。

图像采集时, 除了将摄像机固定某一地点或对其设置一定路径拍摄外, 也常手持相机移动拍摄, 对于手持式拍摄, 相机运动较为复杂, 包含了平移、变焦和旋转等运动方式。在实际应用过程中, 摄像机的运动一般是以上几种运动方式的合成。

2.3.2 图像变换模型

图像拼接的研究是基于这样的假设: 相机固定在空间某一点, 然后旋转一定角度后(要求相邻两幅图片之间应有部分重叠区域)拍摄, 每次拍摄的角度和焦距不应有较大变化。可以将此过程理解为相机在一个由景物构成的圆柱体内部沿特定半径小规模旋转而成像的过程^[20], 该过程的实现即为几何变换: 将待匹配图像变换到参考图像的坐标系中。利用齐次坐标, 建立上述模型的关系表达式为:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2.1)$$

在式(2.1)中, 图像变换矩阵 H 有 8 个参数, 各个参数分别代表摄像机不同的运动方式, 具体对应情况见表 2.1。

表 2.1 H 矩阵各参数的含义

参 数	参数对应的含义
m_0, m_1, m_3, m_4	尺度和旋转变量
m_2	水平方向位移
m_5	垂直方向位移
m_6, m_7	水平和垂直方向的变形量

通过对矩阵 H 做相应的简化，可以到下面不同的变换模型^[16]：

1、刚性变换模型

刚性变换模型具有平移不变性，它不会改变物体原有的形状，即原图像和变换图像中的像素点间的距离相同，并且不改变对应直线的平行性和垂直性。这类模型可看做是平移和旋转的集合，刚性变换矩阵有 3 个自由度（未知量），它的矩阵形式 H 如式所示：

$$H = \begin{bmatrix} \cos\theta & -\sin\theta & m_2 \\ \sin\theta & \cos\theta & m_5 \\ 0 & 0 & 1 \end{bmatrix}$$

(2.2)

那么，将式(2.2)代入(2.1)中，得到刚性变换过程表示为：

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} m_2 \\ m_5 \end{bmatrix}$$

(2.3)

其中， θ 为图像的旋转角度， m_2 和 m_5 为平移量。

2、仿射变换模型

仿射变换模型可以描述摄像机的平移、旋转、缩放运动。它的一个重要性质是“直线性”，即图像中的一条直线经过仿射变换后仍然为直线，且保持其平行关系，改变了的是不同方向的长度比。仿射变换矩阵具有6个自由度，它的矩阵形式H如式(2.4)所示：

仿射变换模型具有“保直性”，经过此类变换后图像的直线性质不变，且直线间的平行关系也不变，改变了的是不同方向的长度比。这类模型可看做是平移、旋转和变焦运动的集合，仿射变换矩阵具有6个自由度，它的矩阵形式H如式(2.4)所示：

$$H = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ 0 & 0 & 1 \end{bmatrix}$$

(2.4)

那么，将式(2.4)代入(2.1)中，得到仿射变换过程表示为：

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} m_0 & m_1 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} m_2 \\ m_5 \end{bmatrix} \quad (2.5)$$

3、投影变换模型

相对于上述两种变换方式，投影变换具代表性，它基本包含了摄像机的所有运动形式。刚性变换模型和仿射变换模型可以看做是投影变换模型的简化，因此研究投影变换模型具有广泛的代表性。投影变换模型的特点是直线经过投影变换后仍然是直线，但其平行关系发生改变。其变化矩阵具有8个自由度，它的矩阵形式H如式(2.6)所示：

$$H = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & 1 \end{bmatrix} \quad (2.6)$$

那么，将式(2.6)代入(2.1)中，得到投影变换过程表示为：

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2.7)$$

2.4 图像配准

图像拼接的质量主要取决于图像的配准精度。图像配准，简而言之，是指对同一场景在不同的时间，从不同的视角，相同或不同传感器拍摄的两幅有重叠区域的图像进行处理的过程^[7]，它从几何上校准参考图像和待配准的图像。它的核心问题是寻找一个空间变换，使图像之间相互重叠部分的坐标点对准，并缝合成一个新的大画面的视图。配准算法既要保证配准的精度，又要使计算量不至于过大，这是目前图像拼接技术的关键步骤，也是本文研究的重点。

2.4.1 图像配准关键要素

由于图像配准比较复杂，对于有着复杂空间变换的图像配准目前还没有很完美的解决方法。国内外许多学者对此进行了大量的研究工作，针对不同的应用领域，提出了不同的图像配准方法，虽然算法各异，但都可以归纳为特征空间、相似性度量、搜索空间及搜索算法的不同组合^[21]。

1、特征空间

特征空间是由参与配准的图像特征构成，特征包括图像的灰度值、边缘、纹理、角点、统计特征、显著特征等^[22-24]。选取怎样的特征空间取决于参与匹配的图像特征，选取的特征信息应该是明显并易于提取的，并且在参考图像和待匹配图像上有足够多的均匀分布。选择合理的特征空间可以改善配准算法的性能，减少配准算法的计算量，减小搜索空间、噪声和其它不确定因素对匹配算法的影响，同时也会影响配准算法对不同成像方式获得的

图像的适应能力。

2、搜索空间

搜索空间是图像匹配中待估算参数的集合。图像的变换可以分为以下三类：全局变换、局部变换和运动场形式的变换。全局变换一般以矩阵代数理论为基础，用一个参数矩阵来描述图像的变换，典型的全局几何变换包括以下的一个或几个形式：平移、旋转、各向同性或各向异性的缩放、二次或三次多项式变换等；局部变换也称为弹性映射，允许变换参数有位置依赖性，也就是说，不同的位置具有不同的变换参数模型，变换参数只在特定的关键点上有定义，而在区域到区域之间进行插值；运动场方法的变换通常有直接匹配或基于特征的匹配两种。

3、相似性度量

相似性度量是两个图像达到配准与否的衡量尺度，也是剔除误匹配对的必经途径。常用的相似性测度共分为三种^[25-26]：距离度量、相关度测量和角度测量法。对于区域相关算法，通常采用相关函数作为相似性度量，如SSD函数、互相关、归一化相关系数（NCC）、相位相关等。对于特征配准算法，通常采用各种距离函数作为特征的相似性度量，如欧氏距离、Hausdorff距离、街区距离等。本文采用基于SIFT特征匹配算法，因此采用欧氏距离作为相似性度量。

4、搜索方法

搜索方法是在搜索空间中找到图像变换参数的最优估计，使得相似性度量达到极大值的方法。最佳搜索方法能最大限度地降低算法计算量，搜索过程实际上是最优化过程，按照图像参数是否可以直接计算，将搜索过程分为以下两种：参数可以直接计算的寻优和参数需要通过优化搜索的寻优，本文主要是针对后者进行研究。

2.4.2 图像配准方法

同一场景在不同条件下变换成的二维图像有较大的差异，主要是由于两幅图像可能是在不同时间、不同视角、不同传感器得到，并且目标景物的变化以及传感器本身存在缺陷。在这种情况下，匹配时间短、匹配精确度高、算法配准率高、算法复杂度低、可移植性好以及适应性强的匹配算法成为人们追求的目标。

目前，国内外已产生了不少的图像配准方法，各种方法都是面向一定范围的应用领域，也具有各自的特点，但并没有任何一种方法能对所有类型的图像都能获得最佳效果，各个领域的匹配方法都是对各自具体的应用背景结合实际情况量身定制的技术，大多数的匹配方法依赖于图像本身；但是不同领域的匹配技术在理论方法上又具有很大的相通性。根据图像信息的匹配方法不同，可以将图像配准分为三类：基于图像灰度的匹配，基于变换域的图像匹配、基于图像特征的匹配。

1、基于灰度信息的图像匹配^[27]

基于灰度信息的方法是以两幅图像重叠区域在RGB或CMY颜色系统对应灰度级的相

似性为标准查找图像的配准位置。该方法直接利用整幅图像的灰度信息进行配准,不需要对图像进行分割和特征提取,算法思路比较简单,易于实现,并且配准精度高、鲁棒性好。但这种配准方法往往依赖光照一致性的假设,且迭代要求初始化,对有光照、对比度变化和含噪图像不够稳健。要计算图像的所以灰度信息,计算量大,而且对于具有旋转、仿射等变换的图片不能进行处理,对图片的要求较高。

2、基于变换域的图像匹配^[28]

基于变换域的方法是先将图像转换到变换域中,在变换域下实现图像的配准。其中一个最常用的方法是相位相关法,此时的变换域为频域。相位相关法的优点在于计算速度较快,而且对于具有轻微旋转和缩放变换的图片也能够进行处理,相对于基于灰度信息的匹配方法适用性更广泛。它的缺点在于算法复杂度较高,而且要求待匹配图片具有较大的重叠区域,否则拼接可能失败。

3、基于特征的图像匹配^[29-31]

基于特征的配准方法通过像素导出图像的特征,图像中包含很多可以利用的特征信息,常用的特征包括闭合区域、边缘、特征点(角点、高曲率点等)、特征结构以及统计特征(如矩不变量、重心)等,然后以图像特征为标准,对图像重叠区域的特征集进行搜索匹配。

基于特征的图像匹配技术是近年来图像拼接技术的热点,主要包括以下三个步骤:

(1) 特征提取

根据参与配准图像的特征来决定使用何种特征进行匹配。选取的特征具体上来说上必须考虑以下三点:第一,必须是待匹配的两幅图像所共有的特征,如果参考图像和待匹配图像具有相似的尺度和灰度信息,我们可以考虑选取灰度作为特征空间;第二,特征集必须包含足够多的分布均匀特征,否则可能出现误匹配;第三,选取的特征必须易于进行下一步工作即特征匹配,以减少匹配算法的计算量。

(2) 特征匹配

在基于特征的图像匹配中,特征提取之后,可以直接进行特征匹配,就是在上步提取的特征集之间建立一个对应关系,可以直接根据特征周围像素建立此对应关系,但这样既不稳定还会存在大量冗余信息。对其进行改进,给每个特征赋予一定的描述符,使特征匹配过程成为搜索相近的特征描述符的过程。本文采用的尺度不变特征变换(SIFT)就是这样的一类描述符,该算子包含灰度信息和空间位置信息,在光照和尺度等变化下仍可保持一定的稳定性,本文将针对如何降低由于SIFT描述符造成的计算量大问题提出解决方案。

(3) 运动关系求解

运动关系求解的过程实际上是构造变换模型的过程,这里的运动关系通常是对应变换矩阵。得到特征匹配对后,就该构造变换模型,也就是通过特征匹配对确定参与匹配图像的变换关系,再将两者变换到参考图像的坐标系。从两幅图像上直接得到图像信息获取时摄像机的复杂运动关系几乎是不可能的,故求解变换矩阵可以通过迭代法实现,且每次迭代都是先随机选取一些特征匹配对求解一个初始的对应矩阵。

基于特征的方法的优点在于：第一，能够处理两幅图像间存在比较大的未对准情况；第二，它只利用了图像的某些特征，极大地减小了计算量、降低了配准时间；第三，由于是在既定的特征空间上进行匹配，因此对图像的光照和噪声的影响不是很敏感，对灰度和噪声的变化具有鲁棒性。缺点在于只利用了图像的一小部分特征，在特征提取和特征匹配存在错误的情况下，该方法配准容易失效。所以基于图像特征配准方法的关键在于如何提取和选择鲁棒的特征，以及采取何种方法进行特征匹配。

2.4.3 配准算法评价体系

由于图像类型、波段和数据信息量存在很大不同，甚至外部的噪声干扰，研究的重点和需求也各不相同，所以评价的标准不可能用一个指标来界定，通常需要一个评价体系综合评定，本节对常用的评价指标进行简要的介绍。

1、匹配时间：对于两幅待匹配图像，从原始图像的输入开始，到最后找到最好的匹配对，整个过程消耗的时间就是匹配时间。

2、匹配精确度：衡量匹配误差的一个评估标准。如果匹配误差在一个像素内，则称为像素级配准；如果配准误差小于一个像素，称为亚像素级配准，目前应用上常常要求匹配精确度在亚像素级范围内。

3、适用性：是指算法能不能对在图像采集中，由于传感器不同引起的差异有很好的共用性。

4、算法复杂度：算法复杂度分为时间复杂度与空间复杂度，分别指程序运行所需的时间、空间。当今计算机的内存与硬盘通常都很大，所以非特殊情况下，算法复杂度就是指时间复杂度，这是由于图像处理时效性要求较高，也就是说可能要求图像在传输的时候同步实现其拼接。时间复杂度与匹配时间的区别在于它不但需要考虑程序的运行时间，还要考虑在最坏的情况下，算法可能出现的所有问题。

5、算法配准率：多次匹配操作中，达到配准精度要求的正确匹配的次數与所有匹配操作次數的比例。

6、可移植性：包括软件、硬件可移植两方面，是指算法软件编程或移植到各类硬件系统上的可行程度。

2.5 图像融合技术

图像配准后，由于图像采集所带来的光照、视野、环境、图像源等差异，若将图像的重叠区域直接平均叠加，会在两幅图像重叠部分产生明显的缝合线或是重叠区域有重影现象，为了有效地消除拼接痕迹，图像融合技术应运而生。

图像融合是数据融合的一种，它是指利用计算机图像处理技术对不同条件下采集到的图像信息按照一定的准则加以自动分析、优化组合，以完成所需的决策和任务评估而进行的图像信息处理过程。图像融合算法应该满足：第一，融合方法应具有渐变性，以保证拼

合区域的平滑过渡；其次，图像融合处理应当只针对于拼接的重叠区域有效，不能损失原始图像的非重合区域的信息；最后，算法复杂度不能太高，不能影响了图像拼接的速度。图像融合算法应该满足：

根据图像的特征层来划分，图像融合可分为以下三类^[32-34]，各类图像融合的层次如图 2 所示。

1、像素级融合：

像素级融合也称数据级融合，是指原始图像信息不需要经过图像预处理，直接在图像数据层上进行数据关联融合处理。像素级融合通常采用集中式融合体系，这是最低层次的图像融合。

2、特征级融合：

特征级融合先对采集到的原始图像信息进行特征提取，再对提取到的特征信息进行综合分析和处理。特征级融合既有采用分布式的融合体系也可以采用集中式的融合体系，属于中间层次的图像融合。

3、决策级融合

决策级融合是最高层次的图像融合，主要是建立在认知模型方法的基础上，对多幅图像信息进行逻辑推理或统计推理，包含两个步骤，一是对各幅图像的基本处理，其中包括图像预处理、特征提取、识别或判决。再通过关联处理进行决策层融合判决，最后得到联合判断结果。

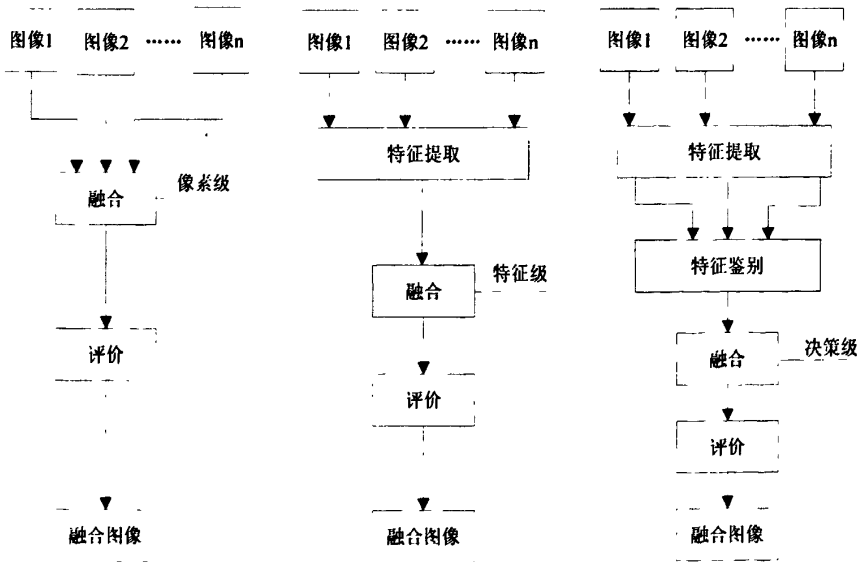


图2.2 图像融合的层次

2.6 本章小结

本章主要介绍了图像拼接的基本理论，包括图像拼接的各个流程：图像采集、图像预处理、几何校正、图像配准、图像融合以及各流程中的关键技术，对于决定图像拼接质量的图像配准技术进行了重点研究。实际应用过程中，应该根据图像类型及配准需求合理选

择图像拼接算法。

第三章 基于 SIFT 特征的图像配准算法

3.1 引言

图像拼接中最核心的技术环节就是图像配准问题，它直接关系到图像拼接算法的成功和效率。基于特征点的图像配准是目前应用最为广泛的一种图像配准方法，它实质上是通过对求取图像特征点集来寻找两幅图像之间的映射关系从而实现图像的配准。因为通过特征点提取算法得到的特征点数量较少，所以计算量小，能够兼顾实时性和有效性。

目前图像拼接算法中，研究得比较多的特征提取方法有：Harris角点算法、SUSAN角点算法、SIFT特征提取算法等。每种算子都有自身的一些特点，有着各自的使用场合，其中，SIFT特征描述符在对光照变化、旋转变换、尺度缩放、几何变形、模糊和图像压缩等6种情况下性能最好。近年来，SIFT特征匹配已成为国内外计算机视觉领域的研究热点和难点，它在很多领域，如图像配准、目标识别、全景图拼接，视频检索等，都有着广泛的应用。

对于基于特征点匹配的图像拼接方法，因为通过特征点提取算法得到的特征点的数量很少，所以图像拼接的效果很大程度上取决于是否能够提取到精确的特征点。Harris 角点检测算法是一种比较有效的特征点检测方法，它对于仅仅存在平移、选择和小尺度变换的图像匹配精度较高，这在很大程度上限制了它的应用。而尺度不变特征变换算法恰好解决了这个问题，SIFT 特征对于旋转变换、尺度变换、亮度变化都具有不变性，并且对于噪声、视角变化、仿射变换等具有良好的鲁棒性。

3.2 基于特征点的图像配准流程

基于特征点的图像配准算法基本流程是：

(1)特征点检测：特征点是图像信息最抽象的描述，在光照、视角以及场景变换的情况下，特征点相对来说不易受影响，所以采用特征检测算子和描述符提取图像的特征，具有一定得鲁棒性。

(2)特征点匹配：是指找出待配准的两幅图像中正确匹配的特征点对，通过分别提取两个或多个图像的点特征，对特征点进行参数描述，然后运用所描述的参数进行匹配。

(3)建立投影模型：通过建立投影模型，将图像绘制在投影模型上。其方法是根据匹配特征点之间的映射关系来表现图像之间的投影位置，从而将待拼接图像映射到指定坐标空间。

本文是基于 SIFT 特征点实现对手持数码相机采集到的原始图像进行拼接，算法流程如下图所示：

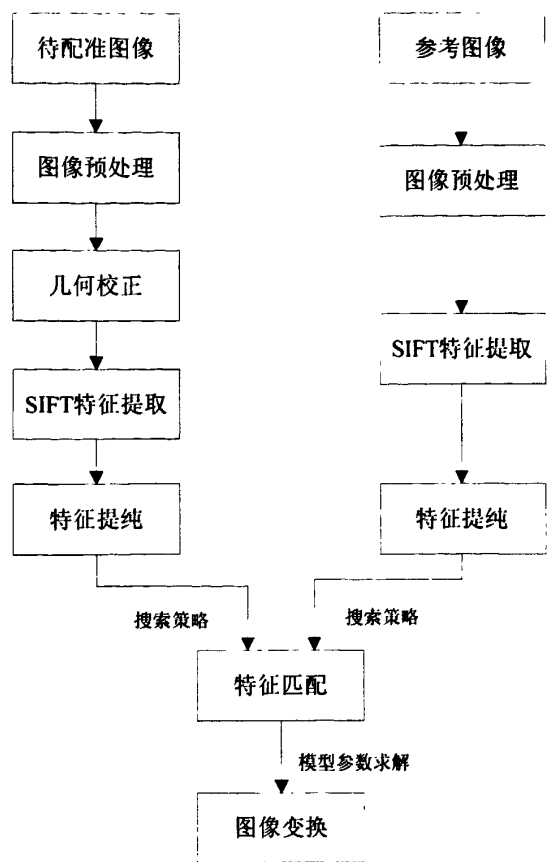


图 3.1 图像配准流程

3.3 SIFT 特征点提取

3.3.1 经典算法

David.Lowe 在 1999 年提出 SIFT 算子，并在 2004 年总结了现有的基于不变量技术的特征检测方法，正式提出了一种基于尺度空间的，对图像平移、旋转、缩放、甚至仿射变换保持不变性的图像局部特征描述算子^[31]。SIFT 特征的生成一般包括以下几个步骤：

1、构建尺度空间，检测关键点

为了使特征具有尺度不变性，特征点的检测通常是在多尺度空间完成的。尺度空间理论考虑了图像各结构在多尺度中的演化，其目的在于模拟图像信息的多尺度特征。

尺度空间理论基本原理是利用高斯卷积核对初始图像进行尺度变换，获得图像多尺度下的尺度空间表示序列，再对这些序列进行尺度空间特征提取^[35-36]。一幅图像在不同尺度下的尺度空间定义为图像 $I(x,y)$ 与高斯核 $D(x,y,\sigma)$ 的卷积：

一幅图像在不同尺度下的尺度空间定义为图像 $I(x,y)$ 与高斯核 $D(x,y,\sigma)$ 的卷积：

$$L(x,y,\sigma)=I(x,y)\otimes D(x,y,\sigma)$$

(3.1)

其中 L 为图像的尺度空间, (x, y) 为图像的像素坐标, σ 为尺度坐标。为了更快更好地在尺度空间检测到稳定的特征点, 在实际SIFT特征点提取中, 将图像金字塔与尺度空间结合起来, 得到高斯差分图像, 即DoG图像。具体方法是将输入图像通过不同尺度 σ 的高斯核函数连续平滑和下采样, 然后将各个图像层相邻尺度的两个高斯图像相减, 得到DoG图像。在图像尺度空间内遍历所有的点, 判断其与各个相邻点的大小, 如下图所示, 每个采样点要与它同尺度的8个相邻点和上下两相邻尺度各9个相邻点相比较, 如果该采样点为局部极值点(灰度极大值或极小值), 则视为候选特征点, 这样就能保证检测点在尺度空间和二维图像空间都是极值点, 也就初步确定了关键点所在的位置和对应的尺度。

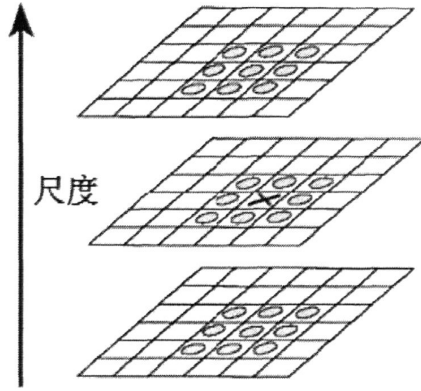


图 3.2 尺度空间检测极值点示意图

假设尺度空间坐标为 σ , 图像层坐标为 O , 亚像素坐标为 S , 则有:

$$\sigma(i, s) = \sigma_0 2^{i+s/S} \quad (3.2)$$

其中 σ_0 是图像基准层尺度, 图像层 $i \in i_{\min} + [0, \dots, O-1]$, 像素坐标 $s \in [0, \dots, S-1]$ 。另外,

第 i 组的空间坐标为 x_i , 可以得出:

$$x_i = 2^i x_0, i \in O, x_0 \in [0, \dots, N_0-1] \times [0, \dots, M_0-1] \quad (3.3)$$

(M_0, N_0) 为 $i=0$ 时的分辨率, $i \neq 0$ 时分辨率 (M_i, N_i) 由式(3.4)求出:

$$N_i = \left\lceil \frac{N_0}{2^i} \right\rceil, M_i = \left\lceil \frac{M_0}{2^i} \right\rceil \quad (3.4)$$

2、特征点过滤并进行精确定位

经过第一步得到极值点后, 接下来就是确定稳定特征点的位置、尺度等信息。经上面步骤产生的极值点并不都是稳定的特征点, 其中还存在对噪声敏感的低对比度的点以及不稳定的边缘响应点, 这两类点会影响匹配的准确度, 因此在特征点匹配前, 必须将其去除。

剔除低对比度的特征点时, 将DoG空间函数 $D(x, y, \sigma)$ 泰勒展开最高至平方项:

$$D(x,y,\sigma)=D+\frac{\partial D^T}{\partial x}X+\frac{1}{2}X^T\frac{\partial^2 D}{\partial x^2}X \tag{3.5}$$

其中， D 及其微分通过样本点来计算， $x=(x,y,\sigma)^T$ 为样本点的偏移量。

通过对该函数求导并将其导数值设为0，可得 x 的极值 $\hat{x}$ ：

$$\hat{x}=-\frac{\partial^2 D^{-1}}{\partial x^2}\frac{\partial D}{\partial x} \tag{3.6}$$

如果极值 $\hat{x}$ 在任一方向上大于1/2，就意味着样本点偏移量的极值与其它样本点非常接近，此时样本点可能会被改变，这就需要用插值来取代该样本点，具体做法是将偏移量 $\hat{x}$ 与其样本点相加，从而得到极值位置处的插值估计值，以确定关键点的准确位置和尺度。对于极值点 $\hat{x}$ ，其方程的值为：

$$D(\hat{x})=D+\frac{1}{2}\frac{\partial D^T}{\partial x}\hat{x} \tag{3.7}$$

当极值点对比度值 $D(\hat{x})$ 小于某一个阈值时，去除该极值点，在本文里，当 $|D(\hat{x})|<0.03$ ，则视为低对比度点过滤掉。

因为高斯差分算子会产生较强的边缘响应，需要剔除不稳定的边缘响应点。剔除边缘特征点，我们主要是利用处于图像边缘的特征点在DoG空间函数 $D(x,y,\sigma)$ 的 $D(x,y,\sigma)$ 的曲率值在横跨边缘部分较大，在垂直边缘的方向较小的性质。主曲率可以通过一个 2×2 的海森(Hessian)矩阵 H 求出。

$$H=\begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{3.8}$$

DoG空间函数 D 的主曲率值和海森矩阵 H 的特征值成正比例，所以直接计算 H 的特征值可以剔除边缘特征点，但前人通过实验研究发现，单单考虑矩阵 H 的两个特征值之间的比值，也可以完成上述判别。假设 α 为最大特征值， β 为最小特征值， α 、 β 的比值 r 可以通过下面公式计算：

$$Tr(H)=D_{xx}+D_{yy}=\alpha+\beta \tag{3.9}$$

$$Det(H)=D_{xx}D_{yy}-D_{xy}^2=\alpha\beta \tag{3.10}$$

其中 $Tr(H)$ 是矩阵 H 的迹，即 H 主对角线元素之和， $Det(H)$ 是矩阵 H 行列式的值。

如果 $Det(H)$ 为负值，那么曲率值的符号就不相同，此时该点不视为关键点，将其剔除。令

$\alpha=\gamma\beta$ ，则有：

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(\gamma + 1)^2}{\gamma} \quad (3.11)$$

由式(3.11)可以看出, 公式中最后只剩下变量 γ , 而不必计算 α 和 β , 根据最值原理, 当两个特征值一致时 $(\gamma + 1)^2 / \gamma$ 最小, 并与 γ 成正比。因此, 要判断主曲率是否小于阈值 γ , 只需要对下式进行判定:

$$\frac{Tr(H)^2}{Det(H)} < \frac{(\gamma + 1)^2}{\gamma} \quad (3.12)$$

当式(3.12)不成立时就说明边缘响应点不稳定, 过滤掉该点。本文取 $\gamma = 10$, 即如果主曲率间的比值大于10, 则认为该点是边缘响应点而被剔除。

3、为特征点分配方向值

上述两步确定了关键点的位置与尺度, 满足了尺度不变性, 为了使得检测出来的特征点还满足旋转不变性, 必需为特征点确定主方向, 以实现算子对尺度和方向的无关性。特征点的主方向即为该特征点邻域梯度的主方向, 具体求解为对关键点邻域范围内像素的梯度方向用直方图进行统计, 直方图的峰值所对应的方向即为特征点的主方向。如果还有其它峰值能量约为主峰值能量的80%, 则将其作为该关键点的辅方向。

关键点梯度的模值 $m(x, y)$ 和方向 $\theta(x, y)$ 计算公式如下, 其中 $L(x, y)$ 为关键点所在尺度的图像层。

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (3.13)$$

$$\theta(x, y) = \alpha \tan^{-1} 2(L(x+1, y) - L(x-1, y)) / (L(x, y+1) - L(x, y-1)) \quad (3.14)$$

在实际应用中, 在以关键点为中心的邻域内采样, 并用直方图统计邻域像素的各个梯度方向。梯度直方图的范围为0到360°, 每10°表示一个方向, 一共有36个方向。方向直方图的峰值代表了关键点处邻域梯度的主方向, 这样就确定了关键点的主方向。图3.3是采用7个方向时使用梯度直方图为关键点确定主方向的示例。为了增强匹配的健壮性, 只保留峰值大于主方向峰值80%的方向作为该特征点的辅方向。据统计, 15%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)。这种邻域方向性信息联合的思想增强了算法的抗噪能力, 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。

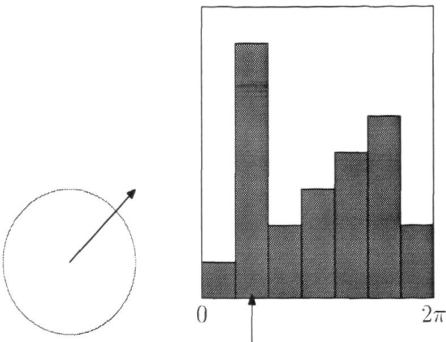


图3.3 由梯度方向直方图确定特征点主方向

4、生成特征描述符

至此，我们确定了特征点的位置、尺度和方向，接下来为每个特征点建立特征描述符，使其不随各种变化而改变，比如光照、视角、环境变化等。特征描述符应该有较高的独特性，以便于减小特征点的误匹配率。

生成特征描述符的过程如下图所示，图3.4(a)中央点为视为关键点的位置，图中蓝色圆圈代表高斯加权的范围（对越靠近关键点的像素梯度方向权值越高）。以特征点为中心取8*8的邻域作为采样窗口，将采样点与特征点的相对方向通过高斯加权后归入包含8个 bin 的方向直方图，最后获得4*4*8的128维特征描述符，图3.4(b)为关键点特征向量示意图，图中仅表示2*2个种子点，每个种子点含8个方向向量信息。

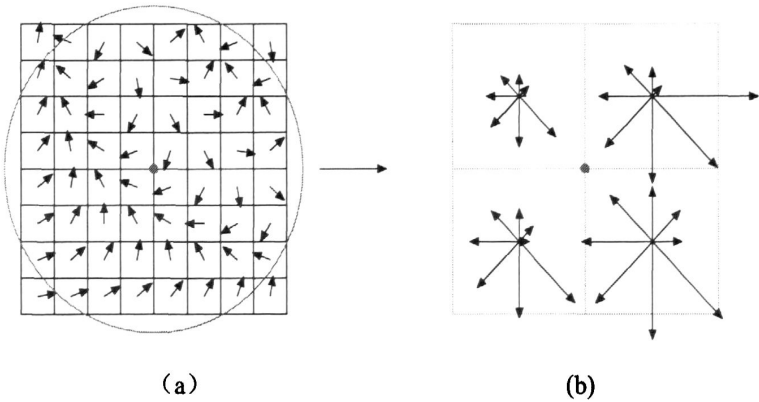


图 3.4 由关键点邻域梯度信息生成特征向量

3.3.2 算法后期改进

在特征点提取部分，经典 SIFT 算法具有以下不足：第一，相对于特征点的数量来说，其计算复杂度高且运算量大，速度较慢。这是因为提取 SIFT 特征是在多尺度空间完成，需要与高斯核进行数次卷积，而且为了产生精确的特征描述符，需要对直方图进行多次加权运算。第二，为了增强算法的鲁棒性，SIFT 算法需要能够检测出尽可能多的关键点，一幅典型的 480×640 的图像可得到 1000 个左右的 SIFT 特征点，如此庞大的特征点集会增加拼接系统中特征匹配的时间负担，降低算法运行效率。

为了减少关键点提取消耗的时间，在保证关键点质量的前提下应尽可能减少其数量。本文在特征提取部分做了如下改进：在构建高斯金字塔时减少图像采样形成的组数，经典 SIFT 算法要求一般产生 4 个图像组，本文综合考虑 SIFT 特征提取速度和精度，选择产生 3 个图像组。经大量实验证明这种改进做法是可行的既保证了关键点的质量，又减少了关键点的数量，在不影响算法的匹配精度条件下节约了算法匹配时间。

3.3.3 实验结果分析

图 3.5 是两幅大小为 480×640 的校园建筑物原始图，图 3.6 为经典 SIFT 特征提取结果，图 3.7 为改进 SIFT 特征提取结果，表 3.1 为两种算法 SIFT 特征点的数目和提取时间的比较。由实验结果可知：改进后算法中，提取的特征点数量和时间都大大减少，从而节约了算法的匹配时间，加快算法的匹配速度。

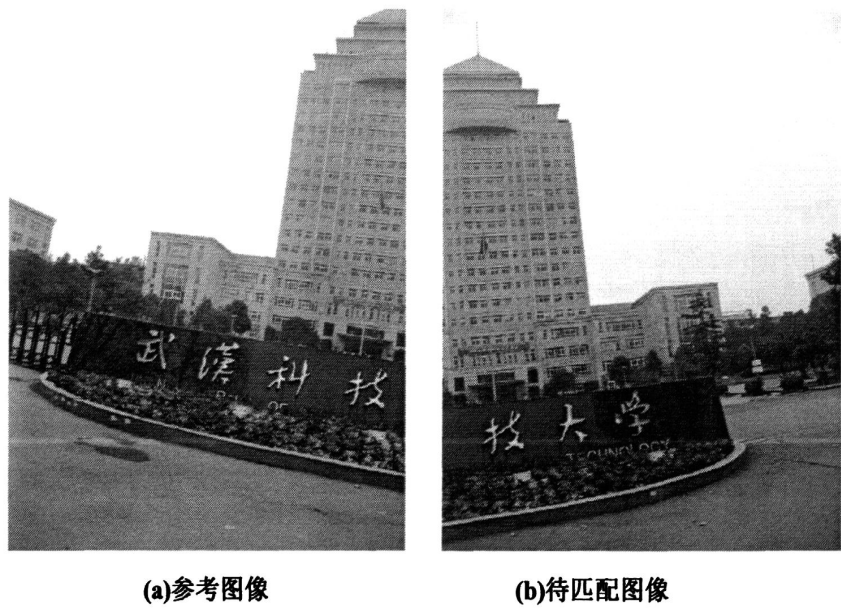


图 3.5 景物原始图

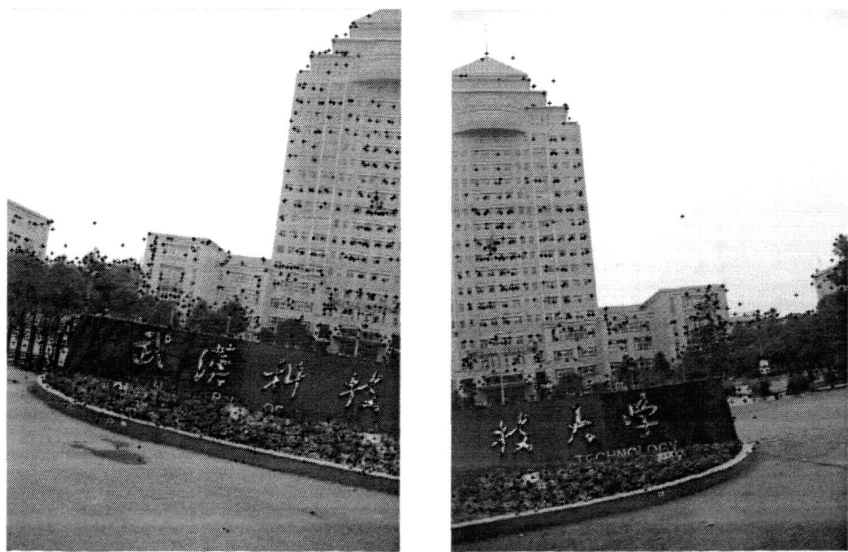


图 3.6 经典 SIFT 特征提取结果

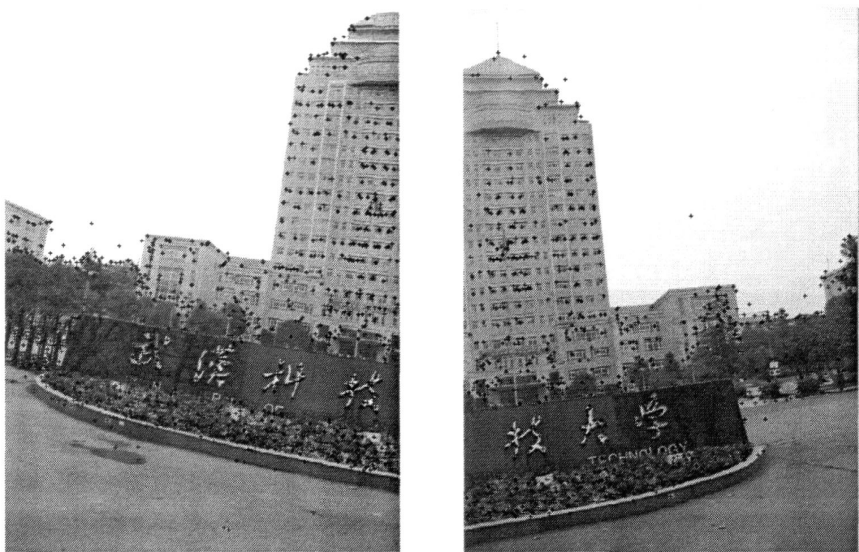


图 3.7 改进 SIFT 特征提取结果

表 3.1 两种 SIFT 算法特征提取性能比较

算法	参考图像 SIFT 点数/提取时间	待匹配图像 SIFT 点数/提取时间
传统 SIFT 算法	771/2.4s	597/1.8s
改进 SIFT 算法	381/1.5s	326/1.3s

3.4 SIFT 特征点匹配

特征描述符对关键点各种属性进行了详尽的表述，比如位置、尺度、方向等，因此特征描述符之间具有很大的差异性，所以可以直接利用特征描述符进行特征匹配。为了减少一个特征点可能存在几个相似匹配点而造成的匹配出错情况，这里采用最近邻和次近邻特征点距离之比来减少误匹配。如果最近的距离和次近的距离比值小于某个阈值 n ，则认为该点对为匹配点对，否则丢弃。降低阈值，SIFT匹配点对数目会减少，但更加稳定，综合各个因素考虑，本文算法中 n 取0.5。

3.4.1 Kd-Tree 算法

Kd-Tree也叫K维搜索树^[37]，其中K表示空间维数，是二叉树在维数上的扩展。Kd-Tree的每一层将特征空间分成两部分，树的顶层节点按一维进行划分，下一层节点按另一维进行划分，以此类推，各个维循环往复。划分时要保证两边子树节点数大致相等的，即每个节点，存储在子树中一半的点落入左子树，而另一半落入右子树，这样就可以保证二叉树提供的选择机会最大。当某个节点中的总特征点数少于给定的最大点数时，划分结束。

因为很多节点，甚至某些子树无需考虑，Kd-Tree的优点是能够减少查询量，并且

Kd-Tree提供了一种查询的高效机制,从而大大降低了寻找最佳匹配的计算量。因此Kd-Tree搜索算法在K较低时效率非常高,但对于K较高时其执行效率较低,一般对于维数大于10时,表现出较差的搜索效率,此时基本等同于穷举法。

3.4.2 BBF 搜索算法

近似最近邻算法^[38], Best-Bin-First(BBF), 是对Kd-Tree算法的改进,通过上节分析,我们知道 Kd-Tree搜索算法大部分时间都用来查询节点,但只有小部分节点满足最近邻条件,基于 Kd-Tree搜索算法这一缺点,我们引入BBF方法。该方法采用一个优先级队列使搜索依次从节点与被查询节点距离由近及远的顺序进行,它限定Kd-Tree中叶子节点数,即限定了搜索的最大次数,能快速的找到最近邻和次近邻点,极大提高了搜索效率。

BBF算法之所以更快速是因为仅考虑那些次近最邻近点的距离0.8倍以内的最邻近点,这样就避免了多个距离相近邻接点的麻烦。并且上文提到,Kd-Tree搜索算法对10维以上层次效率较低,而SIFT特征是一个128维特征向量,所以BBF搜索策略较为实用,因为其改变了常规Kd-Tree的搜索结果,对最近邻搜索算法在高维度空间上有了较大的提高。

3.5 SIFT 特征点提纯

3.5.1 RANSAC 算法

随机采样一致性算法是计算机视觉领域中最应用最广的稳健估计方法,属于假设-验证估计方法,在基础矩阵估计、特征匹配、运动模型选择等领域中都得到了广泛的应用。

RANSAC是一种鲁棒的变换估计算法,利用特征集合的内在几何约束关系进一步剔除错误的匹配。下面以直线拟合为例来阐述RANSAC算法的主要思想^[39]:首先随机地选择两个点,这两个点确定了一条直线,在这条直线一定距离范围内的点称为这条直线的支撑。随机选择重复数次,具有最大支撑特征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称为内点,反之则为外点。由这个内点集合线性解出新的直线,再根据这条新直线寻找它对应的内点,不断地重复这样的随机采样,直到某次采样使内点数量最大,那么得到的直线估计就是该集合的最优估计。根据算法描述可以很快判断,随机选取的包含外点的初始点集确定的直线不会获得很多特征点支撑,因此点1、点10作为外点被剔除,如图3.8所示。

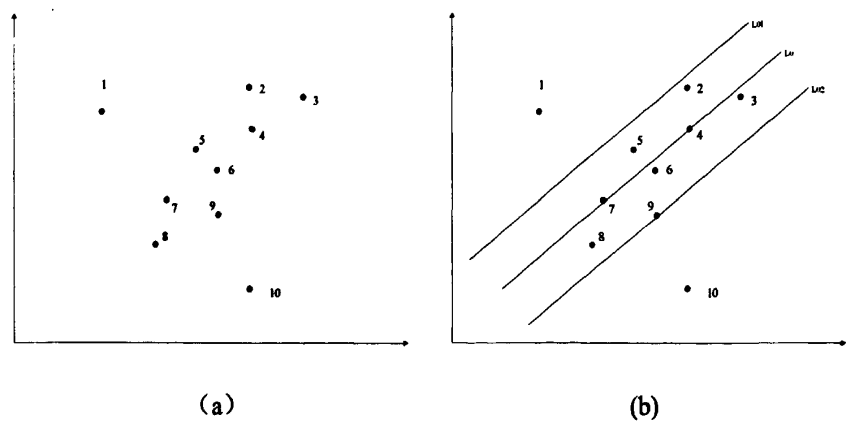


图 3.8 RANSAC 直线拟合模拟

由上述分析, 可将RANSAC算法归纳为: 假设 M 个数据点组成的数据点集 Q , 在 Q 中有一定比例的数据点满足特定参数的某种模式, 则此部分数据点被称为内点, 剩余的数据点被称为外点。算法目标就是要从参数空间中求出模式参数, 该模型的参数至少需要 n 个数据点求出($M > n$), 该模式参数可以使代价函数取最大值。

- 1、从 Q 中随机选取含有 n 个数据点的样本 S , 样本应该足够小。
- 2、由所选取的 n 个数据点计算出模式参数, 从而确定模型 H 。
- 3、对 Q 中其余的 $M-n$ 个数据点, 计算出它们与模型 H 之间的距离, 记录在 H 的某个误差允许范围内的 Q 中数据点的个数 C 。

重复以上三个步骤 K 次, 对应最大 C 值的模型即为所求模型, 数据集 Q 中的这 C 个数据点为内点, 其余的 $M-C$ 个数据即为外点。由于充分使用了所有的测量数据, 找到精确的内点, 并利用内点来进行参数估计, 进而得到了最佳结果。

显然, RANSAC算法有3个需要确定的量: 内点与外点的距离阈值、随机采样次数 K 和最大内点个数 C 。

理想的随机采样是搜索所有可能的组合, 但这样计算量巨大, 在实际应用中不可能做到。根据图像变换模型不同, 需要的匹配点对不同, 若为仿射变换, 需要4对匹配点, 若为投影变换, 则需要的匹配点对至少是8对。假设模型参数估计需要的 r 对匹配点, 那么可以确定一个恰当的采样次数 K , 以保证采样的 r 对匹配点都是内点的概率足够高, 设以 P 来表示此概率, 大量实验证明, 实际采样时只要保证 $P > 95\%$ 就能满足要求。

3.5.2 改进 RANSAC 算法

使用RANSAC算法对数据点进行处理后, 得到的内点数据集中, 仍然存在不满足要求的点, 此情况下, 若从内点数据集中直接取出满足仿射变换的3对点进行后续处理, 则很大程度上会出现由于两幅图像变换关系定位不准而导致拼接图像存在鬼影现象。

在此基础上本文采用二次导向匹配的方法提取仿射变换矩阵, 首先用RANSAC确定配准误差阈值, 然后将此阈值的90%大小作为配准阈值提取最终的仿射变换矩阵。区别于传

统RANSAC算法，本文采用3对点提取仿射矩阵并检验仿射的效果。其具体方法如下：

1、在程序执行初始阶段，设置最高循环次数K解决可能因阈值设置不合理导致程序运行出现死循环的现象；设置平均几何配准误差阈值E1，在本文里初始阈值设为0.08，其中平均几何配准误差阈值表示为：

$$E_n=\frac{1}{2n}\sum_{i\in n}(d(x_i,Hx_i)+d(x_i+H^{-1}x_i)) \tag{3.15}$$

2、从内点数据集中任意取 6 个点，用 3 个点拟合变换关系，再用另外 3 对点检验配准误差的大小，通过这 3 对点根据式(2.4)(2.5)计算两幅图像之间的仿射变换矩阵 H。

3、再次从内点数据集中取 6 个点，求出仿射变换矩阵 H 的平均几何配准误差 En，并将结果与 E1 相比较，当 En>E1 时重复上步，直到满足程序运行循环退出条件，若 En<E1，则得出 H，退出循环。

4、对于每次得到的 En，选取其最小值，并将该值大小的 90%作为新的阈值 E2，然后反复执行第二步和第三步。

传统 RANSAC 算法只从内点数据集中抽取一个点计算仿射变换矩阵性能的好坏，改进后的 RANSAC 算法抽取 3 个数据点对进行验证，这样就减少了差错出现的可能性，即提高了算法匹配精度和算法配准率，使拼接效果趋于理想。

3. 5. 3 实验结果分析

第一组实验源图像为图 3.5 中的两幅武汉科技大学学校正门图像，对源图像用 3.3 小节中的改进算法提取 SIFT 特征，再用 BBF 搜索算法进行候选匹配，此时还存在比较多的错误和冗余匹配点，利用本文提出的改进 RANSAC 算法进行提纯，提纯后得到两幅图像间的变换矩阵为 H_b ，而基于传统的 RANSAC 算法得到的两幅图像间的变换矩阵为 H_a 。具体参数表示如下：

$$H_a = \begin{bmatrix} 0.986354 & -0.000233 & -208.956143 \\ -0.001638 & 0.979386 & 0.894526 \\ -0.000011 & -0.000005 & 1.000000 \end{bmatrix}$$
$$H_b = \begin{bmatrix} 0.986354 & -0.000233 & -210.437269 \\ -0.001638 & 0.979386 & 0.807483 \\ -0.000010 & -0.000005 & 1.000000 \end{bmatrix}$$

图 3.9 为传统 RANSAC 算法拼接结果，图像分辨率为 936×708，图 3.10 改进 RANSAC 算法拼接结果，图像分辨率为 937×724。图 3.5 中两幅原始图像视角变换较大，在教学楼顶部，图 3.9 在主楼顶部有一明显弯曲变形，重叠部分衔接出现偏差，而图 3.10 则较好的解决了这一问题。

通过第一组实验可知，本文提出的算法匹配更为准确，得到的配准结果也较为理想。从算法的执行时间和正确匹配率两个方面比较两者的求解效率，比较结果如下表 3.2 所示。



图 3.9 传统 RANSAC 算法拼接结果 1



图 3.10 改进 RANSAC 算法拼接结果 1

表 3.2 传统 RANSAC 算法与改进的 RANSAC 算法性能比较 1

算法	运行时间(s)	正确匹配率(%)
传统 RANSAC 算法	3.5	79.3
改进 RANSAC 算法	3.9	94.1

另外一组实验是取自学校主楼 5 幅分辨率为 480×640 图像序列，如图 3.11 所示。图 3.12 为传统 RANSAC 算法拼接结果，图 3.13 为改进 RANSAC 算法拼接结果，两幅拼接后图像分辨率皆为 1400×632 。比较两幅图像得知，在图 3.13 中心区域出现了花坛的阴影，通常称为“鬼影”现象，而在图 3.14 中，“鬼影”现象完全消除，拼接效果较好。表 3.3 为从执行时间和正确匹配率方面比较两个算法的性能。

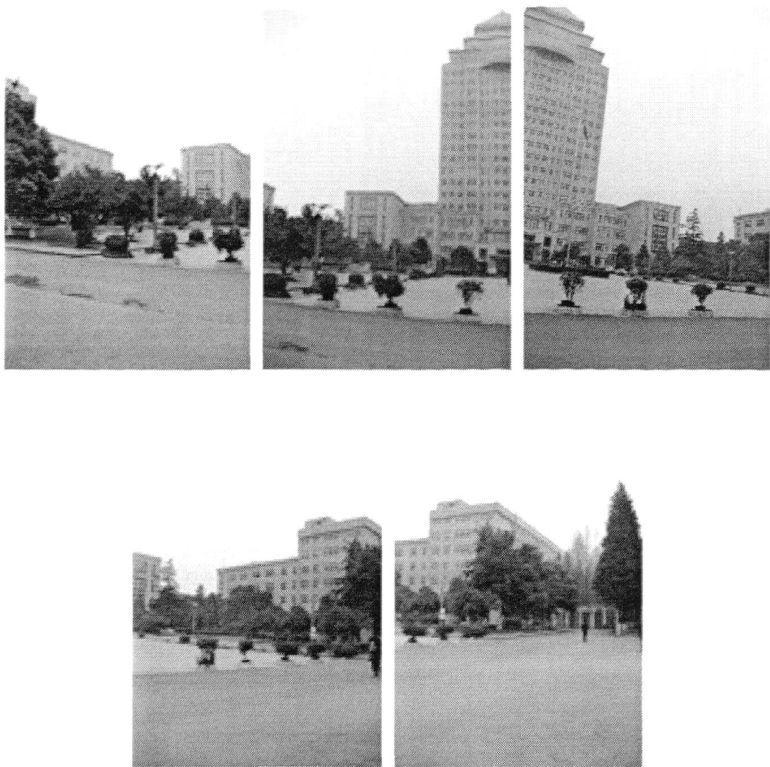


图 3.11 教学楼图像序列



图 3.12 传统 RANSAC 算法拼接结果 2



图3.13 改进RANSAC算法拼接结果2

表 3.3 传统 RANSAC 算法与改进的 RANSAC 算法性能比较 2

算法	运行时间(s)	正确匹配率(%)
传统 RANSAC 算法	8.4	76.4
改进 RANSAC 算法	9.1	95.2

从表 3.2、3.3 可以看出，改进 RANSAC 算法极大地增加了特征的正确匹配率，可是由于增加了处理的点对和程序的循环次数，所以增大了匹配时间和算法时间复杂度，但相对来说增加的时间不多，一般情况下为 10%。总体来说，改进 RANSAC 算法较好的折中了算法的效率和速度。

3.6 本章小结

本章首先介绍了基于SIFT特征的图像配准经典算法，对于该算法中存在的特征匹配时间较长和错误匹配的不足，在SIFT特征点提取方面，综合考虑提取速度和精度，在构建高斯金字塔时选择3个图像采样组，在SIFT特征点匹配方面，在Kd-Tree搜索算法基础上，采用BBF搜索算法提高关键点的搜索效率，在SIFT特征点提纯部分，采用改进的RANSAC过滤外点，去除误匹配对。实验证明本文算法可以较好的解决经典SIFT算法存在的上述问题。

第四章 图像融合

4.1 引言

由于待拼接的各个图像可能不是在同一时刻拍摄，会存在亮度上的差异。此外，因为使用普通相机进行图像采集，待拼接图像中可能出现边缘失真现象，而图像拼接正好要充分利用到图像的边缘部分。以上原因造成配准后的图像在强度或颜色上不连续，存在明显的拼接缝隙，必须对图像的重叠部分进行适当的处理。图像融合就是为了消除图像之间的拼接缝隙，实现拼接图像之间的平滑过渡。

4.2 图像融合经典方法

像素级图像融合是最基本的处理手段，是其它图像融合的必经阶段，也是研究最多的一种，目前主要有以下几种方法：

1、直接平均法

配准后图像的重合部分像素点的灰度值采用其灰度的平均值，直接平均法作用类似于低通滤波器，实现速度很快。但在重合部分会有明显的带状痕迹，使得最终的拼接图像存在明显的拼接缝隙。

设 f 为融合后的图像， f_1 和 f_2 分别表示待拼接的原始图像，则有：

$$f(x, y) = \begin{cases} f_1(x, y) & (x, y) \in f_1 \\ [f_1(x, y) + f_2(x, y)] / 2 & (x, y) \in f_1 \cap f_2 \\ f_2(x, y) & (x, y) \in f_2 \end{cases} \quad (4.1)$$

2、加权平均法

加权平均法^[40]是从直接平均法改进而来，对重合部分的图像灰度值先加权再叠加平均。

设 f 为融合后的图像， f_1 和 f_2 分别表示待拼接的原始图像，则有：

$$f(x, y) = \begin{cases} f_1(x, y) & (x, y) \in f_1 \\ \omega_1(x, y)f_1(x, y) + \omega_2(x, y)f_2(x, y) & (x, y) \in f_1 \cap f_2 \\ f_2(x, y) & (x, y) \in f_2 \end{cases} \quad (4.2)$$

式(4.2)中的 ω_1 ， ω_2 分别是两幅原始图像重叠区域对应像素的权值，且 $\omega_1 + \omega_2 = 1$ ，其中 $0 < \omega_1 < 1$ ， $0 < \omega_2 < 1$ 。选择适当的权值可以实现重合部分的平滑过渡，基本上消除拼接缝隙。该方法直观简洁，速度较快，是比较常用的一种算法。

根据选择不同权重分布函数函数 $\omega_1(x, y)$ ， $\omega_2(x, y)$ ，加权平均法分为：

(1) 帽子函数加权平均法

帽子函数加权平均法基于下面的假设：图像像素对最终合成图像的贡献跟该像素与中心的距离成反比。所以在该方法中，图像中心区域像素占有较大的权重，边缘区域像素占较小的权重。权值函数 $\omega_i(x,y)$ 采用帽子函数（hat function），如式(4.3)所示，

$$\omega_i(x,y) = (1 - \left|\frac{x}{width_i} - \frac{1}{2}\right|) \cdot (1 - \left|\frac{y}{height_i} - \frac{1}{2}\right|) \tag{4.3}$$

其中 $width_i$ 和 $height_i$ 分别表示图像的宽和高，帽子函数表现为三角形，其函数图像如图4.1所示。

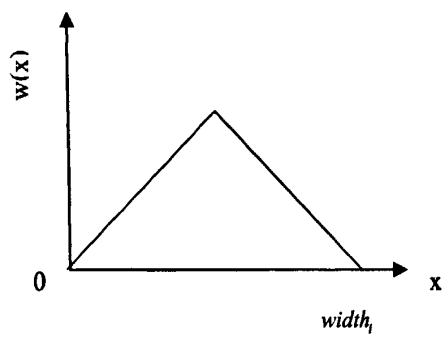


图4.1 帽子函数示意图

为了满足 $\sum_i \omega_i = 1$ ，对式(4.3)进行归一化处理，并代入式(4.1)，得到帽子函数加权平均算法的数学表达式表示如下：

$$f(x,y) = \frac{\sum_i \omega_i(x,y) f_i(x,y)}{\sum_i \omega_i(x,y)} \tag{4.4}$$

(2) 渐入渐出法

渐入渐出法是由 Szeliski 提出的，是一种特殊的加权平均法，因此也满足式(4.1)，只是这里的权重值 ω_1, ω_2 习惯用 d_1, d_2 表示，且 d 由重叠部分宽度决定，满足 $d_i = 1 / width$ ，其中 $width$ 表示重叠部分的宽度。在重叠部分中， d_1 由 1 慢慢减小到 0， d_2 由 0 慢慢增大到 1，由此实现在重叠部分由 f_1 渐进的平滑过渡到 f_2 ，如下图 4.2 所示。

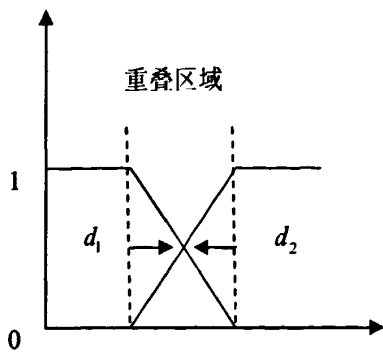


图4.2 渐入渐出法融合示意图

3、中值滤波法

中值滤波法的主要思想是利用中值滤波器来解决图像重合部分的像素值的突变^[41]。具体做法是：在图像的边界附近的狭长地带，对每一点都与其相邻点的像素灰度值进行比较，如果两者的比值超过设定的某一阈值时，对该像素点进行中值滤波，这样就消除了任意点的像素值的突变，从而保持了光强的连续性。中值滤波法可以较好的解决图像细节模糊问题，保持较高频的图像信息或者突出待拼接图像中变化的目标，提炼出图像的原始背景。

4、多分辨率样条融合法

多分辨率方法是将图像分解成一系列具有不同分辨率、频率特性及方向特性的子带图像，然后在对各子带图像采用加权平均融合法，最后使用重构算法生成原分辨率下重叠区域的图像^[42-43]，具体过程如图4.3所示。该方法优点在于不但可以有效保留图像细节信息，还可以使得图像低频信息实现平滑过渡，拼接后的图像背景较为连贯。

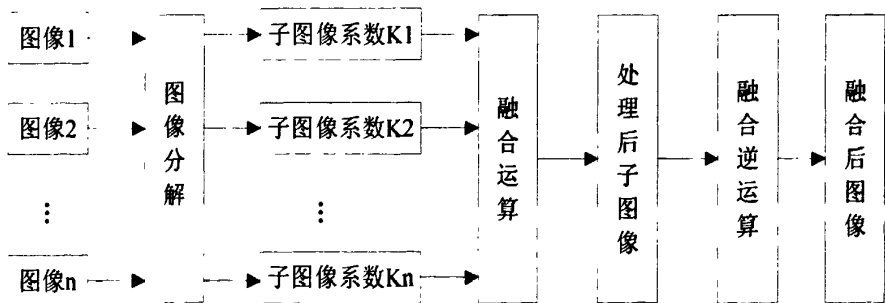


图 4.3 多分辨率样条融合法

多分辨率方法可以使具有不同强度的图像平滑过渡，所以融合质量高，但由于要对所有频率带的边界狭长地带都进行融合运算，工作量相对来说较大，计算时间长，在一般要求下的图像拼接中应用较少。

4.3 实验结果分析

在实际应用过程中，必须综合成像条件、图像特征、应用场合及需求精度来考虑，选择合适的图像融合方法，使处理的速度和效果两方面能达到均衡。

加权平均法是从图像的中心出发，对整个图像灰度信息都进行了加权平均，该方法直观简洁，速度较快，是比较常用的一种算法。本文融合部分使用渐入渐出加权平均法。

图4.4为2幅480×640武科大主楼图片，图4.5是采用直接拼接的方法得到的拼接图像，未经过图像融合处理，从图中可以明显的看出:由于两幅图片的光照条件不同，重叠区域存在明显的过渡带。经由渐入渐出加权平均法图像融合处理后的拼接结果如图4.6所示，拼接结果自然无缝，图像重叠区平滑过渡，基本实现了无缝拼接。

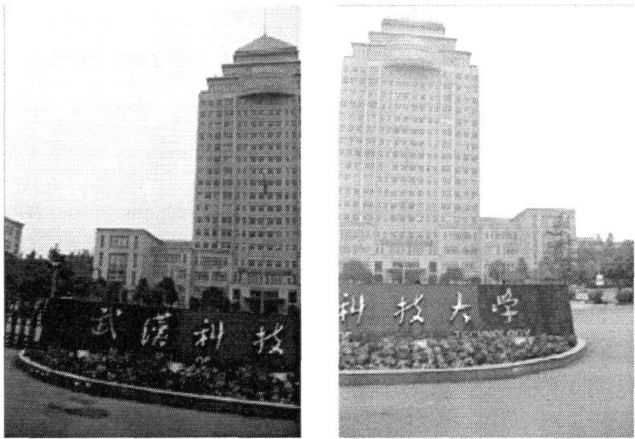


图4.4 原始图像



图4.5 未经图像融合的拼接图像



图 4.6 图像融合处理后的拼接图像

4.4 本章小结

本章重点研究了图像融合的一些经典方法，对其优缺点及实际应用场合进行分析。针对本文实验图像的特点，应用渐入渐出法对配准后的图像进行融合处理，实现了重叠区域的平滑过渡，达到无缝拼接的目的。

第五章 基于 SIFT 特征的全景图像生成

5.1 全景图像拼接流程

全景图像是指在固定的视点，用照相机或摄像机按照一定的方式采集图像，采集之后的图像输入计算机进行图像拼接，融合等处理生成无缝全景图像。全景图像拼接是一项复杂的工作过程，其典型的系统一般分为具有样本的采集、模板提取、搜索配准点、全局配准和无缝拼接这几个模块，本文图像拼接算法总体框架如图5.1所示。

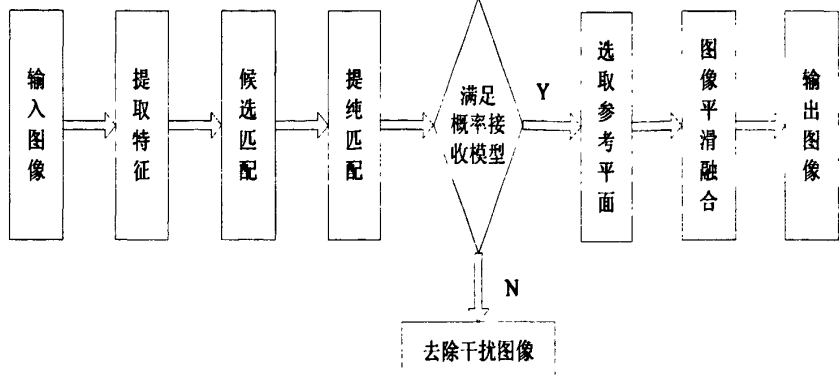


图 5.1 基于 SIFT 特征的全景图像拼接流程

5.2 投影模型选择

投影模型一般有平面、圆柱面、球面或多面体^[44]，选取适合的投影模型，使全景图像呈现在投影平面上，表现出真实场景效果。其中，球面和多面体模型能反映视野全方位的空间信息，柱面模型只能反映视点前后左右的空间信息，不能反映上下变化的情况。相对于球面和多面体模型在局部图像合成算法方面难度较大，处理时间长，圆柱面全景图像构造简单，并且在很多情况下并不需要所有的空间信息，因此柱面模型是应用最多的投影模型。

柱面投影是一种透视投影，就是将多幅待拼接图像投影到一个圆柱面上，即需要得到从视点上观察图像在柱面的成像^[45]。

对于手持相机拍摄照片进行拼接，可选取柱面或球面投影模型，在水平方向的图像拼接柱面和球面的效果区别不大，有垂直方向图像进行拼接时，考虑采用柱面投影模型。本文采用柱面投影模型，将所有的图像都投影到参考图像所在的柱面坐标系下后，将图像间的重叠部分的冗余信息去掉，连接起所有的图像，就得到了360柱面全景图。

5.3 基于概率接收模型的确认匹配

第三章研究了两幅图像的配准算法，而实际中往往有多幅图像需要拼接。这首先需要

确定两两图像相邻匹配的情况，本文引入概率模型的方式来确定图像的匹配关系。

通过第三章改进的 RANSAC 算法，我们可以得到满足仿射变化矩阵的候选匹配特征点即内点。候选特征匹配点 i 是否为内点记为事件 f_i ，图像匹配正确与否记为事件 m ，图像重叠区的总特征点个数记为 n_f ，内点数记为 n_i 。因为 $f_i = \{0,1\}$ 且 f_i 和 f_j 相互独立 ($i \neq j$)，所以事件 f_i 服从独立的贝努利分布。则重叠区的内点总数 n_i 可表示为：

$$n_i = n_1 + n_0 \quad (5.1)$$

其中：

$$\begin{aligned} n_0 &= \sum_{i=1}^{n_f} p(f_i | m=0) p(m=0) \\ &= \sum_{i=1}^{n_f} B(n_i, n_f, p_0) p(m=0) \\ &= \sum_{i=1}^{n_f} \frac{n_f!}{n_i! (n_f - n_i)!} p_0^{n_i} (1 - p_0)^{n_f - n_i} p(m=0) \end{aligned} \quad (5.2)$$

$$\begin{aligned} n_1 &= \sum_{i=1}^{n_f} p(f_i | m=1) p(m=1) \\ &= \sum_{i=1}^{n_f} B(n_i, n_f, p_1) p(m=1) \\ &= \sum_{i=1}^{n_f} \frac{n_f!}{n_i! (n_f - n_i)!} p_1^{n_i} (1 - p_1)^{n_f - n_i} p(m=1) \end{aligned} \quad (5.3)$$

其中， p_1 表示当图像匹配正确时对应特征点为内点的概率； p_0 表示当图像不匹配时对应特征点为内点的概率。由全概率定理可得图像正确匹配的概率 $p(f_i)$ ：

$$p(f_i) = p(f_i | m=1) p(m=1) + p(f_i | m=0) p(m=0) \quad (5.4)$$

根据贝叶斯准则可以计算出图像正确匹配的后验概率 $p(m=1 / f_i)$ ：

$$\begin{aligned} p(m=1 / f_i) &= \frac{p(f_i / m=1) p(m=1)}{p(f_i)} \\ &= \frac{p(f_i / m=1) p(m=1)}{p(f_i / m=1) p(m=1) + p(f_i / m=0) p(m=0)} \end{aligned} \quad (5.5)$$

图像正确匹配需满足下列条件：

$$p(m=1 / f_i) > p_{\theta \min} \quad (5.6)$$

将式(5.5)代入式(5.6)有：

$$\frac{p(f_i / m=1)p(m=1)}{p(f_i / m=1)p(m=1) + p(f_i / m=0)p(m=0)} > p_{\theta \min} \quad (5.7)$$

式(5.7)可进一步简化为:

$$\frac{p(f_i / m=0)p(m=0)}{p(f_i / m=1)p(m=1)} < \frac{1}{p_{\theta \min}} - 1 \quad (5.8)$$

将式(5.2)和式(5.3)代入式(5.8)有:

$$\frac{p(m=0)}{p(m=1)} \cdot \left(\frac{p_0}{p_1}\right)^{n_i} \cdot \left(\frac{1-p_0}{1-p_1}\right)^{n_f-n_i} < \frac{1}{p_{\theta \min}} - 1 \quad (5.9)$$

将具体数值代入式(5.8)进一步计算, 其结果可简化为:

$$n_i > \alpha + \beta \cdot n_f \quad (5.10)$$

当内点数和特征点总数满足上述条件时, 图像匹配正确; 否则匹配错误。其中参数 α 和 β 由 p_0 、 p_1 、 $p(m=1)$ 和 $p(m=0)$ 决定。本文各参数选择分别为 $p_{\theta \min} = 0.99$, $p_0 = 0.5$, $p_1 = 0.2$, $p(m=1) = p(m=0)$, 则匹配确认条件为:

$$n_i > 1.37 + 0.05 \cdot n_f \quad (5.11)$$

5.4 参考平面的选择

对于多幅图像的配准, 需要确定一个合适的参考平面, 然后将局部配准得到的运动关系转换为各图像与参考平面的运动关系。参考平面的选择在多幅图像拼接中其中关键的作用, 因为随机选取某幅图像作为参考平面有可能使配准累积误差太大使得序列图像拼接失败。

实际应用中, 参考平面的选取要重点考虑参考图像与待匹配图像的重叠情况, 包括重合部分的面积与特征匹配对的数目。当重合部分所含特征匹配对较少时, 应侧重考虑重合部分的面积大小对参考平面选取的影响; 对于特征点数目较多的图像, 则应侧重考虑重合部分里特征匹配对数目的影响。参考图像得选择用数学表达式^[54]表示为

$$T_i = \sum_{j \in I(i)} (S_{ij} + \omega M_{ij}) \quad (5.12)$$

式中 $I(i)$ 表示与图像 I_i 有重合部分的图像的集合, S_{ij} 为重合部分的面积, M_{ij} 为重合部分特征匹配对的数量, ω 为权重系数, ω 随图像含有的特征点个数增加而增加。每幅图像分别经过以上运算, 得到 T_i 值最大对应的图像即为参考图像, 该图像对应的平面为参考平面。

在确定参考平面后, 第二步就是计算待匹配图像到参考平面的对应矩阵。对应每幅待

匹配图像 I_i ，找到与该图像特征匹配对最多的具有重叠部分的图像作为最佳相邻图像 I_j 。

再计算出 I_i 到 I_j 的对应矩阵 H_{ij} ，则图像 I_i 到参考平面的对应矩阵 H_i 可表示为

$$H_i = H_{ij} \cdot H_j \tag{5.13}$$

5.5 全景图生成

实验采集的图像序列共有 10 幅图像，从左至右、从上到下分别为第 1 幅到第 10 幅图像，各个图像分辨率为 384×512 ，如图 5.2 所示。对于多幅图像，引入概率接收模型，计算每两幅图像匹配对应的匹配特征点个数和内点个数，得到各个图像的最佳相邻图像，去除无关图像，如本例中的第 5 幅图像，再进行图像拼接的后续步骤。实验结果如图 5.3、5.4 所示，两幅结果图像分辨率分别为 1400×286 ，其中，图 5.3 为未经融合处理的拼接结果，由于光照差异，图像中出现了明暗交替的现象，存在多个过渡带，图 5.4 为图像融合后的处理结果，可以看出拼接缝隙消除，基本实现了重叠区域的平滑过渡。

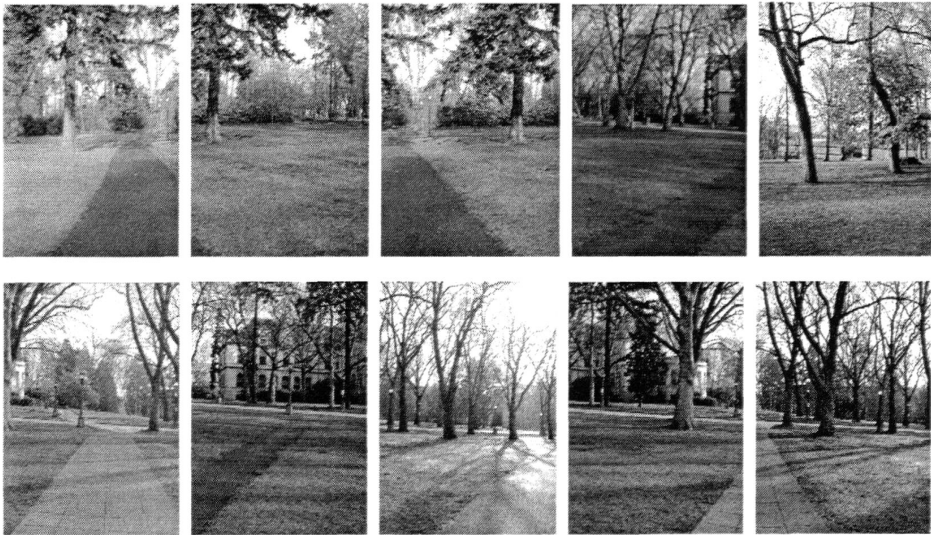


图 5.2 待拼接图像序列



图 5.3 未经图像融合的拼接图像



图5.4 图像融合处理后的拼接图像

5.6 本章小结

本章首先简要介绍了全景图像生成的流程，重点研究了全景图像生成的关键问题：投影模型和参考平面的选择，对于使用手持相机拍摄的景物图像进行拼接，本文选择柱面投影模型。最后在两幅图像拼接的基础上，实现了若干幅图像的基于SIFT特征的全景图像自动拼接，并用实验证明了该算法的可行性。

第六章 总结与展望

6.1 论文工作总结

本文主要研究了基于特征点的图像拼接的关键技术，对经典的基于SIFT图像配准算法进行了深入分析，针对该算法中存在的特征匹配时间较长和错误匹配的不足，提出了一种改进和优化方法，基本实现了图像的精确配准。完成配准后，采用渐入渐出加权平均法对拼接结果进行图像融合，最后实现了多幅景物图像的全景图自动拼接。

基于SIFT特征的全景图拼接方法具体流程如下：

- 1、对待拼接原始图像进行图像预处理和几何校正；
- 2、提取图像 SIFT 特征算子，综合考虑 SIFT 特征提取速度和精度，选择产生 3 个图像组构建图像高斯金字塔；
- 3、用BBF搜索算法进行特征匹配，得到候选匹配特征对；
- 4、采用一种改进的RANSAC算法进行匹配提纯，得到鲁棒的特征匹配对；
- 5、本文引入概率模型对多幅图像的两两匹配关系进行计算，确定参考图像，排除干扰图像；
- 6、最后对配准的图像进行图像融合处理形成全景图，由于对颜色和亮度等都经过了平滑处理，拼接结果没有明显的缝隙。

6.2 展望

图像拼接技术的发展是一个不断完善的过程，本文在理论和实际工作当中都做了一些力所能及的工作，但是仍只涉及到其中的一小部分，此外，由于理论知识上的欠缺和技术上的难点还有大量问题函待研究和解决：

- 1、拼接图像中出现“鬼影”现象。在图像重叠部分所占比例过高、多个相似图像出现误匹配以及图像场景中有多个运动目标干扰时都可能造成“鬼影”现象。在本文中，由于校园里行人的运动和重叠区域较大，导致在其临近领域存在明显的“鬼影”现象。
- 2、图像变形。匹配的图像之间存在 H 矩阵，拼接后导致图像有伸缩扭曲的情况，全景图像的拼接更有可能因为若干个图像变形的叠加，导致图像拼接失败，需要寻找一种更好的变换方法改善图像变形现象。
- 3、图像内容的完整性和拼接的平滑过渡之间存在矛盾。简单的图像融合方法本质上是以牺牲输入图像的灰度信息来换取拼接缝隙的改善。本文选用的渐入渐出融合算法较为简单，复杂度较低，在一些曝光差异较大的地方还存在明显的缝合痕迹。
- 4、全景图像拉直问题。采用图像拼接技术实现的全景图会出现弧形的效果，尤其是柱面全景图，需要考虑采用数学手段将图像拉直到统一平面上。

参考文献

- [1] 钟力, 胡晓峰. 重叠图像拼接算法[J]. 中国图像图形学报, 1998, 3(5):367-369.
- [2] Yehuda Afek, Brand A. A two phase digital photo mosaic system[C].*International Conference on Imaging Science, Systems, and Technology*, 1999, 151-154.
- [3] Richard Szeliski. Video mosaics for virtual environments[C].*IEEE Computer Graphics and Applications*, 1996, 16(2):22-30.
- [4] S.Peleg, B.Rousso and A.Acha. Mosaicing on Adaptive Manifolds[J]. *IEEE Transactions on pattern and machine intelligence*, 2000, 22(10):1144-1154.
- [5] S.Gumustekin, R.Hall. Mosaic image generation on a flattened gaussian sphere[C]. *In Proceeding of IEEE Workshop on Applications of Computer Vision*, 1996, 50-55.
- [6] S.Gumustekin, R.Hall. Image registration and mosaicing using self-calibrating camera[C].*International Conference on Image Processing*, 1998, 818-822.
- [7] M.Brown, DG Lowe. Recognising Panoramas[C].*Proceedings of IEEE International Conference on Computer Vision*, 2003, 1218-1225.
- [8] 件建宁, 郭宝龙. 一种基于兴趣点匹配的图像拼接方法[J]. 计算机应用, 2006, 26(3):610-612.
- [9] 李晓娟, 郭宝龙. 一种基于兴趣点方向特征的图像拼接方法[J]. 计算机工程与应用, 2007, 43(12):26-28.
- [10] 孙立峰, 张茂军. 虚拟空间会议系统研究与实现方法[J]. 小型微计算机系统, 2000, 21(5):462-465.
- [11] 张茂军, 孙立峰. 虚拟会议空间的研究与实现[J]. 计算机工程, 2001, 27(1):10-12.
- [12] C.Kuglin, D.Hines. The phase correlation image alignment method[C]. *In Proc IEEE Conference on Cybernetics and Society*, 1975, 163-165.
- [13] E.Castro, C.Morandi. Registration of translated and rotated images using finite Fourier transforms[J], *IEEE Transactions on pattern and machine intelligence*, 1987, 9(5):700-703.
- [14] B.Reddy, B.Chatterji. An FFT-based technique for translation, rotation, and scale-invariant image registration[J]. *IEEE Transactions on Image Processing*, 1996, 8(5):1266-1271.
- [15] R.Gonzalez, R.Woods. 数字图像处理(MATLAB版) [M]. 北京:电子工业出版社, 2004.
- [16] L.McMillan, G.Bishop. Plenoptic modeling: An Image-based Rendering System[A]. *Computer Graphics Proceedings (SIGGRAPH)*, 1995, 39-46.

- [17]O.Faugcras.S, Laveau.3-D Reconstruction of Urban Scenes from Sequences of images.*Computer Vision and Image Understanding*, 1998, (3):292-309.
- [18]P.S.Heckbert. Survey of Texture Mapping.*IEEE Computer Graphics and Applications*,1986,(11):56-62.
- [19]李晓娟.图像拼接技术研究[D], 西安电子科技大学, 2007年
- [20]李志刚.边界重叠图像的一种快速拼接算法[J]. 计算机工程, 2000, 26(5):37-38.
- [21]L.Brown. A survey of Image Registration Techniques[J]. *ACM Computer Surveys*,1992, 24(4):325-376.
- [22]H.Moravec. Visual Mapping by a Robot Rover[C]. *International Joint Conference on Artificial Intelligence*, 1979, 598-600.
- [23]S.Smith, M.Brady. SUSAN-A New Approach to Low Level Image Processing[J]. *International Journal of Computer Vision*. 1997, 23(1):45-78.
- [24]W.Forstner. A feature based Correspondence Algorithm for Image Matching[J]. *International Archives of Photogrammetry and Remote Sensing*, 1986, 26:150-166.
- [25]刘宝生, 闫莉萍, 周东华. 几种经典相似性度量的比较研究[J]. 计算机应用研究, 2006, 11(23):1-3.
- [26]S.Chen. Measures of Similarity between Vague Sets[J]. *Fuzzy Sets and Systems*, 1995, 74(2):217-223.
- [27]Szeliski R, Heung-Yeung Shum, Creating Full View Panoramic Image Mosaics and Environment Maps. *SIGGRAPH 97 Conference Proceedings*, 1997,3(1):251-258
- [28]B. Reddy, B. Chaterji. A FFT Based Technique for Transition, Rotation, and Scale Invariant Image Registration[J]. *IEEE Trans. Pattern Analysis and Machine Intelligence*,1996,5(8):1266-1271.
- [29]K.Mikolajczyk,T.Tuytelaars,C.Schmid,A.Zisserman,J.Matas,F.Schaffalitzky,T.Kadir,L.V.Gool,A comparison of affine region detectors,*International Journal of Computer Vision*, 2005.
- [30]D. Lowe. Object Recognition from Local Scale-Invariant Features[C]. *International Conference on Computer Vision*, 1999, 1150-1157.
- [31]D.Lowe. Distictive Image Features from Scale-Invariant Keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2):91-110.
- [32]郭雷, 李晖晖, 鲍永生. 图像融合[M]. 北京:电子工业出版社, 2008.
- [33]A.Pinz, M.Prantl , H.Ganster and H.Koppborotschnig. Active Fusion: A New Method Applied to Remote Sensing Image Interpretation[J]. *Pattern Recognition Letters*, 1996, 17(13): 1349-1359.
- [34]C.Genderen, J.Pohl, L.Van. Multisensor Image Fusion in Remote Sensing:Concepts,Methods and Applications[J]. *International Journal of Remote*

- Sensing*, 1998, 19(5):823-854.
- [35] T.Lindeberg. Scale-Space for Discrete Signals[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1980, 20:7-18.
- [36] J.Koenderink. The Structure of Images[J]. *Biological Cybernetics*, 1984, 363-396.
- [37] J.H.Friedman, J.L.Bentley, R.A.Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM.ansactions on Mathematical Software*.1977,3(3): 209-226.
- [38] Jeffrey S.Beis, D.G.Lowe. Shape indexing using approximate nearest neighbour search in high-dimensional spaces[C] . *In Conference on Computer Vision and Pattern Recognition*, Puerto Rico, 1997, 1000-1006.
- [39] C.Fish. A Random Sample Consensus: A Paradigm Formodel Fitting with Application to Image Analysis and Automated Cartography[J]. *Communication AssociationMachine*, 1981, 24(6): 381-395.
- [40] Szeliski R. Image mosaicing for tele-reality applications. *IEEE Computer Society on Applications of Computer Vision*, Sarasota, Florida, 1994, 44-53.
- [41] J.Koenderink. The Structure of Images[J]. *Biological Cybernetics*, 1984, 363-396.
- [42] 刘婧瑜. 多传感器图像融合方法研究及实现[D]. 哈尔滨工业大学, 2006.
- [43] 李晖晖. 多传感器图像融合算法研究[D]. 西北工业大学, 2006.
- [44] 林诚凯, 李惠, 潘金贵. 一种全景图生成的改进算法[J]. *计算机工程与应用*, 2004, 35:69-71.
- [45] 仵建宁, 郭宝龙. 图像拼接技术研究[D]. 西安电子科技大学, 2006.

致谢

本论文是在吴谨教授的精心指导下完成的，值此论文完成之际，谨向吴老师表示最诚挚的感谢。从课题的研究方向、研究内容、研究思路等方面的指引到最后论文的修改都倾注了导师的心血。在三年的学习过程中，吴老师渊博的知识、敏捷的思维、谦逊的为人使我在学习知识和做人方面都受益匪浅。导师对我的谆谆教诲必将永远鞭策着我在人生道路上不断进取。再一次衷心地感谢我的导师吴谨教授。

同时要感谢三年来朝夕相处的肖浩华、胡婷婷、陈奕奕、陈宇、姚欣、文志科、明幼林、康晓晶等同学，正是他们在学习中对我的帮助和影响，使我不断进步。共同的求学生活中结下的深厚情谊，令人难以忘怀。

感谢我的父母和家人，他们给予我这么优越的学习条件，他们的爱和鼓励使我能够积极面对学习和生活中的每一个困难。

感谢实验室的其他老师和同学在论文完成期间给予作者的关怀、鼓励和帮助。

由于本人能力有限，论文中难免出现疏漏之处，敬请各位老师和同学们批评指正。

最后感谢各位评审专家在百忙之中抽出宝贵的时间对我的论文进行审阅，衷心谢谢各位评审赐教和指正。

研究生期间发表的论文

- [1] 郑辉,吴谨. 基于小波分频与直方图均衡的图像增强算法. 现代电子技术[J]. 2010, 8.

基于SIFT特征的全景图像拼接算法研究

作者：[郑辉](#)
学位授予单位：[武汉科技大学](#)

本文链接：http://d.g.wanfangdata.com.cn/Thesis_Y1739528.aspx