



# 基于特征点检测的图像配准算法研究

## Research on Image Registration Algorithm Based on Feature Point Detection

(申请中国空间技术研究院工学硕士学位论文)

培 养单位：中国空间技术研究院西安分院

学 科： 信息与通信工程

研究方向： 航天器数据传输与处理

研 究 生： 张晨光

导 师： 周 詮 (研究员)

2019 年 4 月 1 日



## 摘 要

图像配准是数字图像处理领域中一项非常重要的技术，是模式识别、图像融合、图像拼接等领域中关键的预处理步骤，在遥感、医疗、军事中都有着广泛的应用。图像配准的目的是拟合出两幅图像在空间上的变换关系，根据变换关系将待配准图像与参考图像对齐。配准后的图像与参考图像相似度越高，则配准效果越好。在图像配准算法中基于特征点的算法是目前关注的重点。本文围绕基于特征点的 SIFT 算法进行深入研究，主要包括：

（1）针对 SIFT 算法实时性较差的问题，提出了一种快速图像配准算法。根据 SIFT 算法中构建的图像高斯金字塔的特点，采用下采样对金字塔的规模进行缩减，从而减少了计算量，缩短了算法的运行时间。实验表明，相比于 SIFT 算法，该算法的运行时间可减少 64%。

（2）针对 SIFT 算法没有考虑图像颜色特征和视觉显著性的缺点，本文将 SIFT 算法与显著性区域检测结合，提出了一种基于显著性检测的图像配准方法。该算法首先利用彩色图像的亮度和颜色信息进行显著性区域检测，得到灰度图像形式的显著性图像，然后利用 SIFT 算法进行特征匹配。实验结果表明，本文算法的均方根误差从 18.38 减小到 0.56，准确度明显提高。另一方面，对显著性区域进行图像分割后，再进行特征点检测，能够明显地去除图像背景中的噪声，改善配准准确度。

（3）将本文提出的图像配准算法应用于基于图像集的压缩系统中，有利于压缩性能的提高。

**关键词：**图像配准；特征点；SIFT；显著性区域检测；差值图像



## Abstract

Image registration is a very crucial technology in the field of digital image processing. It is a key pre-processing step in the fields of pattern recognition, image fusion, image stitching, etc. It has been widely used in remote sensing, medical and military. The purpose of image registration is to fit the model of spatial transformation of the two images, and align the image which is going to be registered with the reference image according to the transformation relationship, and a registered image is obtained. The higher the similarity between the registered image and the reference image, the better the registration effect are. Feature points based algorithms in image registration algorithms are currently the focus of attention. This paper focuses on the SIFT algorithm, which is based on feature points. The main contents of this essay contain:

(1) The theoretical ideas and technical points of image registration are studied and analyzed. According to the classification of image registration algorithms, different kinds of algorithms are expounded and compared. Aiming at the widely used feature point-based registration algorithm, several mainstream algorithms were selected for experimental comparison, and the reference direction was determined for future research work.

(2) Aiming at the problem of poor real-time performance of SIFT algorithm, a fast image registration algorithm is proposed. According to the characteristics of the image Gaussian pyramid constructed in the SIFT algorithm, the scale of the pyramid can be decreased by downsampling, therefore we can reduce the amount of calculation and shorten the running time of the algorithm. Experiments show that the algorithm can reduce the running time by 64% compared to the SIFT algorithm.

(3) Apply the image registration algorithm proposed in this paper to the image set based compression system, and leads to the improvement of compression performance.

**Key words:** image registration; feature points; SIFT; saliency detection; hash algorithm; difference image



# 目 录

摘 要 .....	I
ABSTRACT .....	III
目 录 .....	V
主要符号对照表 .....	VII
<b>第 1 章 绪论 .....</b>	<b>1</b>
1.1. 课题研究背景及意义.....	1
1.2. 图像配准技术概述.....	2
1.3. 国内外研究现状.....	3
1.4. 本文的组织结构.....	8
<b>第 2 章 对基于特征点的图像配准算法分析.....</b>	<b>11</b>
2.1. 引言 .....	11
2.2. 特征点配准算法流程.....	11
2.3. 常用特征点配准方法介绍.....	13
2.4. 实验结果与分析.....	17
2.4.1 实验内容 .....	17
2.4.2 实验指标 .....	17
2.4.3 实验结果 .....	18
2.5. 本章小结 .....	21
<b>第 3 章 基于 SIFT 算法的快速图像配准算法.....</b>	<b>23</b>
3.1. 引言 .....	23
3.2. 所提算法框架 .....	23
3.3. 金字塔的构建和极值点的寻找.....	24
3.4. 对于单应矩阵的修正.....	26
3.5. 实验结果与分析.....	31
3.6. 本章小结 .....	34
<b>第 4 章 基于显著性区域的高准确度图像配准算法.....</b>	<b>35</b>
4.1. 引言 .....	35
4.2. 所提算法框架 .....	35

4.2.1. 基于显著性图像上的特征点检测 .....	35
4.2.2. 基于显著性图像的图像分割 .....	36
4.3. 基于 LAB 颜色空间的显著性区域检测 .....	37
4.4. 基于聚类算法的图像分割 .....	38
4.5. 实验结果与仿真 .....	41
4.6. 本章小结 .....	45
<b>第 5 章 图像配准在图像压缩中的应用 .....</b>	<b>47</b>
5.1. 引言 .....	47
5.2. 图像集压缩中的图像配准算法框架 .....	47
5.3. 基于哈希算法选择可压缩的图像对 .....	48
5.4. 基于图像间冗余信息的图像压缩 .....	51
5.5. 实验结果与分析 .....	55
5.6. 本章小结 .....	58
<b>第 6 章 总结与展望 .....</b>	<b>59</b>
6.1. 工作总结 .....	59
6.2. 工作展望 .....	60
<b>参考文献 .....</b>	<b>61</b>
<b>致 谢 .....</b>	<b>65</b>
<b>硕士学位论文原创性声明 .....</b>	<b>67</b>
<b>硕士学位论文使用授权书 .....</b>	<b>67</b>
<b>个人简历、在学期间发表的学术论文与研究成果 .....</b>	<b>69</b>



## 主要符号对照表

缩写名词	英文全称	中文译名
AKAZE	Accelerated-KAZE	加速 KAZE
AOS	Additive Operator Splitting	加性算子分裂算法
BRIEF	Binary Robust Independent Elementary Features	二进制鲁棒性独立基本特征
BRISK	Binary Robust Invariant Scalable Keypoints	二进制尺度不变特征点
BBF	Best Bin First	最优节点优先
CSIFT	Colored SIFT	彩色尺度不变特征转换
EBRs	Edge-Based Regions	基于边缘区域
$E$	Essential matrix	本质矩阵
FAST	Features from Accelerated Segment Test	快速分割测试中的特征点
FED	Fast Explicit Diffusion	快速显式扩散
$F$	Fundamental matrix	基础矩阵
$H$	Homography matrix	单应矩阵
IBRs	Intensity extremal Based Regions	基于密度极值区域
LSH	Locality Sensitive Hashing	局部敏感哈希
LDB	Local Difference Binary	局部差异二进制
MAD	Mean Absolute Differences	平均绝对差算法
MSERs	Maximally Stable Extremal Regions	最大稳定极值区域
ORB	Oriented FAST and Rotated BRIEF	方向性的 FAST 和旋转性的 BRIEF
PCA	Principal Component Analysis	主成分分析
$PSNR$	Peak Signal to Noise Ratio	峰值信噪比
RANSAC	Random Sample And Consensus	随机抽样一致
$RMSE$	Root Mean Square Error	均方根误差
ROI	Region Of Interest	感兴趣区域
SAD	Sum of Absolute Differences	绝对误差和算法
SAR	Synthetic Aperture Radar	合成孔径雷达
SIFT	Scale-Invariant Feature Transform	尺度不变特征转换
SSDA	Sequential Similarity Detection Algorithm	序贯相似性检测算法
SURF	Speeded Up Robust Features	快速鲁棒性特征点



## 第 1 章 绪论

### 1.1. 课题研究背景及意义

在我们的话语体系中有很多俗语与视觉有关：百闻不如一见、眼见为实、眼睛是心灵的窗户，这些俗语都表现了视觉的重要性。我们看书看报，看电视、电影，都离不开视觉的参与，事实上，人主要通过视觉获得外界色彩、形状、距离等直接关系到生存和正常生活的信息，通过视觉获得的信息量远远超过了听觉、触觉等感官。为了揭示视觉的奥秘，生理学做了许多研究，比如 T.N.Wiesel 凭借在视觉系统的信息加工中的研究获得了 1981 年的诺贝尔奖<sup>[1]</sup>。现在我们知道，外界发出或者反射的光经过晶状体等投影在视网膜上，然后经过神经系统传递到大脑，大脑就得到了对图像的感知。在这个过程中，图像由光信号转换为了电信号。

生理学中图像表示形式由光信号转换为电信号，而在计算机数字图像处理中我们使用一个二维函数来表示数字图像，离散的坐标值作为自变量，离散的灰度值作为因变量。对数字图像的处理起源于上个世纪 20 年代，标志是通过海底电缆从伦敦到纽约传输了第一张数字照片<sup>[2]</sup>。在传输图像中，在发送端对图像进行数字编码，接收端再进行解码，所以数字图像处理也可以看作是一种信源编码。数字图像处理的核心是通过计算机进行计算，目的是得出图像的一些属性值，或者使得图像满足人眼的直观感受或使得图像适于进一步的图像处理。传统的数字图像处理包括图像去噪<sup>[3]</sup>、图像增强、图像分割<sup>[4]</sup>、图像压缩等。近来大热的计算机视觉的终极目标是实现图像理解<sup>[5]</sup>，也可以认为是一种高级的数字图像处理。

在数字图像处理之中，图像配准是一项关键的技术。因为随着大数据时代的到来，人工智能、机器学习的火热发展，我们面对的是海量的数据，其中很大一部分是智能手机、监控摄像头等传感器获得的图像，我们可能需要将数目庞大的不同图像结合起来处理，找出其共同特征或者找出差异，比如进行图像匹配、图像融合。然而不同的传感器获得的图像效果可能千差万别，从分辨率，到成像角度，到图像大小都有差异，所以在进行具体的图像处理之前往往需要进行图像配准<sup>[6]</sup>。图像配准的输入是两幅图像，二者之间因为成像角度、成像时间等存在相对形变，图像配准的目的就是拟合出二者之间的变换关系，将其中一幅变形图像（待配准图像）变换到与另外一幅图像（参考图像）相同的状态。

图像配准在遥感、模式识别、计算机视觉等领域被广泛地应用。在增强现实技术中，借助图像配准实现视觉融合，增强对环境的感知<sup>[7]</sup>；在医学分析中，依靠单一医学图像很难准确识别病灶，需要不同的医学图像融合获取更多的信息，而图像融合的

关键就是利用图像配准将不同属性的医学图像对齐，哈工大的付浩威针对医学图像配准中的非线性不定问题进行了研究<sup>[8]</sup>；在遥感分析中，受到气流、云层、光照等影响，卫星或者飞机上的相机在每次拍摄时不能保证具有相同的成像条件，需要利用多幅图像进行对比或者拼接，同样需要图像配准技术；在模式识别领域，如车牌识别<sup>[9]</sup>、人脸识别<sup>[9]</sup>、运动目标跟踪和三维重建，虽然基于监督训练的方法已经被广泛应用，但基于传统图像配准的方法也仍然占据一席之地。

## 1.2. 图像配准技术概述

在数学上，图像配准指的是两个二维函数之间在空间和灰度上的反变换。图像配准用数学公式表示如下，其中二维函数  $f_1(x, y)$  和  $f_2(x, y)$  分别指参考图像和待配准的变形图像。

$$f_2(x, y) = g[f_1(h(x, y))] \quad (1-1)$$

式 1-1 中  $h$  表示二维空间坐标变换， $g$  表示灰度变换。广义配准的目标是要找出坐标和灰度的最佳变换参数，但通常意义上的配准特指空间位置坐标的变换，灰度变换可以认为是图像配准之前的预处理部分。本文所研究的图像配准正是狭义的配准，只进行空间上的配准。

1992 年，L.G.Brown 在其论文<sup>[11]</sup>中对图像配准技术进行了系统的总结，他认为配准技术分主要有四个内容：特征空间、搜索空间、搜索策略和相似性度量。图像配准和图像匹配的主要研究内容基本相似，都需要在两幅图像之间寻找具有对应性的部分，不同的是图像配准需要在匹配的基础上估计出二者之间的变换关系。所以二者在关键技术上有相同点，故在下文对技术的介绍中对二者不进行区分。

### 特征空间

特征空间对应于输入空间，又区别于输入空间。输入图像本身可以看作是输入空间，但计算机更适合处理的是矩阵或者向量，对图像进行特征提取就得到了特征空间。特征空间是对图像抽象化和数字化的表示，相比于输入空间，特征空间对图像形变的容忍性更高，数据量更小。一般选取的特征可以是点、线、区域这种浅层次的特征，也可以是更深层次的高阶特征如统计信息（信息熵、矩等）。在深度学习中，一般使用卷积神经网络进行特征提取，得到特征图作为特征空间。Saining Xie 等设计了神经网络，可以比传统的 Canny 边缘检测算法更精确的边缘特征<sup>[12]</sup>。

### 搜索空间

搜索的目的是匹配，具体来讲是根据特征的描述向量在另外一幅图中搜索得到合适的匹配对。为了进行高效率的搜索，可以对搜索空间进行划分。搜索空间的最终目的是靠匹配对成功拟合出最优的空间变换模型和其参数值。

### 相似性度量

相似性度量有两方面。一方面是两幅图像各自特征空间之间的相似性，不同的特征空间使用的相似性度量一般不同，常用的相似性度量有信息论中的互信息、联合熵，像素点之间的各种意义上的距离等。另外一方面是配准（变换）之后的结果与参考图像的相似程度，是对配准效果的评估，也可以认为是配准的效果函数或者功效函数，从而使用在深度学习中。

### 搜索策略

确定了特征空间中相似性度量就可以开始搜索，在寻找的过程中使用不同的策略会对搜索的效果产生很大的影响。策略不同，其采用的代价函数（损失函数）也不同。最简单的搜索策略就是贪婪搜索和穷举法，实际应用中可以选取效率更高的牛顿法、梯度下降法等。

## 1.3. 国内外研究现状

图像配准在上个世纪七十年代由美国提出并运用于导弹制导，将目标图像和实时图像配准，提高了命中目标的精度。图像配准在之后得到了国内外的关注，并在各个领域得到了广泛应用。研究重点集中在医学图像、遥感图像、现实增强和模式识别等领域。下面根据图像配准的分类，介绍各自的研究历程。

L.G.Brown 的总结中将图像配准分成两类：基于灰度的图像配准和基于特征的图像配准。本文参考了谢凤英等主编的资料<sup>[13]</sup>，将图像配准分成三类：基于灰度信息、基于变换域和基于特征的图像配准方法。

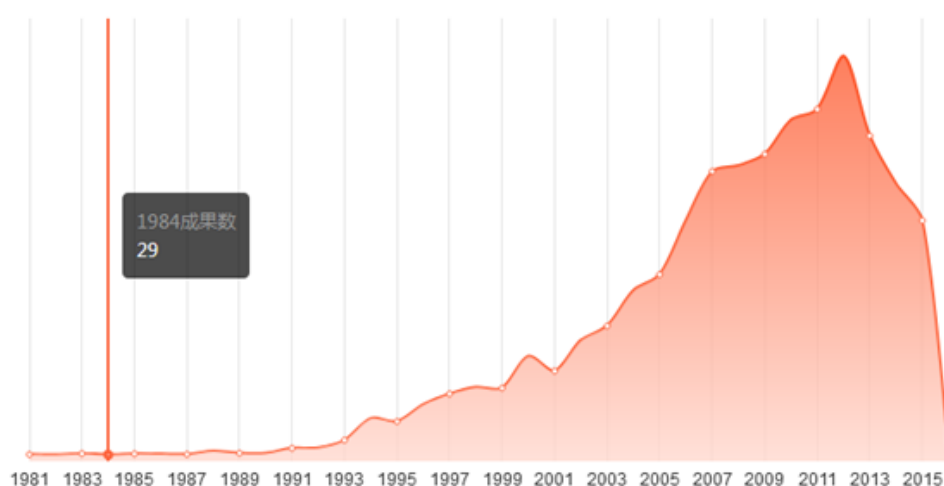


图 1-1 百度学术统计的图像配准(Image Registration)在各个年份的成果数

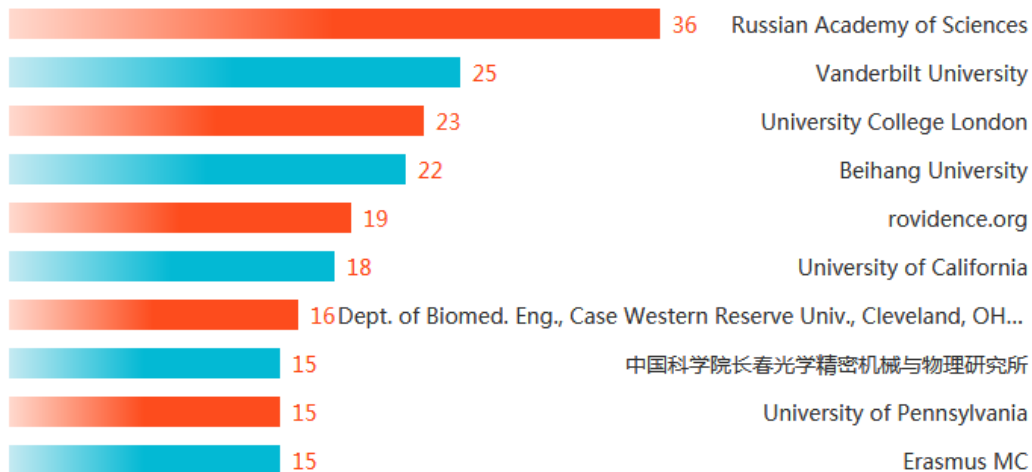


图 1-2 百度学术统计的各个机构的研究成果数

基于灰度信息的方法利用两幅图像之间灰度信息的相似性进行搜索。不同的匹配算法之间区别主要是相关性度量的不同，平均绝对差算法（Mean Absolute Differences, MAD）以模板与图像窗口的平均绝对差作为相似性度量。该算法是模式识别中常用的匹配算法，计算简单但计算量大，且容易受噪声影响。但是 MAD 仍然可以作为其他算法的补充，燕磊等<sup>[14]</sup>在 2017 年就将 Oriented FAST 和 MAD 结合起来用于图像配准。绝对误差和算法（Sum of Absolute Differences, SAD）在 MAD 算法基础上将相似性测度改成了模板与子图之间的 L1 距离，并且不再求均值。郭龙源博士等将 SAD 作为费用函数，在三个通道分别处理，对彩色图像的匹配进行了改进<sup>[15]</sup>。序贯相似性检测算法（Sequential Similarity Detection Algorithm, SSDA），由 Barnea 和 Silverman 于 1972 年提出，可以比 MAD 算法可以快几十到几百倍。张维琪等利用 OpenMP 编程实现了多核多线程的自适应阈值 SSDA 的图像配准并行算法<sup>[16]</sup>。1948 年，Shannon 首次提出互信息的概念，Viola<sup>[17]</sup>等人在 1995 年将互信息引入图像配准，解决异源图像的配准。曹源等以互信息作为相似性测度，对粒子群算法进行改进，与烟花算法结合起来进行图像配准<sup>[18]</sup>。基于灰度信息的方法因为灰度信息直接来源于图像的像素本身，所以优点是不需要对图像进行复杂的特征提取等预处理，缺点是运算量大；没有利用图像的颜色信息；对非线性变换不适用；对噪声较敏感。

基于变换域的算法在变换域内完成图像的配准。图像由空域到变换域的变换方法主要有傅里叶变换，小波变换等。傅里叶变换的优势是可以利用 FFT 快速实现，并且在变换过程中可以去除与频域不相关的噪声的干扰。在 20 世纪 80 年代前后，在傅里叶变换的基础上出现了小波变换。相比于傅里叶变换，小波变换不仅可以分析出信号的频率分量大小，还可以分析得到频率的位置信息，不仅适用于平稳信号，还可以用于非平稳信号（如遥感图像）。利用小波分解可以产生各种分辨率的图像，我们可以综合利用这些图像进行图像配准方法的研究。湖南大学伍君提出了基于小波变换提

取边缘特征点的图像配准方法和基于小波变换与互信息测度相结合的图像配准方法 [19]。

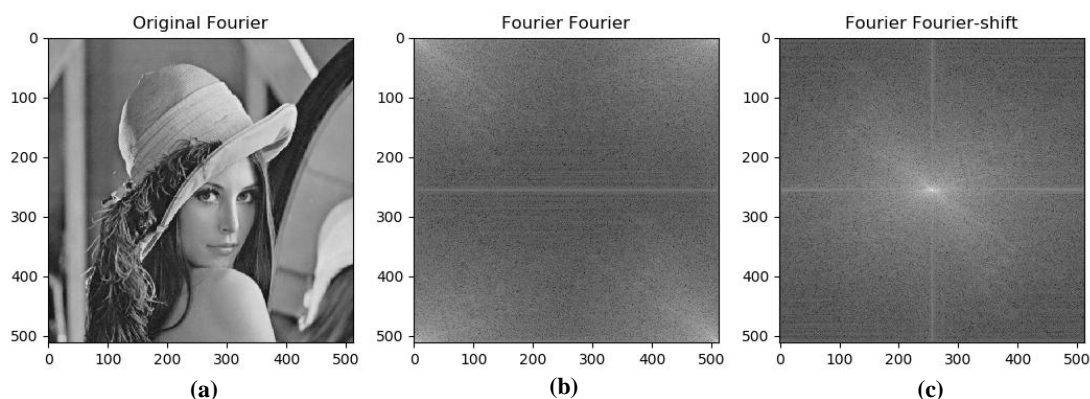


图 1-3 图像的傅里叶频谱表示

基于特征的图像配准方法利用特征提取，将对整个图像的分析转化为对图像特征的分析，降低了分析的维度，因此能够大大减小图像处理过程的运算量。特征是在高层次对图像信息的提炼，所以对灰度变化、空间变换等有很好的鲁棒性。特征一般要满足两个条件：提取的特征具有代表性；提取的特征要能满足仿射、尺度不变性。提取的特征又可分为全局特征和局部特征。全局特征以矩特征为代表，如 Hu M K 提出的 Hu 不变矩特征<sup>[20]</sup>、基于 Zernike 多项式的 Zernike 矩特征<sup>[21]</sup>等。

基于全局特征的配准算法计算量大，对图像质量和类别要求较高：要求图像遮挡部分少，相同部分多。基于局部特征的算法更为常见。局部特征包括局部区域特征和局部点特征。目前较常用的局部区域检测算法包括：基于边缘区域（Edge-Based Regions, EBRs）检测算法、最大稳定极值区域（Maximally Stable Extremal Regions, MSERs）检测算法、基于密度极值区域（Intensity extremal Based Regions, IBRs）检测算法、显著性区域（Salient Region）检测算法等。较为经典的边缘算法包括 Hough 特征提取<sup>[22]</sup>、Canny 算子边缘检测等<sup>[23]</sup>。Troglio G 等人提出的将分水岭分割算法和广义 Hough 变换结合的特征提取算法，使用在了数据量庞大的行星图像配准分析中<sup>[24]</sup>。

局部点特征是另外一种局部特征。1977 年，Moravec<sup>[25]</sup>利用灰度自相关函数，把在多个方向都有亮度变化的像素点判定为角点。Harris 在 1988 年提出了 Harris 角点提取算法<sup>[26]</sup>，观察像素点周围小窗口内的各方向灰度变化的强度。在具体计算中，可以用协方差矩阵及其特征值表示所有方向上的强度变化。

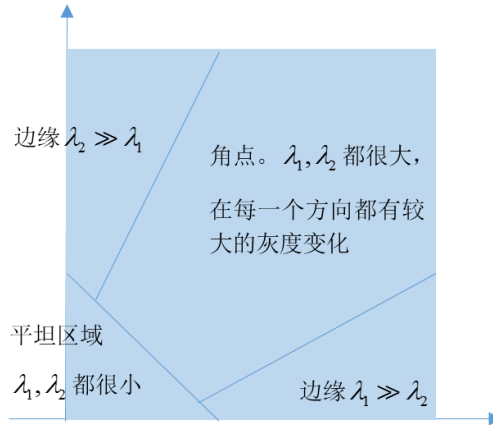


图 1-4 Harris 角点检测，由协方差矩阵的特征值判断是否是角点

Harris 角点解决了对旋转和灰度的不变性，但是没有尺度不变性。20 世纪 90 年代，Lindeberg<sup>[27]</sup>提出了图像的尺度空间理论。Lowe 将尺度空间引入了局部特征点的检测中。Lowe 从 1999 年到 2004 年，Lowe 相继发表的三篇论文<sup>[28-30]</sup>，系统地提出并完善了 SIFT（Scale Invariant Feature Transform）算法。Lowe 利用构造高斯拉普拉斯金字塔的方式，来快速求解高斯拉普拉斯空间中的极值点，加快了特征提取速度。在特征点周围 16\*16 的区域，通过梯度直方图统计的方式得到了特征点的主方向，并形成对特征点进行描述的 128 维特征向量。SIFT 算法在图像旋转、尺度变换、和视角变换等条件下都具有良好的不变性，在目标识别、图像拼接、目标跟踪等领域扮演着重要角色。

针对有较多斑点噪声的合成孔径雷达（Synthetic Aperture Radar, SAR）图像，SIFT 表现较差，这是因为 SAR 图像中的噪声属于相关乘性噪声，服从瑞利分布，而 SIFT 针对的是光学图像，噪声是加性噪声，服从高斯分布。BFSIFT<sup>[31]</sup>是由中科院的王山虎基于 SIFT 提出的一种改进算法，通过双边滤波，将标准的高斯尺度空间用各向异性尺度空间来代替，可以抑制 SAR 图像中平坦区域上的斑点噪声。在匹配策略上，BFSIFT 采用了双向匹配，减少了误匹配对。但是双边滤波的耗时影响了算法的实时性。

2006 年，Bay 在 SIFT 算法的基础上，提出了一个实时性较强的算法，即 SURF（Speeded Up Robust Features）<sup>[32]</sup>。Bay 通过引入积分图像和模板近似，并在特征描述阶段采用了一阶 Haar 小波响应，计算速度比 SIFT 快 3 倍左右。除 SIFT 和 SURF 算法外，比较重要的类 SIFT 算法包括 PCA-SIFT<sup>[33]</sup>、ASIFT（Affine-SIFT）<sup>[34]</sup>、ORB<sup>[36]</sup>、CSIFT(Colored SIFT)<sup>[37]</sup>等其他改进的 SIFT 算法<sup>[38][39]</sup>。PCA-SIFT 利用了主分量分析技术（Principal Component Analysis, PCA），将高维空间的样本投影到低维的特征空间，寻找向量中起主要作用的分量，从而降低了特征描述子的维度，提高了特征的匹配速度。但它的缺点是需要事先对大量典型图像进行学习，不适用于异源图



像。ASIFT 是基于 SIFT 算法在 2009 年提出的一种新算法，着重解决了 SIFT 等特征提取算法在图像仿射变换角度较大时配准较差的问题。

传统的 SIFT、SURF 等特征检测算法都是基于线性的高斯金字塔进行多尺度分解来消除噪声和提取特征点。但高斯分解牺牲了局部精度，容易造成边界模糊和细节丢失。Alcantarilla<sup>[40]</sup>等人在 2012 年发表的论文中提出了基于非线性尺度空间的 KAZE 特征。在论文的各项测试中，KAZE 算法比 SIFT、SURF 等算法具有更好的稳定性、旋转不变性、尺度不变性。唯一的不足是其计算复杂度过高，比 SIFT 算法运算时间更长。

国内的研究紧跟国际接轨，并且做出了一些贡献。西北工业大学李晖晖等针对参考图像和待配准图像对之间存在的仿射变换问题，提出了一种基于 SIFT 特征和图像正确匹配特征点对角度相对距离的图像配准算法<sup>[41]</sup>，实现了特征点之间的精确匹配。潘建平等引用相对距离理论，在 SURF 算法的基础上改进了图像配准算法<sup>[42]</sup>。侯坚、张明利用蚁群算法对 SIFT 进行优化<sup>[43]</sup>，首先采用内核投影算法对 SIFT 特征描述子进行降维，然后采用优化的蚁群算法针对初匹配点进行提纯，提高了匹配率。北航的安勇、郑虹提出了基于 K 近邻点的 SIFT 特征点匹配<sup>[44]</sup>。电子科技大学的肖哲等针对大脑 CT 图像提出了结合 SIFT 特征和灰度特征的描述子，并且通过欧式距离和余弦相似度得到匹配图<sup>[45]</sup>。Magic Leap 公司基于点特征建立了训练数据，在神经网络下实现了对单应矩阵的估计<sup>[46]</sup>。

基于特征点的配准技术研究已成为图像的配准研究领域中最热门和最突出的研究方向。基于特征点的配准技术具有以下三个优点：

- (1) 在图像中提取出来的特征点与原图像的像素点相比，数目得到了极大的减少，运算量也得到了极大的减小。
- (2) 特征点的鲁棒性更强，对光照、噪声和仿射变换具有适应性。
- (3) 单应矩阵有 8 个自由度，原则上根据四对匹配对就可以计算出透视变换矩阵。通过特征点可以直接进行拟合，减少了计算量。

针对图像配准，尤其是基于特征点的图像配准的研究取得了一些重要的成果，但仍然存在一些亟待解决的问题：

- (1) 首先是配准的精度。配准后的图像与参考图像应该尽可能地相似，从像素级别的精度发展到亚像素级别的精度，还需变换模型的准确性和依赖插值技术的进步。
- (2) 配准的速度。移动平台如智能手机等的普及，受限于计算性能和使用场景，需要实时性高的配准，这就需要算法具有快速性的特点。
- (3) 更加智能的配准算法。现有的多种算法都是对特定图像的优化，难以有一个普适性的算法适合多种类型的图像，并且对于明显完全不同的图像，应该及时做出判断，避免不必要的计算。

(4)特征点的数目和质量的问题。理论上只需要四对匹配点对就可以完成变换模型参数的估计。但是由于特征点本身的精度有误差,而且在匹配过程中可能存在误匹配,所以需要大量的特征点来进行拟合。但是过多的特征点会造成搜索空间的增大,增加计算量。所以质量好的特征点可以一定程度降低特征点的数量,如何提高质量,降低数量会是一个研究方向。

(5)三维重建的配准。现在已有的算法针对的都是二维平面图像的配准,但是随着VR等技术的发展,在三维立体视觉中的图像配准、图像匹配将变得具有必要性。

#### 1.4. 本文的组织结构

图像配准作为图像处理中关键的预处理步骤,一直是国内外的研究者的课题之一。本文的主要研究内容是基于特征点的图像配准算法,选取其中的 **SIFT** 算法作为改进基础,针对其缺点进行优化。最后将本文提出的更具准确度的配准算法使用在图像压缩中,可以得到更大的压缩比。

研究目的是对该类方法有一个综合性的评测,对算法实时性差的特点进行改进;对一些没有充分利用图像信息,配准效果稍差的算法进行优化。

全文共分六章,各章节将按照如下安排:

第一章为绪论,阐述了视觉、图像处理、图像配准的意义。从数学角度定义了图像配准,分别对三类图像配准算法的研究现状进行了介绍。并对各个章节的内容进行了概述。

第二章的重点在分析基于特征点的图像配准方法。首先对基于特征点的图像配准算法的流程做了分析,介绍了几种常用的特征点检测算法和描述方法。设计实验对主要的几种配准算法进行了对比和分析。

第三章是本文的主要研究内容之一,针对 **SIFT** 算法运算量较大的特点和金字塔的特点,提出了基于下采样的减少计算量的方法。本章根据下采样的对象和插值方法的使用设计了对照实验,并与 **SURF** 算法进行了比较,可以看出改进之后的方法有一定的优越性。

第四章将 **SIFT** 算法与显著性区域检测方法进行了结合。首先介绍了显著性区域检测的意义,介绍了几种主要的算法。然后选取 **FT** 区域性检测方法补充了 **SIFT** 算法对于颜色分量的忽略,实验证明在显著性图像上的特征点检测可以得到更多的高质量点。针对基于显著性检测的图像分割做出了改进,可以保留更多的主体信息。

第五章主要研究的是图像配准算法的应用。本文将图像配准应用在了对图像数据集的压缩中。针对图像集中存在包含相同场景的图像,选取其中一幅为待配准图像,利用变换矩阵描述其余图像,并保存配准后的图像与参考图像的差值,此时得到的差

值图像为稀疏矩阵，可以进一步压缩。利用这种应用可以有效减少数据集的规模，降低存储系统的负担。

第六章是对本文的总结。研究生阶段在图像配准方面做出了一定的工作，但受限于时间和个人能力，还有很多待完善之处，有待继续研究。



## 第2章 对基于特征点的图像配准算法分析

### 2.1. 引言

基于特征点的配准方法因为其鲁棒性更好，计算量较小，可以根据特征点形成的匹配点对直接对变换模型的参数进行估计的特点，受到了国内外众多学者的研究。基于特征点的配准算法的流程大致相同，区别主要是关于特征点的检测和描述方法。为了对这类配准方法有一个综合性的认识，找出研究的侧重点，有必要对现存的主流算法做一些了解。

本章首先介绍了配准算法的大致流程，针对特征点检测和描述的方法介绍了几种算法。对常见的基于特征点检测的配准算法设计了对比实验，确立了 SIFT 算法作为日后的改进基础。

### 2.2. 特征点配准算法流程

SIFT 算法<sup>[30]</sup>给出了一套完整的图像配准方法，从特征点的检测方法到如何使用描述子数字化地表示特征点，并通过两幅图像之间的特征点通过搜寻查找得到匹配点对，通过匹配点对估计出两幅图像之间的变换模型参数，最后根据变换矩阵完成图像配准。在 SIFT 算法之后出现了许多经典的图像配准算法，同样采用了相同或者相似的配准流程。

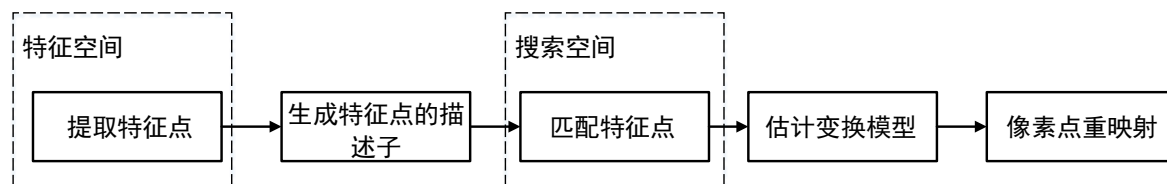


图 2-1 图像配准算法的大致流程

特征点的选取非常重要，因为我们的操作对象不再是原始图像而是从图像提取出的高维特征。这些特征点要能准确地反映图像的特点，即具有显著性；不同的特征点应该包含不同的信息，即独特性；对同一幅图像每次检测出的特征点应该是一致的，这是特征点的可重复性；在图像经过噪声、光照、形变的影响后特征点要保持不变，这是特征点的稳定性。满足以上特性的点就可以称为特征点。特征点的好坏直接决定了算法的效果，好的特征点可以提高配准算法的速度和精度。不同图像配准算法之间

主要的不同点之一就是关于如何提取特征点。

相对于原始图像，特征点所含的信息较少，但与普通的点坐标相比，特征点为了保证尺度不变性和旋转不变性，还会包含尺度因子和主方向等，同时为了具有鲁棒性，还会加入特征点邻域的一些统计信息，这些信息进行编码，就得到了描述子。描述子是特征点的数字化表示，有利于之后两幅图像的特征点的匹配过程。不同图像配准算法之间主要的不同点除了提取特征点的方法之外就是生成描述子的方法。一般通过特征点空间邻域上的信息来生成描述子，其中以二进制字符串表示的描述子具有适合计算机快速计算的特点，常用的生成描述子的方法有 **BREIF**<sup>[47]</sup>。**BRIEF** 在特征点周围按照一定分布函数采样得到一些点对，通过比较每对像素点的灰度值可以直接生成二进制的字符串。以特征点周围两个随机的点  $x, y$  为例， $p$  指对应点处的灰度值，则根据下式可以得到比特 0 或者比特 1。

$$\tau(p; x, y) := \begin{cases} 1: p(x) < p(y) \\ 0: p(x) \geq p(y) \end{cases} \quad (2-1)$$

特征点的匹配是根据描述子的相似程度进行两幅图像之间特征点的匹配，在另外一幅图像中找出对应的特征点。匹配的过程与相似度准则的选取有关，而相似度准则的选取与描述子的形式有关。相似性测度目前常用的方法有两种：距离测度法、相似性函数法。不同的距离测度法选取比较了向量之间不同种类的距离，如欧氏距离、马氏距离、汉明距离、切比雪夫距离、闵可夫斯基距离等。相似性函数法主要是对向量的方向进行分析，如夹角余弦法、相关系数法等。确定了相似性测度，根据如何在另外一幅图中搜索得到最相似的特征点又有不同的搜索策略。优秀的搜索策略可以提高算法效率，缩短运行时间，减少误匹配。常使用的有：穷尽搜索、分层搜索、Powell 算法、遗传算法、蚁群算法。

搜索策略还与匹配策略有关。匹配策略指的是在什么条件下认为两个特征点之间是匹配关系。有三种匹配策略：固定阈值、最近邻、最近邻距离比率（**Nearest Neighbor Distance Ratio, NNDR**）。以距离测度法为例，在固定阈值策略中，当两个特征点之间的距离小于某个阈值就认为二者是匹配的；在最近邻策略中，认为距离最近的点是可以匹配的，这样保证了每一个特征点都会形成匹配对；在最近邻距离比率策略中，不仅考虑最近邻点，还要找到次近邻点，当两个点与特征点的距离的比率大于某个值时，认为可以达成匹配。第三种匹配策略经常会用到 **K** 近邻算法，可以通过 **BBF**（**Best Bin First**）算法来构建 **KD** 树来实现。**BBF** 算法可以高效地高维空间中找到最近邻的近似解<sup>[48]</sup>。

变换模型的估计指的是通过特征点的匹配对，估计出两幅图像之间的变换模型。在我们选取好变换模型的类型后，所估计的其实是变换模型的参数。常见的变换模型包括仿射变换、透视变换、多项式变换等。以透视投影变换模型为例，因为透视变换

矩阵有 8 个自由度，理论上通过 4 对匹配点对就可以联立方程组求解出  $3 \times 3$  大小的透视变换矩阵。但是由于匹配对中的特征点存在误差，或者存在误匹配对，所以需要使用时使用这些样本数据进行大数据拟合。因为本文对全局像素采用了同一个变换模型，所以拟合的目的使得变换模型尽可能满足大部分像素点或者使得平均的误差最小。拟合的方法一般有最小二乘法和随机抽样一致性算法（Random Sample And Consensus, RANSAC）算法<sup>[49]</sup>。

利用拟合出的变换模型（矩阵）将待配准图像变换到参考图像的过程是一个像素点重映射的过程。这个过程不是简单的将已有像素点映射到新坐标下，因为这样可能会造成新图像中某些像素的缺失，所以在数字图像处理中使用的是根据新图像中像素点的坐标反推出原有图像中与其有关的像素点，根据这些像素点计算得到新图像中的像素值。这个过程其实是插值的过程。

### 2.3. 常用特征点配准方法介绍

在 SIFT 算法提出之后出现了很多著名的基于特征点检测的算法，如 SURF、ORB、KAZE 等。因为图像本身由于自身的成像特点性质会千差万别，图像变换也多种多样，配准算法本身的应用场景对算法的要求也不同，众多研究者做出了各自的改进。下面介绍几种公认的优秀算法。

#### SIFT

尺度不变特征转换（Scale-Invariant Feature Transform, SIFT）是一种在尺度空间中寻找极值点作为特征点的算法，检测出的 SIFT 特征点具有尺度不变性和平移、旋转不变性。除了特征点检测，SIFT 还提供了一系列的特征点描述和匹配算法。极值点的检测在差分高斯金字塔中进行，通过尺度空间函数进行三维二次函数曲线拟合寻找真正的极值点。特征点的方向信息通过其邻域内像素点的梯度统计来实现。数字图像中像素点是离散的，梯度通过差分的形式表示。梯度幅值和方向如式 2-2 和 2-3 所示。

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

$$\theta(x, y) = \alpha \tan(2(L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \quad (2-3)$$

对梯度方向进行量化，对同一范围内的幅值进行叠加，取直方图的主峰值对应的梯度方向作为特征点的主方向。SIFT 算法将尺度和方向信息集中数字化表示，得到了特征点的描述子。先将特征点旋转到主方向，然后对特征点的矩形邻域进行分块（ $4 \times 4$ ），在每个子块内计算梯度直方图（8 方向量化），这样就生成了具有独特性的 128 维（ $4 \times 4 \times 8$ ）的向量，再进行光照的归一化就得到了特征点的描述子。

## SURF

为了使得 SIFT 算法满足实时性的要求，一般需要借助 GPU 等硬件的协助。2006 年 Bay 提出了 SURF (Speeded Up Robust Features) [32]，从算法原理上进行简化，加快了算法的速度。SURF 对高斯滤波后的图像的像素点的 Hessian 行列式进行简化和近似，运算速度可以比 SIFT 算法快 3 倍左右。经过高斯滤波后再进行 Hessian 的计算，其公式如下：

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (2-4)$$

其中， $L_{xx}$ 、 $L_{xy}$ 、 $L_{yy}$  为高斯滤波后图像  $g(\sigma)$  在各个方向的二阶导数，也可看成是高斯二阶微分与图像的卷积。

为了计算简便，使用盒子滤波器代替传统的高斯卷积模板。在盒子滤波器的每一个子区域中系数相同，为了方便计算区域内的像素和，SURF 算法使用了积分图，这样滤波操作也可以转化为加减法，速度得到了进一步的加快。

SURF 确定主方向和生成描述子的方法与 SIFT 不同。在 SURF 中，统计特征点圆形邻域内的 Haar 小波响应。每个扇形内的响应叠加得到新的向量，向量长度越大，说明在这个方向变化情况越剧烈，选取长度最大的方向作为主方向。生成描述子是在特征点旋转到主方向之后进行的，在特征点  $4 \times 4$  大小的矩形邻域中计算 Haar 小波响应，每个子区域得到四个统计量： $\sum dx, \sum |dx|, \sum dy, \sum |dy|$ ，这样每个特征点得到了  $16 \times 4 = 64$  维的向量作为描述子。相比 SIFT 而言，SURF 算法的描述子维度降低到 1/2，且由于 Haar 小波响应的计算也可以利用积分图像，进一步加快了速度。

## BRISK

2011 年的 ICCV 会议上出现了一种快于 SURF 的算法：BRISK (Binary Robust Invariant Scalable Keypoints)。

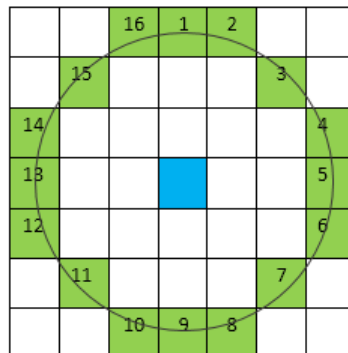


图 2-2 FAST 特征点检测法

BRISK<sup>[50]</sup>在特征点检测上使用了 2008 年得到完善的 FAST (Features From



Accelerated Segment Test)检测方法<sup>[51][52]</sup>,使用下采样建立了图像金字塔,实现了 FAST 特征点的尺度不变性。在构建该金字塔时只使用了下采样而没有使用高斯滤波,所以相当于是对 SIFT 中金字塔的简化与缩小。使用了 BRIEF 描述方法中的均匀采样,为了避免混叠效应,在采样之前先对特征点的邻域的采样点进行高斯滤波:以特征点为圆心的同心圆族中,同一个圆上的采样点采用相同的滤波系数,半径越大的同心圆上使用更大的滤波系数。将采样点对按照阈值划分为长距离子集和短距离子集,使用长距离子集中估计得到的局部梯度确定了特征点的主方向,在将特征点旋转到此方向后,使用短距离子集中的采样点对得到二进制的描述子。短距离子集,长距离子集按照式 2-4 定义,  $p_i, p_j$  是两个随机的采样点对。

$$\begin{aligned} S &= \left\{ \left\| (p_i, p_j) \right\| < \delta_{\max} \right\} \\ L &= \left\{ \left\| (p_i, p_j) \right\| < \delta_{\min} \right\} \end{aligned} \quad (2-5)$$

## ORB

在 BRISK 算法提出的同年, OpenCV 实验室提出了 ORB (Oriented FAST and Rotated BRIEF)算法<sup>[36]</sup>。相比于 SIFT 和 SURF, ORB 比 SIFT 快两个数量级,比 SURF 快 10 倍,并且 ORB 在使用中不必担心专利的问题。作者在安卓手机上进行了验证了 ORB 算法可以真正实现实时性的特征点检测和匹配。与 BRISK 算法一样, ORB 在特征点检测和特征点的描述方法上分别借鉴了速度最快的 FAST 算法和 BRIEF 算法,针对没有尺度不变性和没有旋转不变性的缺点进行了改善。FAST 与图像金字塔架构结合起来实现了尺度不变性;对于 FAST 特征对边缘敏感的问题,使用 Harris 角点滤波器剔除检测到的边缘;针对 FAST 没有旋转不变性的缺点,ORB 没有像在 SIFT 中一样使用梯度直方图,也不像 SURF 一样借助 Haar 小波,而是选择了质心的方法。对于每个特征点的邻域,邻域会有一个质心,特征点和质心一般不会重叠,特征点到质心构成的矢量就是成为该特征点的主方向。ORB 基于 BRIEF 提出了两种描述子,一种是 steer BRIEF,另外一种为 rBRIEF。tests 是 BRIEF 中点对的集合,通过两点的比较得到二进制字符串。steer BRIEF 记录了旋转之前选择的 tests 坐标保证旋转不变性,但损失了 BRIEF 本身方差和均值的特性。为了不增加新的计算量,没有使用主成分分析方法 (Principal Component Analysis, PCA),而是在 rBRIEF 中使用了贪婪查找的方法找到了方差较大,均值为 0.5,且保证了采样点对之间相关性较低。在寻找最近邻点匹配时,ORB 使用了局部哈希 (Locality Sensitive Hashing, LSH)<sup>[53]</sup>来提高效率。实验表明 LSH 比构建 KD 树还要快。在户外和室内的数据集中,ORB 优于 SIFT 和 SURF。但是在涂鸦图像中 SIFT 的表现更好。

## KAZE

虽然 SIFT, SURF 等算法在图像特征提取和匹配方面表现良好,但是这些算法的特征点都基于在线性尺度空间。2012 年 Alcantarilla 等提出了一种基于非线性尺度空间的特征点检测算法 KAZE<sup>[40]</sup>。线性尺度空间通过高斯核构建,在线性尺度空间中,相同尺度下图像中各个点的尺度是一样的,都会因为高斯函数的低通滤波受到相同程度的平滑,其中也包括图像中的边缘特征。边缘特征包括图像内容的一些轮廓,边界等。边缘特征包含了更多的细节信息,所以我们在构建尺度空间的同时可以保留这些边缘特征。非线性尺度空间的使用是 KAZE 与 SIFT、SURF 最大的不同,它保证了图像边缘在尺度变换中的不被模糊化,从而极大地保留了图像细节信息。非线性尺度空间的构建基于非线性扩散滤波<sup>[54]</sup>:

$$\frac{\partial L}{\partial t} = \text{div}(c(x, y, t) \cdot \nabla L) \quad (2-6)$$

其中  $\text{div}$  和  $\nabla$  分别表示散度和梯度,函数  $c(x, y, t)$  表示扩散的传导函数,参数  $t$  是尺度参数。为了尽量使平滑发生在区域内而不是区域间的边界上,减少在边缘处的扩散,Perona 提出了基于图像梯度幅值的传导函数<sup>[55]</sup>:

$$c(x, y, t) = g(|\nabla L_\sigma(x, y, t)|) \quad (2-7)$$

其中  $L_\sigma$  是原始图像在高斯平滑后的梯度。非线性偏微分方程没有解析解,文章中使用了加性算子分裂算法 (Additive Operator Splitting, AOS) 进行求近似解<sup>[56]</sup>。采用任意步长来构造稳定的非线性尺度空间。

## AKAZE

KAZE 在构建非线性空间的过程中很耗时,针对这个 KAZE 的作者提出了 KAZE 的加速版本 AKAZE (Accelerated-KAZE)。AKAZE<sup>[57]</sup>的配准表现和 KAZE 接近,但是 AKAZE 在速度方面加快了几个数量级。AKAZE 将快速显式扩散 (Fast Explicit Diffusion, FED) 加入到图像金字塔中,通过循环求解,可以更快更准确地构建非线性的尺度空间,且易于实施。在描述子方面,AKAZE 使用了更高效的基于 LDB (Local Difference Binary)<sup>[58]</sup>的 M-LDB (Modified Local Difference Binary),利用了从非线性尺度空间提取的梯度 (gradient) 和强度 (intensity) 信息。M-LDB 是旋转和尺度不变的,并且对内存的要求更低。因为 M-LDB 可以利用之前的特征点提取时计算得到的导数,所以减少了生成描述子阶段的计算量。

## 2.4. 实验结果与分析

### 2.4.1 实验内容

实验平台为普通的 PC 机, 采用 Intel Core 处理器, 主频为 2.66GHz, 操作系统为 Windows7。开发环境为 Visual Studio 2013, 使用 C++ 语言编程。实验图像来自比利时鲁汶大学特征检测项目组公布在网络上的开源资料。

编程实现了图像的亮度变化, 高斯模糊变化, 旋转变化和尺度缩放变化, 亮度变换的实现方法是在对每一个像素值的每一个通道加一个相同的分量, 所以这里的亮度变换是整体相同幅度的亮度变化。高斯模糊是最常用的一种滤波方法, 模拟的是图像因为成像条件不同造成的模糊失真, 这里保持水平方向和垂直方向的方差为 0, 改变高斯模板的大小进行不同水平的模糊。旋转通过仿射变换实现, 以图像的中心为旋转中心, 不断改变旋转的角度; 保证旋转的时候不进行尺度变换。尺度变换通过 OpenCV 库中的 `resize` 函数完成, 内插方法使用区域内插。

对在这四种变化下的图像使用不同的算法进行了配准的实现。几种算法都属于基于特征点的算法, 而且真实的变换矩阵是已知的, 所以在对配准结果的评价指标中选取了矩阵误差 `HomographyError`、正确匹配对的占比 `correctMatchesPercent`、平均距离 `Meandistance`、峰值信噪比 `PSNR`。对于算法的实时性, 考察了算法对于每一幅图像的平均耗时, 在每一个特征点上的平均耗时, 和每一幅图像的特征点数目的均值。

### 2.4.2 实验指标

矩阵误差描述的是估计出的变换矩阵与真实的变换矩阵之间的差距。事实上, 在编程实现中, 根据匹配点对一开始求解出的变换矩阵与真实的变换矩阵理论上应该是互逆的, 即二者乘积是单位矩阵。实际中乘积与单位矩阵的差值往往不为 0, 求这个差值的无穷范数作为评价误差的指标。矩阵差值越小越好。

正确匹配对的占比描述的是匹配正确的特征点占有所有检测出的特征点的比例。因为种种原因, 会存在误匹配对, 即它们之间本来不是对应点却被匹配在一起, 这无疑会影响最终的配准结果。因为每一个特征点会有一个描述子, 当描述子之间的距离误差小于 3 时认为是正确匹配。

峰值信噪比利用均方差 ( $MSE$ ) 定义, 均方差可以理解为是均方根误差的平方。两个  $m \times n$  大小的单色图像  $I$  和  $K$ , 一个为另外一个的近似, 那么它们的均方差被定义为像素点之间距离的平方之和的均值:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i, j) - K(i, j)\|^2 \quad (2-8)$$

峰值信噪比定义为:

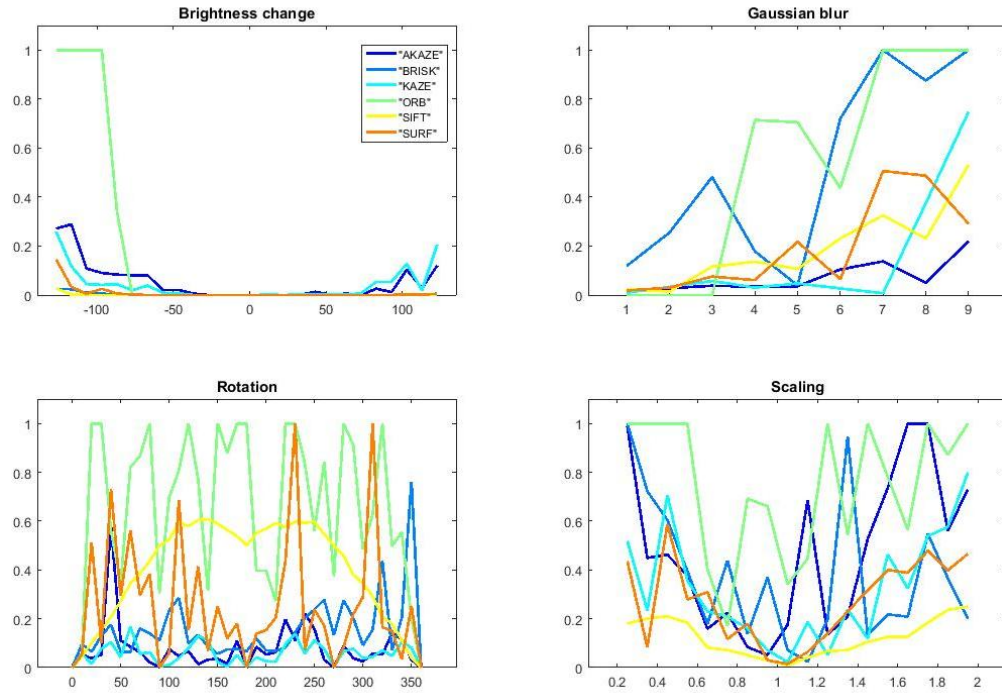
$$PSNR = 10 \cdot \log_{10} \left( \frac{MAX_I^2}{MSE} \right) = 20 \cdot \log_{10} \left( \frac{MAX_I}{\sqrt{MSE}} \right) \quad (2-9)$$

需要注意的是均方根误差中的计算对象是特征点之间的灰度值,而使用峰值信噪比计算两幅图像相似程度时使用的是全部像素点。峰值信噪比中比较的是说明两幅图像之间越相似。一般来讲在 30dB 以上可以认为两幅图像很接近。峰值信噪比的缺陷是没有考虑人眼在观察时对波长、空间频率不同的区域有差异性。

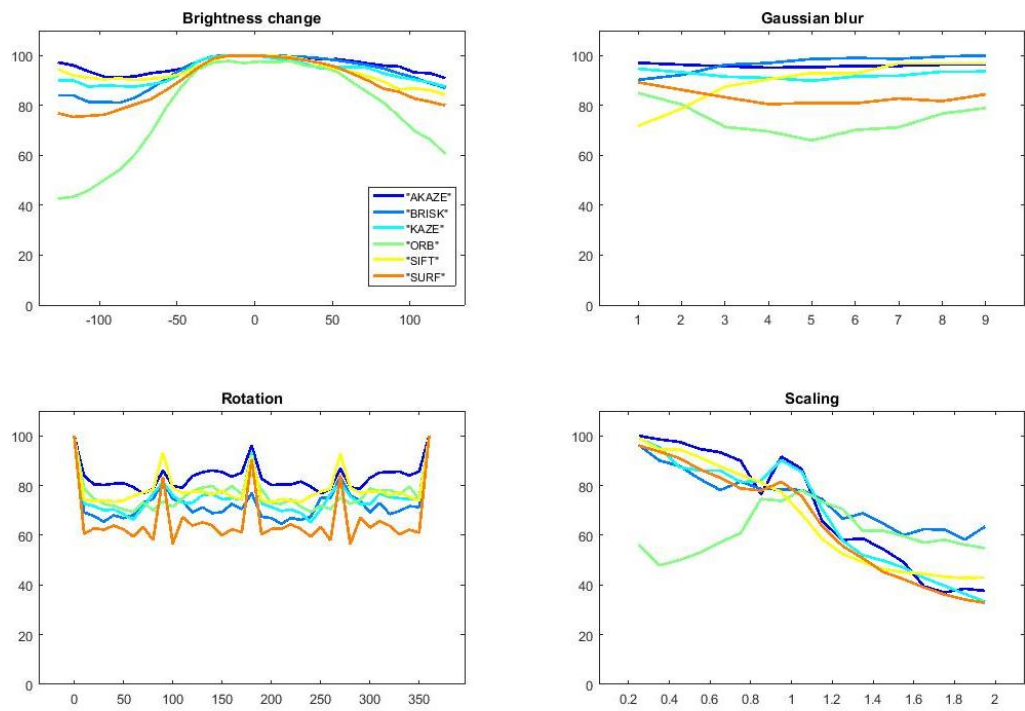
除了配准算法中对于特征点的准确度和匹配的成功率,我们关注的还有算法的实时性。本次实验中统计了不同算法下每次配准的平均运行时间。配准的耗时与特征点数目有关,本文计算了每种算法检测出特征点数目的均值,计算出了不同算法平均到每个特征点下的运行时间。

### 2.4.3 实验结果

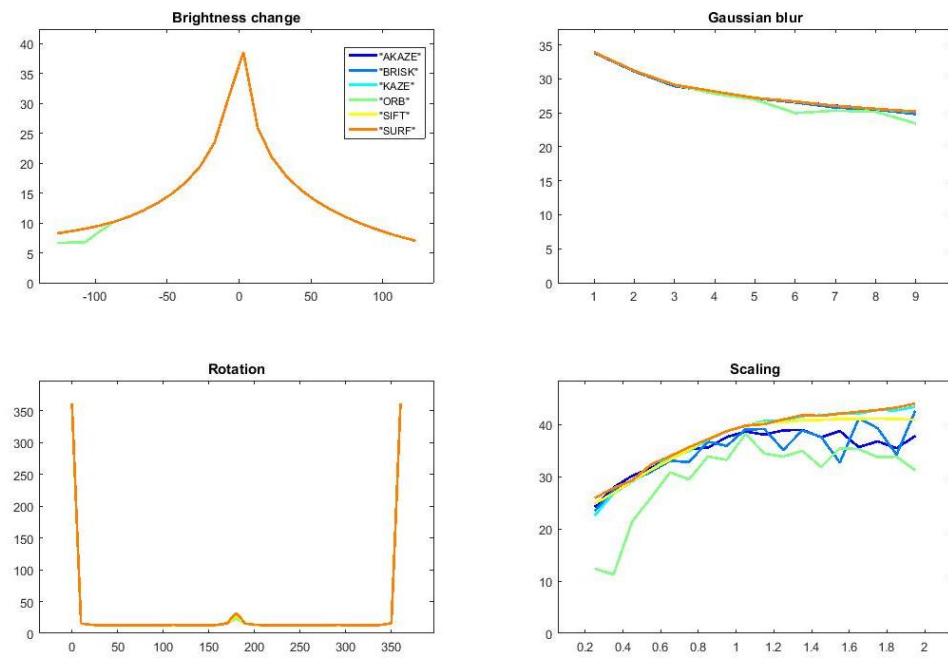
下面是根据统计数据生成的 txt 文件使用 matlab 软件画出的特性曲线。不同图像使用了不同的评价指标,每幅图像中又分成了四个子图,代表了实验模拟的四种变换。不同颜色的曲线代表不同的特征点检测和匹配算法。



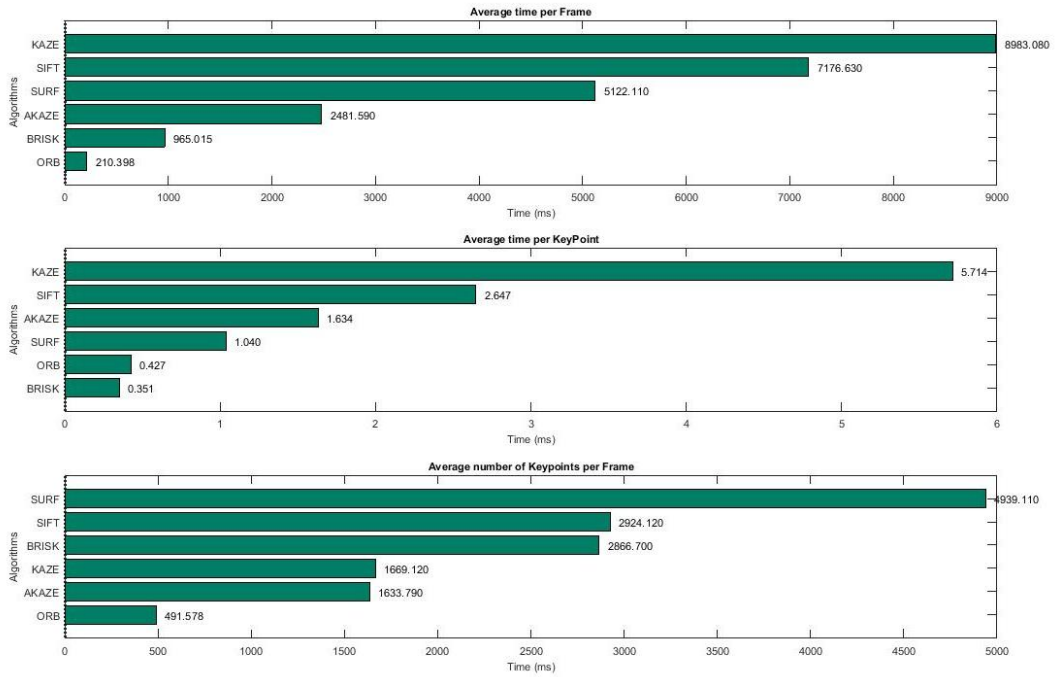
(a) Homography Error



(b) correctMatchesPercent



(c) Meandistance



(d) Speed

图 2-3 展示了几种算法在不同图像变换条件下的表现

根据上图中不同算法的性能表现曲线可以得出以下结论。当评价标准为矩阵误差时，SIFT 和 SURF 算法对于图像亮度变换的表现最好；KAZE 算法和 AKAZE 算法对于高斯模糊变换和旋转变换的效果都最好；当图像发生尺度变换时，表现最好的算法是 SIFT。当评价标准为正确匹配对的占比时，AKAZE 算法对于图像的四种变化几乎都拥有最好的表现，当图像发生亮度变化时 SIFT 算法的表现和 AKAZE 接近。当评价标准为特征点之间的平均距离时，SURF 拥有最好的表现，平均距离接近于 0，除却 SURF，图像亮度变化时，表现最好的算法是 SIFT，高斯模糊和尺度变化、旋转变换时，表现最好的算法是 ORB 算法。当评价标准为峰值信噪比时，几种算法在图像发生亮度变化，高斯模糊，旋转变换下的配准表现几乎一样，当图像发生尺度变换时，表现最好的算法是 SIFT、SURF、KAZE。

在实时性方面，统计了每帧图像的平均耗时、每个特征点的平均耗时和每帧图像中检测出的特征点数的平均值。对于每帧图像的平均耗时，耗时最长的依次是 KAZE、SIFT、SURF、AKAZE，ORB 耗时最短，可以缩短为 KAZE 的 2.3%，AKAZE 可以缩短为 KAZE 的 27.6%。对于每个特征点的平均耗时，最长的仍然是 KAZE 和 SIFT，SURF 略快于 AKAZE，耗时最短的是 BRISK 和 ORB。检测出的特征点数目方面，平均值最多的是 SURF 检测出的特征点，接近 5000 个，其次是 SIFT 和 BRISK，二者数量接近，大约为 3000 个，其次是 KAZE 和 AKAZE，大约为 1700 个，最少的是

ORB, 平均不到 500 个特征点。

## 2.5. 本章小结

本章介绍了基于特征点的图像配准算法的大致流程。属于这一类的算法有 SIFT、SURF、BRISK、ORB、KAZE、AKAZE 等, 本章对这些算法进行了简单的介绍。配准的实验图像一般存在四种变化: 因为光照改变发生的亮度变化, 高斯模糊变化, 旋转变换和尺度变换。本章设计了实验对其进行仿真和比较, 画出了不同算法的特性曲线, 并使用不同的指标从多个角度对以上算法进行对比。可以看到, 其他基于 SIFT 改进的算法的改善有限, SIFT 算法依然有其一定的优越性, 同时因为 SIFT 依然是应用最广泛的算法, 所以在本文之后的部分依然在 SIFT 算法上进行改进。





## 第3章 基于 SIFT 算法的快速图像配准算法

### 3.1. 引言

从前一章的工作中可以看出，不同算法在不同变换下的表现不尽相同，其中以 AKAZE 和 SIFT 的综合性能最好最为稳定。而 SIFT 算法作为里程碑意义的算法，同时具有巨大的改进空间，得到了国内外很多学者的研究。本章针对 SIFT 算法中计算量过大的问题，结合图像金字塔的特点，以下采样的方法为预处理，提出了一种可以实现快速配准的算法。该算法不必改变 SIFT 本身的架构，可以在加快速度的同时保证较好的配准效果，在实验中甚至可以在速度和峰值信噪比两方面同时优于 SURF 算法。

### 3.2. 所提算法框架

本算法框架包括如下几部分：使用下采样进行预处理；构建图像金字塔；使用 SIFT 算法对特征点进行提取和描述；通过匹配得到的特征点对进行变换模型的估计；最终得到单应矩阵来实现图像配准。

根据预处理步骤中下采样的对象可以将本文中的算法细分为两种，两种方法估计出的单应矩阵有所区别。第一种如图 3-1 所示，只对待配准图像进行下采样的预处理。

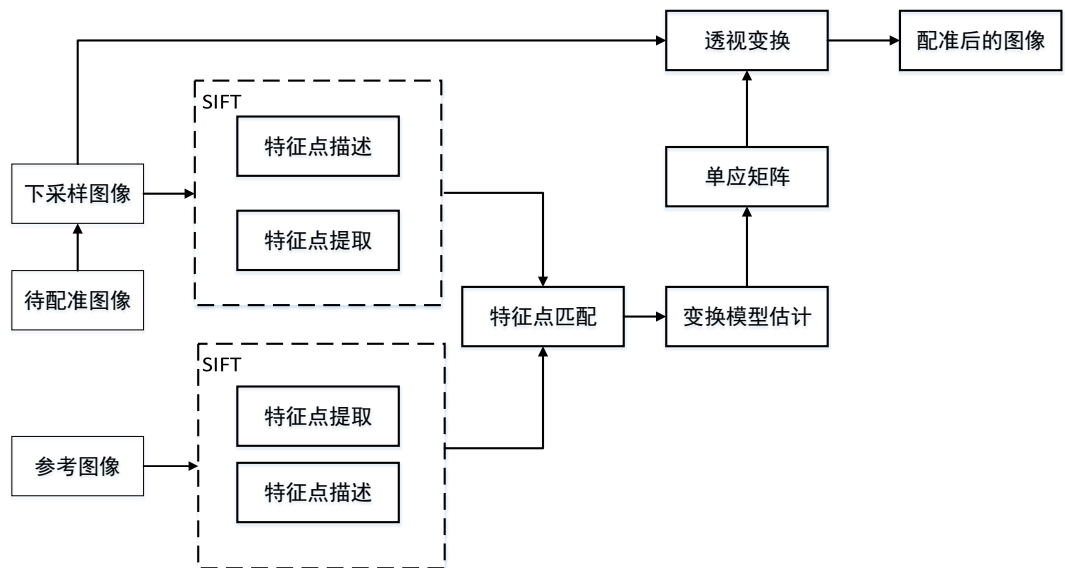


图 3-1 对单幅图像（待配准图像）预处理的快速配准框图

按照图 3-1 估计出的单应矩阵可以直接作用在待配准图像的下采样图像之上，下

采样得到的图像经过透视投影变换和像素点的重映射和插值得到配准之后的图像。

还有一种方法是对两幅实验图像（待配准图像和参考图像）都进行下采样的预处理，构建得到的两个图像金字塔都得到了缩小。在该框架下的算法得到单应矩阵表示的是两幅下采样图像之间的关系。为了得到描述两幅原始的两幅图像之间变换关系的单应矩阵需要对求出的单应矩阵进行修正，使用修正后的单应矩阵和原始的待配准图像得到配准之后的图像<sup>[59]</sup>。流程如图 3-2 所示。

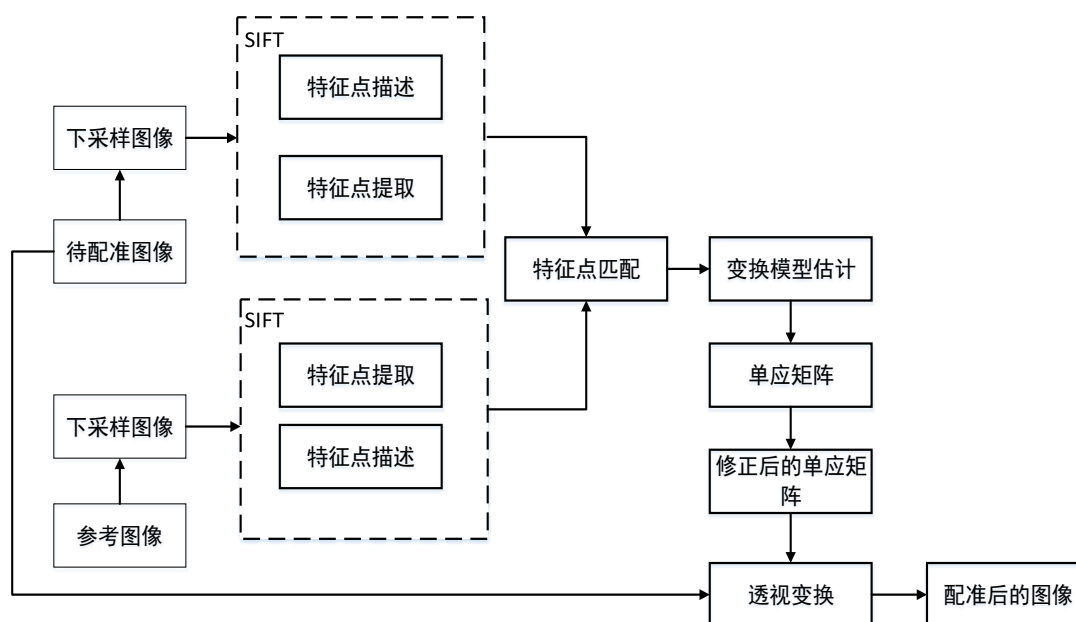


图 3-2 对两幅图像预处理的快速配准框图

### 3.3. 金字塔的构建和极值点的寻找

透视投影有近大远小的特点，这是符合我们对于观察的常规认知的。在计算机图形学和数字图像处理中，为了使得计算机可以将不同尺度大小的图像识别为同一目标，需要使得算法具有尺度不变性，即无论图像处于什么样的尺度下，计算机或者算法都可以在不同尺度下的图像中识别出相同的图像属性或者特征。在 SIFT 算法中解决尺度不变性的方法是建立图像的高斯金字塔，在这个基础上再使用高斯拉普拉斯函数的极值来提取特征点，就得到具有尺度不变性的特征点。

SIFT 算法中利用图像的金字塔模拟了大小不同，清晰度不同的图像，实现了尺度不变性。图像的高斯金字塔中有两个元素，其中一个下采样，另外一个高斯模糊算子。使用高斯滤波是因为远处的物体不仅尺度变小，图像的清晰度也受到了影响，

而高斯核是构建线性尺度空间的唯一线性核，这出自于 Lindeberg 的研究成果<sup>[60][61]</sup>。高斯滤波是图像卷积的一种，使用了高斯核作为卷积模板。以方形图像为例，图像卷积前后的大小关系可由式（3-1）表示。

$$N = (W - F + 2P) / S + 1 \quad (3-1)$$

$N$  是卷积之后的图像边长， $W$  是卷积之前的原图的边长， $F$  是滤波器的边长， $P$  是填充边缘像素的大小， $S$  表示卷积滤波器的滑动步长。一般使用的卷积模板有 Sobel 算子，Prewitt 算子，Roberts 算子，Scharr 算子等。这些算子和高斯滤波算子一样，都是  $3 \times 3$  的大小，在滤波过程中一般选择  $F=3, P=1, S=1$ ，所以在图像高斯金字塔中滤波前后的图像大小保持不变，同时也说明了整个过程中卷积的次数与图像的尺寸大小成正比。所以整个过程中的乘法和加法总次数可以用式（3-2）表示。

$$(F * F + 1) * W * W = (S_F + 1) * S_{in} \quad (3-2)$$

式（3-2）中  $S_{in}$ 、 $S_F$  分别表示卷积运算中输入图像的面积和卷积模板的面积。所以输入图像的面积大小与图像高斯卷积的运算量成正比，而图像高斯金字塔中不同层之间是 2 倍下采样的关系，所以在整个图像高斯金字塔中的运算次数  $Num$  可以表示为：

$$\begin{aligned} Num &= (S_F + 1) * S_{in} * s + (S_F + 1) * \frac{S_{in}}{2} * s + \dots + (S_F + 1) * \frac{S_{in}}{2^O} * s \\ &= (S_F + 1) * S_{in} * s * \left( 1 + \frac{1}{2} + \dots + \frac{1}{2^O} \right) \end{aligned} \quad (3-3)$$

其中  $s$  代表图像高斯金字塔中每组的数目， $O$  表示层数。同一组的不同图像大小相同，高斯模糊因子不同；不同层之间是 2 倍下采样的关系。所以可以得出，整个高斯金字塔上的卷积中运算量与输入图像的尺寸大小成正比。

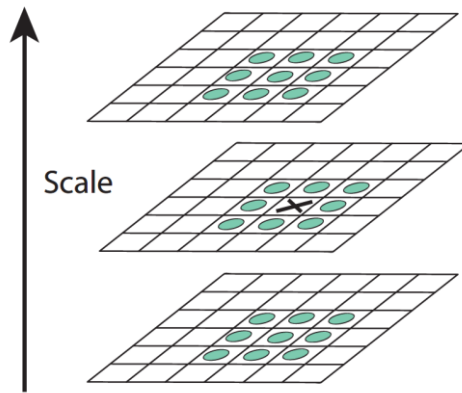


图 3-3 SIFT 算法中在差分尺度空间寻找极值点

图像高斯金字塔的规模不仅与卷积的运算量有关，还与寻找极值点的过程的运算

量有关。图 3-3 是在图像差分金字塔中寻找极值点的过程。**SIFT** 算法在差分高斯金字塔上的相邻尺度的上下两层和本身尺度的空间邻域中寻找极值点，经过泰勒展开逼近得到真正极值点作为特征点。

所以如果可以将金字塔的规模缩小，就可以成比例地减少计算量，降低算法的复杂度。比如 **BRISK** 算法<sup>[50]</sup>中直接舍弃了高斯模糊的过程，改为完全使用下采样建立金字塔，不同层分别按照不同的下采样比例得到。本文在 **SIFT** 算法构建金字塔的方法和 **BRISK** 算法之间取了一个平衡，保留金字塔同一组不同层之间的高斯模糊处理，将原始金字塔中第一层的原图改为原图经过下采样之后的图像。因为 **SIFT** 图像金字塔中主要的处理之一就是下采样，直接使用下采样之后的图像建立金字塔相当于截取了原始金字塔中顶端的一部分，得到的金字塔规模更小。

### 3.4. 对于单应矩阵的修正

#### 相机模型

相机模型有两种，分别是正交投影照相机（Orthographic Camera）和透视投影照相机（Perspective Camera）。正交投影相机模型保证了投影前后大小不变，适用于工业中的制图，而投影透视更符合我们的常识，也是更加常用的模型。

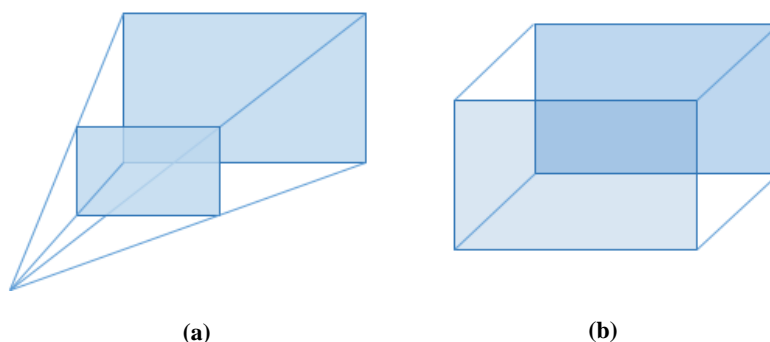


图 3-4 两种投影相机模型

在透视投影模型中又以针孔相机模型最为常用。在针孔相机模型中透镜被理想化为一个点，大大降低了复杂度。这样，两幅图像之间的变换也使用透视投影变换模型来表示。透视投影变换在数学上通过单应矩阵表示。单应矩阵与图像坐标系、齐次坐标和双目视觉有关。

#### 图像坐标系

图像坐标系有四种：世界坐标系，相机坐标系，图像物理坐标系，像素坐标系。

(1) 世界坐标系，也称客观世界坐标系，客观地描述了相机的位置和物体的

位置，单位是物理长度：米。坐标是三维的，用 $(X, Y, Z)$ 表示。世界坐标系使用右手坐标系。

(2) 相机坐标系和世界坐标系相似，只不过更加规则化。相机坐标系规定了坐标原点是相机的光心， $Z$ 轴为相机的光轴，方向指向相机前方，一般取 $X$ 轴与 $Y$ 轴分别为水平和垂直方向。相机坐标系也使用右手坐标系，单位依然是物理单位。相机坐标系与世界坐标系之间是刚体变换，即只存在旋转和平移的关系。 $R$ ， $T$ 分别为旋转和平移向量，那么从世界坐标系到相机坐标系的转换可以用式(3-4)表示。

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + T \quad (3-4)$$

(3) 图像物理坐标系，因为区别于前两个三维空间中的坐标系，所以也称为平面坐标系。图像物理坐标系原点在传感器中心， $X$ 轴与 $Y$ 轴分别与图像的边界平行。图像物理坐标系依然使用物理单位和右手坐标系，但因为相机传感器尺寸一般较小，所以一般使用毫米作为物理单位。相机坐标系到图像物理坐标系的转换关系可以通过图3-5中的相似三角形推导得到。

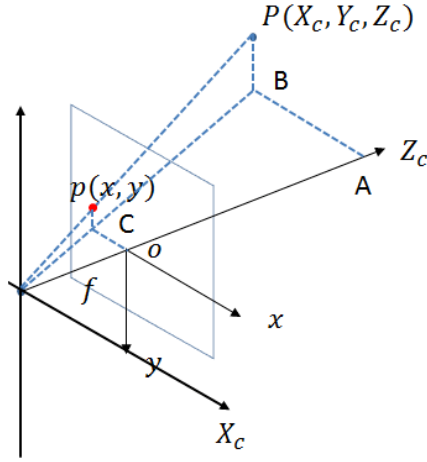


图 3-5 相机坐标系到图像物理坐标系的转换

式(3-5)表示了相机坐标系下的坐标与图像物理坐标系下的坐标的关系。式中 $f$ 指相机的焦距， $x, y$ 指图像物理坐标系下的坐标， $X_c, Y_c, Z_c$ 是相机坐标系下的坐标。

$$x = f \frac{X_c}{Z_c}, y = f \frac{Y_c}{Z_c} \quad (3-5)$$

(4) 像素坐标系是我们常用的坐标系，它与其他坐标系最大的不同就是坐标原点的选取与单位的选取。坐标原点取在图像左上角，单位使用离散值。像素点的坐标表示该像素点与原点相隔了多少个像素。

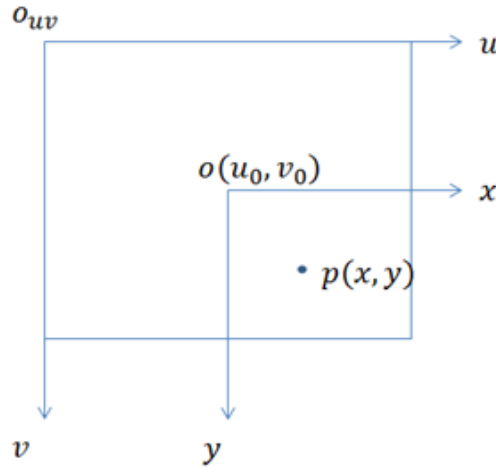


图 3-6 图像物理坐标系到像素坐标系的转换

图像物理坐标系到像素坐标系的转换有两个要素，一个是进行平移变换，利用相机光心的坐标  $(u_0, v_0)$  来表示；还需要进行单位的转换，距离通过包含多少个像素点的离散值来衡量。每个像素大小为  $dx, dy$ ，那么由图像物理坐标系到像素坐标系可以使用式 (3-6) 完成转换。

$$\begin{cases} u = \frac{x}{dx} + u_0 \\ v = \frac{y}{dy} + v_0 \end{cases} \quad (3-6)$$

所以，我们可以从世界坐标转换到像素坐标，这也是数字图像的数字化生成过程。

### 齐次坐标

在上文介绍坐标系中我们使用的是笛卡尔坐标，因为我们经常使用的是欧式空间，但是在计算机图形学应用更广泛的其实是齐次坐标（Homogeneous Coordinates）。对于同一个点，齐次坐标比笛卡尔坐标高一个维度：齐次坐标  $(\hat{x}, \hat{y}, \omega)$  对应笛卡尔坐标  $\left(\frac{\hat{x}}{\omega}, \frac{\hat{y}}{\omega}\right)$ ，一般取缩放系数  $\omega=1$ 。当我们选择投影变换中的透视投影时，平行线在经过变换之后可能会在无穷远处相交，而这是欧式空间和笛卡尔坐标无法表示的。齐次坐标的另外一个重要意义是利用其特殊的高维表示，可以将式 (3-4) 这样包含平移的变换表示为矩阵相乘的形式：

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, R: 3 \times 3, T: 3 \times 1 \quad (3-7)$$

### 双目视觉

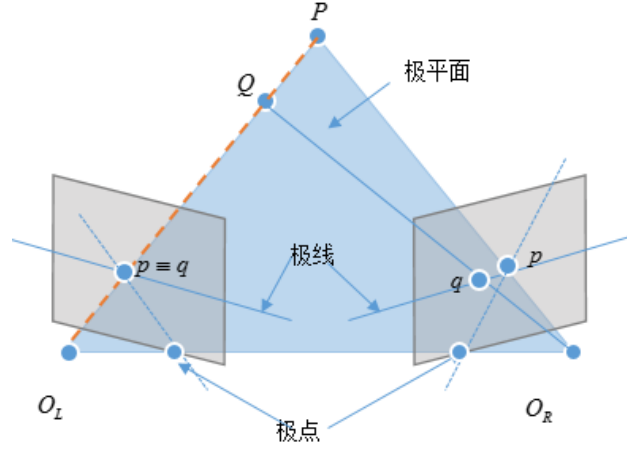


图 3-7 对极几何

双目视觉涉及的知识主要是对极几何 (Epipolar Geometry)，其中以本质矩阵和单应矩阵最为重要，它们表示了同一场景下在两个不同的成像平面上的对应像素点之间的约束关系。图 3-7 是对极几何的示意图。

$P, Q$  及其连线上的空间点在左边的成像平面上重合为一点，在另外一个成像平面是共线的关系。这种约束关系可以通过本质矩阵来表示。本质矩阵的实质是描述了两个相机之间的关系，并用旋转矩阵和平移向量表示，凭借相机相对关系的先验知识建立了同一个三维空间点在不同相机坐标系下的约束关系。本质矩阵  $E$  (Essential matrix) 如式 (3-8)，则两个三维空间点在相机坐标系下的坐标  $P_L, P_R$  有如下关系：

$P_R^T E P_L = P_R^T R S P_L$ 。其中  $T$  表示不同方向下的平移向量， $R$  为旋转矩阵。

$$E = RS = R \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \quad (3-8)$$

在本质矩阵的基础上，为了便于图像处理，将相机坐标系转化成像素坐标系，得到坐标  $\bar{p}_r, \bar{p}_l$ ，这时两个坐标之间的约束关系通过基础矩阵  $F$  (Fundamental matrix) 来表示： $\bar{p}_r^T F \bar{p}_l = 0$ 。

$$F = M_r^{-T} R S M_l^{-1} \quad (3-9)$$

其中  $M_l$  和  $M_r$  是两个仿射矩阵，分别描述了两个相机下由图像物理坐标系到像素



坐标系的转化关系。

### 单应矩阵

当空间中两个点对于两个相机来说处于同一个平面时，约束关系可以得到加强，约束关系通过单应矩阵（Homography matrix）来表示。

$$H = R + T \frac{1}{d} N^T \quad (3-10)$$

其中  $R$  是旋转矩阵， $T$  是平移向量， $N$  是其中一个相机对于平面的法向量， $d$  是相机到平面的距离（景深深度）。特别的，如果能保证平移向量为零，即相机只存在旋转运动，那么单应矩阵将更加简单： $H = R$ 。单应矩阵约束力更强，是因为在两个镜头间的旋转和平移关系的先验基础上，还包含了景深信息  $d$ 。单应矩阵使用的场景是拍摄画面为平面图像，或者相机离目标的距离足够远，使得画面中各个空间点到相机的距离大致相等。

对待配准图像和参考图像都进行了下采样的预处理，若都使用 2 倍下采样的缩放比例，这样可以同时缩小两个图像金字塔的规模，但是这样特征点的检测、特征点的匹配发生在两幅下采样图像之间，估算出的变换矩阵描述的也是两幅下采样图像之间的变换关系。我们当然可以对待配准之后的图像进行 2 倍上采样得到与原始参考图像一样的形态，但是这样可能会因为下采样图像的失真造成配准效果的不佳。所以我们依旧采用了对原始待配准图像进行配准的方法。

为了由原始的待配准图像经过变换得到原始参考图像的形态，我们需要根据得到的变换矩阵进行适当的修正，得到两幅原始实验图像之间真正的变换矩阵。当下采样图像之间的变换矩阵为

$$H_p = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \quad (3-11)$$

在 OpenCV 官方手册提供的函数中，使用 `cv::warpPerspective` 函数进行透视投影变换时依照的映射关系如下式表示。其中  $dst$  表示变换后图像， $src$  表示变换前的图像。正如之前的介绍，在计算机图像处理中，变换后的图像像素是由变换前所映射的图像像素经过运算得到的。

$$dst(x, y) = src\left(\frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}, \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}\right) \quad (3-12)$$

在下采样比例为  $s$  时，可以推导出描述两幅原始图像变换关系的矩阵为：

$$H'_p = \begin{pmatrix} h_{11} & h_{12} & h_{13}/s \\ h_{21} & h_{22} & h_{23}/s \\ sh_{31} & sh_{32} & h_{33} \end{pmatrix} \quad (3-13)$$



这样就可以顺利地得到原始的待配准图像与原始参考图像之间的映射关系, 经过像素点的重映射就可以完成配准。

### 3.5. 实验结果与分析

#### 插值方法

下采样的实质是像素点的重映射, 使用更少的像素点表示原来的图像, 这些像素点的值如何与原图中的像素点相对应就是插值。插值方法的选取不仅直接关系到新建的图像高斯金字塔的质量, 还直接关系到配准的效果, 因为配准的最后一步, 是由下采样图像根据变换模型得到配准后的图像。

常用的插值方法包括: 最近邻插值、双线性插值、双三次插值、B 样条插值、高斯插值。采用最邻近插值法工作量比较小, 算法实现起来也比较简单, 但是插值后图像有明显的锯齿状。双三次卷积法对图像边缘有增强的作用, 插值后图像的质量得到明显的改善, 但是大大增加了计算量。双线性插值法考虑了点周围的像素点的影响, 解决了最邻近插值法中图像连续性不强的缺点, 计算量与双三次卷积法相比也有所下降, 插值后的整体效果比较令人满意, 是 OpenCV 等图像处理库中的默认设置。为了得到更加清晰的图像, 还可以使用超分辨率重建技术<sup>[62]</sup>。

图 3-8 是 OpenCV 库中自带的实验图像, 图 3-10 是使用不同插值方式得到的下采样图像。实验平台和第二章相同, 这里使用 Python3.6 编程, 将 Lena 原图进行十倍下采样之后的仿真, 图像 Lena 分辨率从  $512 \times 512$  大小缩小  $51 \times 51$  大小 (图像长宽为离散的整数值, 这里对按比例缩小之后的数值进行了向下取整)。依次为区域插值 INTER\_AREA、立方插值 INTER\_CUBIC、线性插值 INTER\_LINEAR 和最近邻插值 INTER\_NEAREST。为了便于对比, 在显示时放大到了相同的大小, 但保持其分辨率大小不变。



图 3-8 下采样插值实验中使用的原图 Lena.bmp



图 3-9 不同插值方法的实验图像对比

由图 3-9 可以看到，区域插值之外的三种方法都有明显的马赛克效应，其中又以最近邻插值方法马赛克效应最明显，人物面部失真最为严重。立方插值、线性插值效果接近，立方插值在边缘处的模糊化程度更高。

### 改进后的配准

本实验的实验环境和上文中的实验一致。实验图像来自实现 SIFT 算法的作者 Rob Hess。beaver.png 和 beaver\_xform.png，分别作为待配准图像和参考图像。两幅图像之间存在着空间上的形变，配准的目的是找到能描述二者之间变换关系的单应矩阵，将 beaver.png 变换为 beaver\_xform.png 相同的形态，并且使得二者尽可能地相似。本实验在满足一定相似度的条件下，追求更快的处理速度。结合上文中的插值方法的实验，算法中使用了双线性插值作为下采样的重映射方法，按照两个框图实现了仿真，并且与原本的 SIFT 算法和 SURF 算法进行了对比。

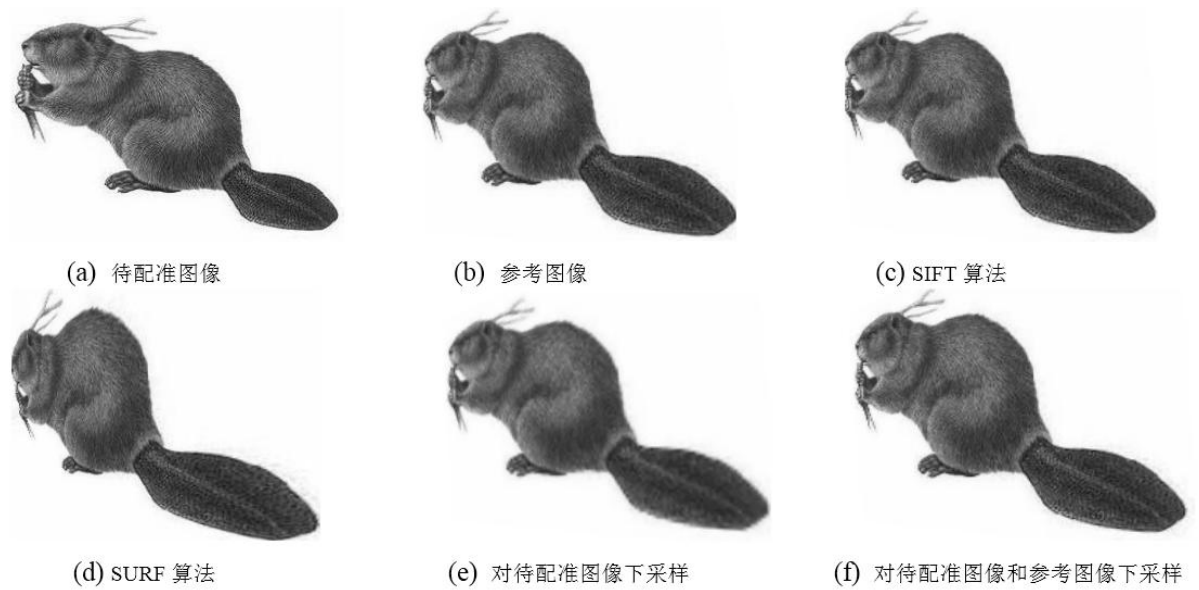


图 3-10 下采样缩小图像金字塔之后的图像配准

直观上可以看出 SURF 算法较差，本文提出的两种改进方法与 SIFT 下的效果相近。为了进一步比较几种不同的算法，包括 SIFT、SURF、只对待配准图像进行下采样的预处理和同时对两幅图像预处理，本文制作了如下表格，主要从运行时间  $t$  和配准后的图像与参考图像之间的峰值信噪比  $R$  两个方面来比较。

表 3-1 不同预处理方法下的图像配准效果比较

参数	SURF	SIFT	本文方法	
			只对待配准图预处理	同时对 2 幅图预处理
$t/s$	0.858	1.421	1.083	0.537
$R/dB$	20.48	38.72	36.82	31.21

下采样利用了图像金字塔对尺度变化具有鲁棒性的特点，缩小了图像高斯金字塔的规模。从表中可以看到，使用下采样的预处理，无论是对单幅图像下采样还是同时对两幅实验图像下采样，都可以在不改变 SIFT 算法本身架构的基础上实现配准的加速。对单幅图像下采样对速度的提升有限，相比于原始的 SIFT 算法可以缩短 23.7% 左右的运行时间，达到 36.82dB 的峰值信噪比；对两幅图像同时进行采样的快速配准可以实现接近 2 帧每秒的处理速度，相比于原始的 SIFT 算法运行时间缩短了 64.8%，同时保证峰值信噪比为 31.21dB，远高于 SURF 算法的 20.48dB。

### 3.6. 本章小结

本章提出了一种基于 **SIFT** 算法的改进算法。针对 **SIFT** 计算量大的特点,分析了图像金字塔的特点,提出了利用下采样作为预处理的方法,从而达到了缩小图像金字塔的效果。本章首先介绍了所提算法的流程框图,对其中的图像金字塔和极值点的寻找做了介绍,研究了与算法核心单应矩阵相关的坐标系和对极几何的知识。

本文对实行下采样的图像和使用的插值方式进行了讨论,分析了不同插值方法的特点,最后采取双线性插值在 **SIFT** 源码和其所提供的测试图像上按照框图进行了实验。实验结果表明,该方法可以较大程度地缩短 **SIFT** 算法的运行时间。在与 **SURF** 算法进行比较后,本章所提出的算法甚至可以得到比 **SURF** 更好的效果。

## 第4章 基于显著性区域的高准确度图像配准算法

### 4.1. 引言

SIFT 算法除了具有运算量较大实时性较差的缺点，还因为 SIFT 的特征点检测发生在灰度图像上，如果实验图像是彩色图像就损失了色彩信息。出现在 2006 年的 CVPR 的 CSIFT<sup>[37]</sup>增加考虑了颜色信息，在颜色不变空间中构建描述子，但是大大增加了计算量。显著性区域检测算法 IG 在显著性检测时利用了图像的亮度和色彩信息，输出的显著性图像是灰度图。本章将 SIFT 的特征点检测与显著性区域检测算法 IG 结合起来，利用了图像原有的色彩信息，提高了 SIFT 配准算法的准确度。

IG 算法的作者基于显著性区域还实现了图像主体部分的分割。在图像配准和特征点的提取中，有时我们关注的对象是图像中的主体目标，这时我们不希望受到图像中背景的干扰。如果能提取得到主体部分，就可以避免提取到背景中多余的特征点。针对原始分割方法得到的分割对象不够完整，本章提出了填充图像背景的方法，通过迭代获得了更加完整的目标分割图。

### 4.2. 所提算法框架

#### 4.2.1. 基于显著性图像上的特征点检测

图 4-1 表示的是基于显著性区域检测的特征点检测和配准。框图主要分为两幅实验图像的显著性区域提取，得到对应的显著性图像；在显著性图像上进行特征点的检测；依靠特征点的匹配对得出两幅图像的单应矩阵；通过单应矩阵完成配准。这里选择了显著性区域检测算法 IG，IG 算法将带来三个优点。（1）IG 算法实行在 Lab 颜色空间上，该颜色空间与设备无关，且拥有更大的色域，保留了图像中的亮度和色彩信息，所以可以弥补 SIFT 算法本身没有考虑色彩信息的缺点。（2）IG 算法通过计算 Lab 空间上每个像素点与均值的距离作为显著性值，使用灰度值的大小表示显著性的大小，最终得到灰度图作为显著性图像。而 SIFT 算法本身是基于灰度图像的，所以可以直接在显著性图像之上进行特征点检测。（3）IG 算法得到的显著性图像与输入的原始图像具有相同的大小，相同的分辨率，所以由两幅显著性图像估计得到的单应矩阵也同样适用于原图，可以直接使用求得的单应矩阵对原始的待配准图像进行配准。

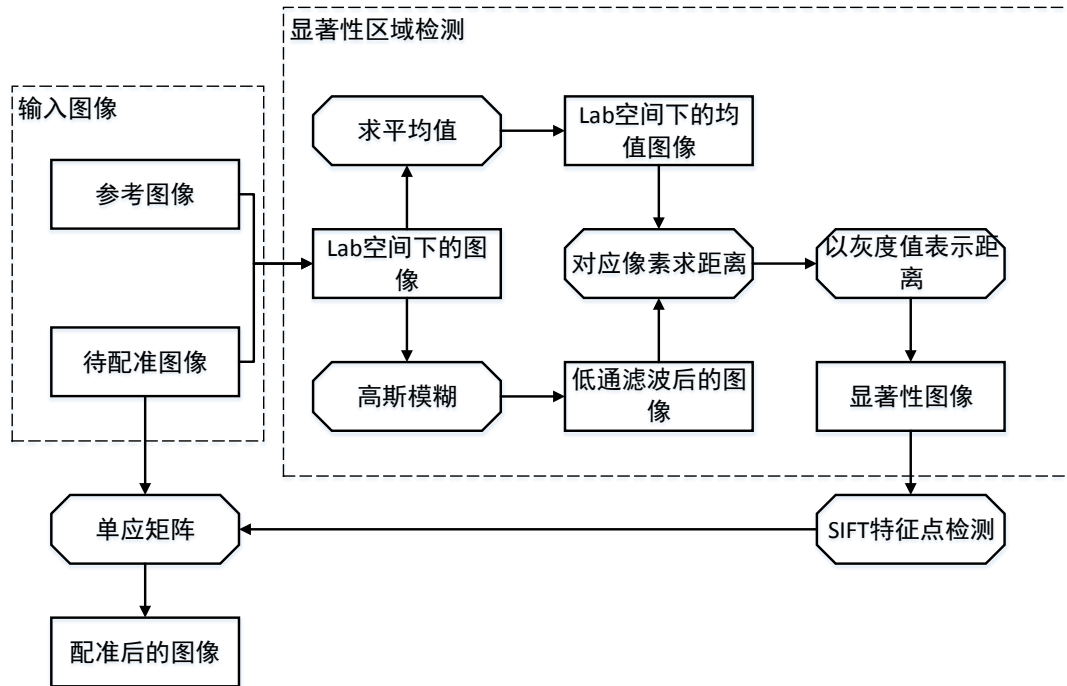


图 4-1 基于显著性图像的特征点检测

#### 4.2.2. 基于显著性图像的图片分割

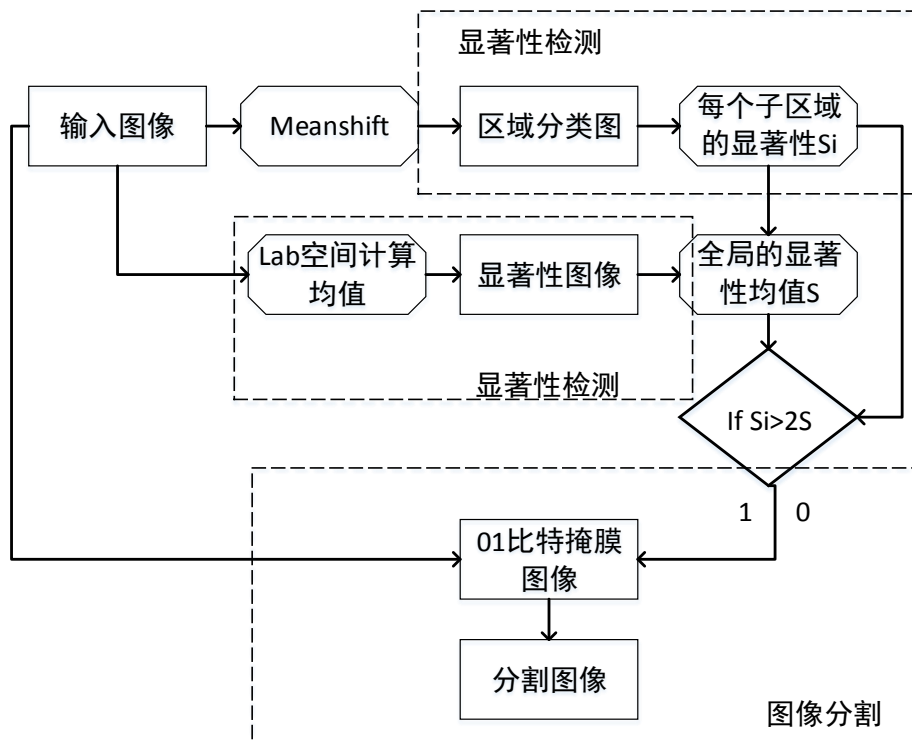


图 4-2 基于显著性图像和 Mean shift 聚类的图像分割

图 4-2 是 Achanta 实现的图像分割的流程图。实现分割的关键是按照框图中的判

断条件，比较整幅图像的显著性图像的均值的 2 倍与聚类后每个子区域的显著性均值，对于满足条件的区域进行保留就可以得到分割后的图像。所以如果可以增加主体目标区域的显著性与全局显著性均值的差距就可以更加突出主体部分。本算法使用显著性图像的灰度均值赋值给三个通道，对原图进行填充，这里选择将原图的长宽填充为原来的 2 倍。

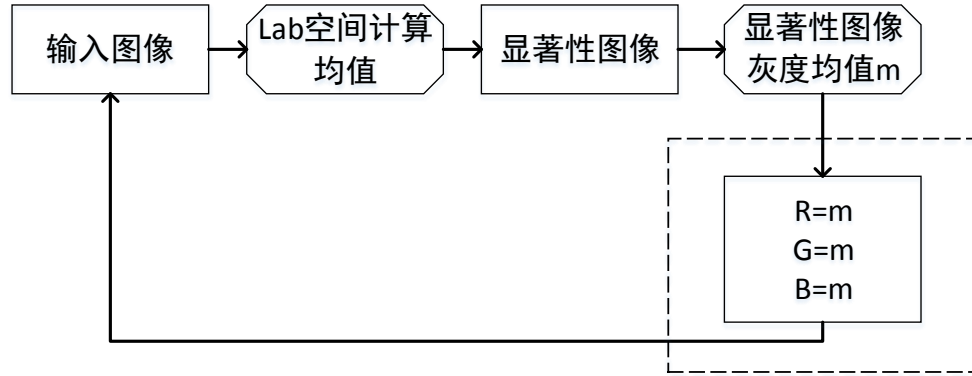


图 4-3 基于显著性图像的灰度均值对原图填充

### 4.3. 基于Lab颜色空间的显著性区域检测

人眼在观察外界时会存在一个所谓的感兴趣区域，在这个区域中的物体会更大概率地吸引人的注意力，使人获得更深刻的印象。显著性区域往往是图像中频率、色彩、对比度有明显不同的区域。这种由图像的数据驱动的视觉认知属于自下而上的注意机制，Itti<sup>[63]</sup>等人最早在人类视觉系统的神经架构和行为习惯的基础上提出了一套完整的视觉注意系统，从颜色、灰度值强度、方向三个方面提取并合并得到了显著性图。显著性检测在图像分割、自适应压缩、图像检索等领域应用广泛。

显著性区域检测得到了很多学者的关注，出现了多种检测显著性区域，将显著性大小可视化得到显著性图像的方法。如 X Hou 和上海交大的 L Zhang 提出的方法可以独立于关于目标物体的先验知识（如类别，特征），通过分析输入图像的对数谱来消除背景<sup>[64]</sup>，还可以在空域中构建显著性图像。Yun Zhai 提出了一种关于时空视频的注意点的技术。Achanta 对显著性区域检测做出了突出贡献，他在 2008 年提出了 AC 算法。AC 算法中的显著性通过像素点周围的邻域决定，具体讲是利用了局部的梯度信息，此外 AC 算法还借鉴了图像金字塔的思想，利用均值滤波得到不同尺度下的显著性值，再叠加得到最终的显著性值。在 2009 年的 CVPR 会议上，Achanta 提出了另外一种经典的显著性区域检测算法 IG，可以输出具有明确边界的显著性物体的全分辨率显著性图像<sup>[65]</sup>，具有更好的精度。

Lab 颜色模型包含 L、a、b 三个要素。L 代表的是亮度，a、b 是两个色彩范围，可以混合得到极为宽大的色彩范围，并且这个颜色模型是与设备无关的。IG 算法利

用 Lab 颜色模型，利用了颜色和亮度的特征，并且易于实现。原图中每个像素的显著性大小表现为 Lab 颜色空间上当前像素和全局平均值的距离：

$$S(x, y) = \|I_{\mu} - I_{\omega_{hc}}(x, y)\| \quad (4-1)$$

$I_{\mu}$  是图像特征均值向量,  $I_{\omega_{hc}}(x, y)$  是高斯平滑之后的像素点,  $\|\cdot\|$  表示二范数。IG 算法可分为四个部分。原始图像上的高斯模糊；获取原始图像的 Lab 模型版本；计算 Lab 向量的平均值；计算像素和平均值之间差值的 L2 范数。完整过程如图所示。

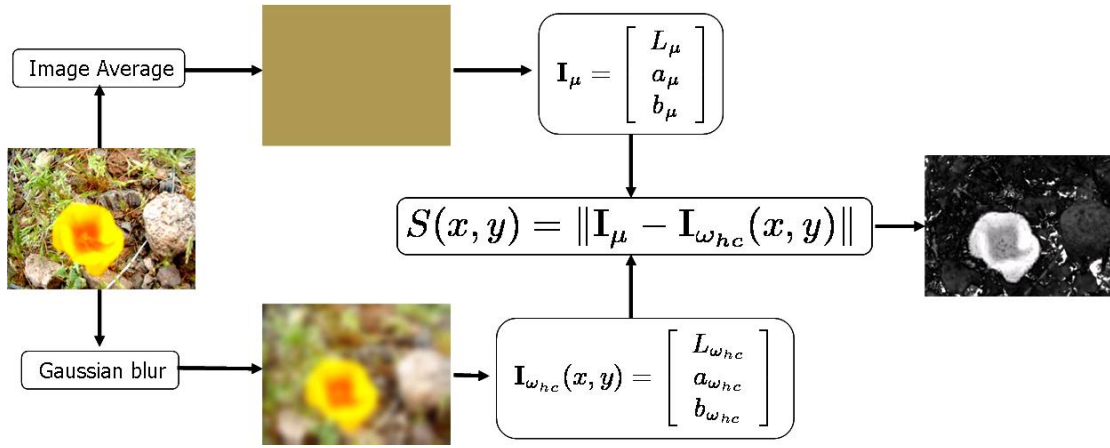


图 4-4 IG 显著性检测方法图示

经过显著性检测得到的显著性图像和原图的分辨率大小一样，以图像的形式表示了各个像素点的显著性。并且在这个过程中我们将三通道的真彩图像转换为了灰度图，而这正是 SIFT 算法所要求的，可以直接使用 SIFT 算法进行特征点的检测和图像配准。

首先，因为显著性图像与原图分辨率相同，代表显著性大小的像素点与原图中的点在空域上是一一对应的，所以根据显著性图像配准估计得到的变换矩阵同样适用于原始的两幅实验图像。其次，得益于 Lab 颜色模型的使用，显著性图像综合考虑了像素点在亮度和彩色空间的信息，在色彩和亮度上与全局的均值差距更大的像素点具有更大的显著性，这与人眼的观察习惯是类似的，从而弥补了 SIFT 算法缺失色彩信息的缺陷。

#### 4.4. 基于聚类算法的图像分割

使用简单的 K 均值（K-means）聚类算法，通过迭代，当各个像素子集中的质心不再发生更新时，我们就可以得到区域分类的图像。可以使用均值漂移（Mean Shift）来替换 K 均值聚类。均值漂移是一种不依靠先验知识的核密度估计法，根据样本的分布，通过迭代估计得到数据的密度分布函数。Mean Shift 被广泛应用在目标跟踪等



领域<sup>[66][67]</sup>。

对于  $d$  维空间, 假设存在  $n$  个随机分布的样本点, 那么在一个以  $x$  为球心, 以  $h$  为半径的高维球状区域, 我们可以得到一个 Mean Shift 向量:

$$M_h(x) = \frac{1}{k} \sum_{x_i \in S_h} (x_i - x) \quad (4-2)$$

式 (4-2) 中的  $S_h$  是以  $x$  为球心, 以  $h$  为半径的高维球状区域。用数学公式可以表示为:

$$S_h(x) = \{y : (y - x)^T (y - x) \leq h^2\} \quad (4-3)$$

各个样本点与中心点之间连线构成的向量通过矢量相加得到的向量和, 再经过计算算术平均就可以得到式 (4-2) 表示的向量。因为是求均值, 各个样本的影响因子是相同的, 为了增强算法的鲁棒性, 提高算法的泛化能力, 可以在中心点周围的圆形区域叠加一个窗口函数, 即核函数。常见的核函数有高斯函数等。

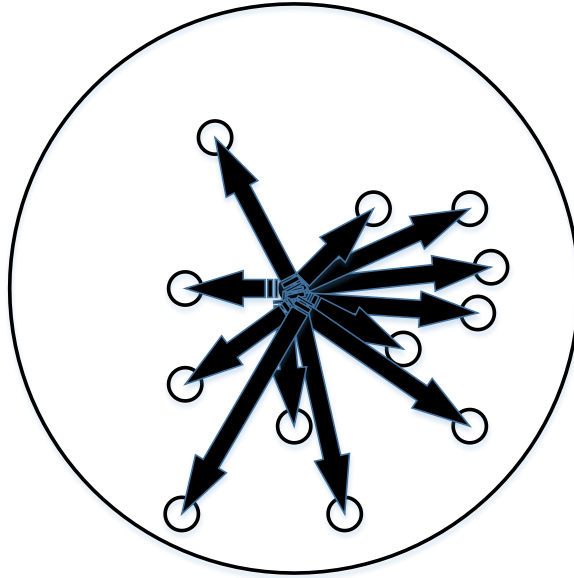


图 4-5 以二维平面为例的 Mean Shift 向量形成示意图

各个样本点与中心点相减得到的向量从中心点出发, 指向样本点, 所以 Mean Shift 向量的本质是一个偏移向量, 指向密度更大的区域方向。所以可以根据 Mean Shift 向量控制球状区域的运动方向, 使之偏移 to 密度更大的方向。

$$x' = x + M_h(x) \quad (4-4)$$

通过式 (4-4), 球状区域的位置得到更新, 在新的位置可以得到一个新的 Mean Shift 偏移向量。如此反复, 当新的偏移向量为 0 时, 各样本到中心点的向量达到平衡, 算法收敛, 此时的中心点位置即为密度最大的位置。Mean Shift 聚类上的具体流程如下表所示:

**算法表 1** 在图像上使用均值漂移聚类算法完成区域划分

**输入：**准备做显著性区域检测的图像，迭代次数 or 偏移阈值（作为终止条件）。

**算法：**

1. 在图像上随机选择一个像素点作为中心点。
2. 找出以中心点为圆心， $h$  为半径的区域内的像素点，构成集合  $M$ 。对区域内的每一个点生成一个计数器，统计其被访问的次数。
3. 对于区域内每一个像素点（样本），得到一个从中心点指向样本的向量，向量长度为像素值之间的欧式距离。向量之和的均值作为偏移向量  $shift$ 。
4. 按照式（4-4），将中心点及其圆形区域进行更新。
5. 重复步骤 2-4，直至偏移向量  $shift$  为 0 或者小于阈值，记下此时的中心点。在此之前所有访问过的样本点都认为属于一类，记为  $a$ 。
6. 重复 1-5，直至所有的点都被访问。如果存在新的类与之前的类之间的中心点的距离小于阈值，则认为二者属于同一类，可以将二者合并。
7. 样本构成的类之间可能存在重叠，可以根据之前由计算器得到的统计结果，访问次数更多的类作为当前像素点的归属类。

结合原始图像的显著性图像，可以分别得到每个子区域的显著值均值。如果平均值是原始图像的平均值的两倍，则将其视为主要对象区域。二进制对应区域为 0 和 255，因此我们得到一个二元掩码，如果我们在原始图像上施加掩码，我们最终将获得主体对象。

SIFT 算法对噪声具有鲁棒性，这是因为噪声一般是细粒度的加性噪声如椒盐噪声，且因为噪声的随机性，在两幅图像中的噪声不可能完全一样，所以在匹配的过程中可以很容易地剔除这一部分点。但是在图像配准中还有一种情形是一些杂质作为场景的一部分一直存在，而我们关注的配准对象可能只是图像中的单一主体目标，主体之外的我们可以称之为背景噪声。在这种情况下，可以将这一图像分割方法应用在图像配准中，在特征点检测阶段就从根本上剔除背景的影响，从而减少图像中背景噪声对于特征点检测的影响，降低无意义的特征点的数量。

此外，本文对于分割算法做了一些改进，主要思路是使用原图的显著性均值对图像进行填充（这里选择将原图面积扩展为原图的四倍），使得图像中主体目标与填充后得到的新图整体的显著性的值差距更大，从而有利于后续的分割。实验表明，这种改进之后的分割方法可以得到更加完整的目标，即召回率更高。因为本文以图像配准为研究对象，在分割之后以其特征点数目等为指标。

## 4.5. 实验结果与仿真

### 显著性区域检测

实验环境继续使用 DELL 计算机，在 Visual Studio 2013 下实现了 IG 算法对图像的显著性区域检测。实验图像来自作者项目主页，都是一些明显包含有主体目标的图像。图 4-6 是选取的一些实验图像及其实验结果，左边一列是三通道的实验图像，右边一列是其对应的显著性图像。



图 4-6 IG 显著性区域检测得到的显著性图像

从图 4-6 可以看到，原图在经过显著性检测之后转换为了显著性图，原图中像素点的显著性大小通过显著性图中的灰度值表示，灰度值越大，亮度越高说明该点在整幅图像中更突出，原图中的主体目标从而得到了突出表示。

### 显著性图像上的图像配准

下面是将显著性检测加于特征点检测的实验结果。(a)(b)分别是来自 Rob Hess 提供的配准实验图像。(c)(d)分别是直接进行特征点检测与匹配和在显著性图上的效果。

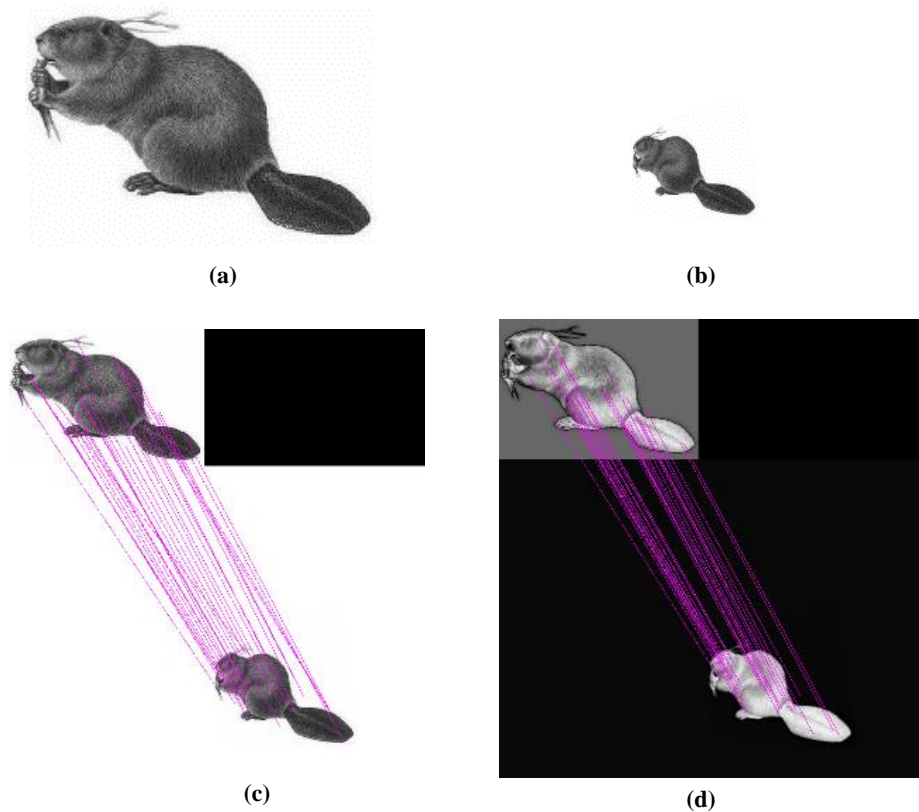


图 4-7 IG 显著性图像上的特征点检测与匹配

均方根误差  $RMSE$  是一种衡量精度的指标, 采用两幅图像所有匹配点(变换后坐标与参考图像对应点)坐标的均方根误差  $RMSE$  来衡量。

$$RMSE = \sqrt{\frac{1}{n} \sum \|(x'_i, y'_i) - f(x_i, y_i)\|^2} \quad (4-5)$$

式 (4-5) 中,  $(x', y')$  为参考图像上的点的坐标,  $(x, y)$  为待配准图像点的坐标,  $f$  表示图像间的变换关系,  $n$  为最终的匹配点数目。  $RMSE$  的形式与标准差类似, 不同的是均方根误差每次比较的是数值对, 而标准差为了表示数据本身的离散程度, 每次以均值作为比较对象。均方根误差数值越小, 则表示精度越高, 效果越好。

表 4-1 是图 4-6 中特征点的数目、运行时间和配准效果的统计结果。N1 和 N2 分别表示从两幅图中提取出的特征点数目。评价标准选取了运行时间  $Time$ 、峰值信噪比  $PSNR$  和均方根误差  $RMSE$ 。

表 4-1 两种配准方法的性能对比

	N1	N2	Time/ms	PSNR /dB	RMSE
原始图像	116	95	1597	38.719	18.381
显著性图像	353	167	2140	43.920	0.556

从实验结果可以看出，基于显著性图像的 SIFT 检测获得了更多的特征点，实现的配准结果也提高了约 5dB，而 RMSE 从 18.381 减少到 0.556。

### 图像分割与特征点检测

下面是图像分割的实验结果，实验环境与实验图像与本章的上一小节相同。

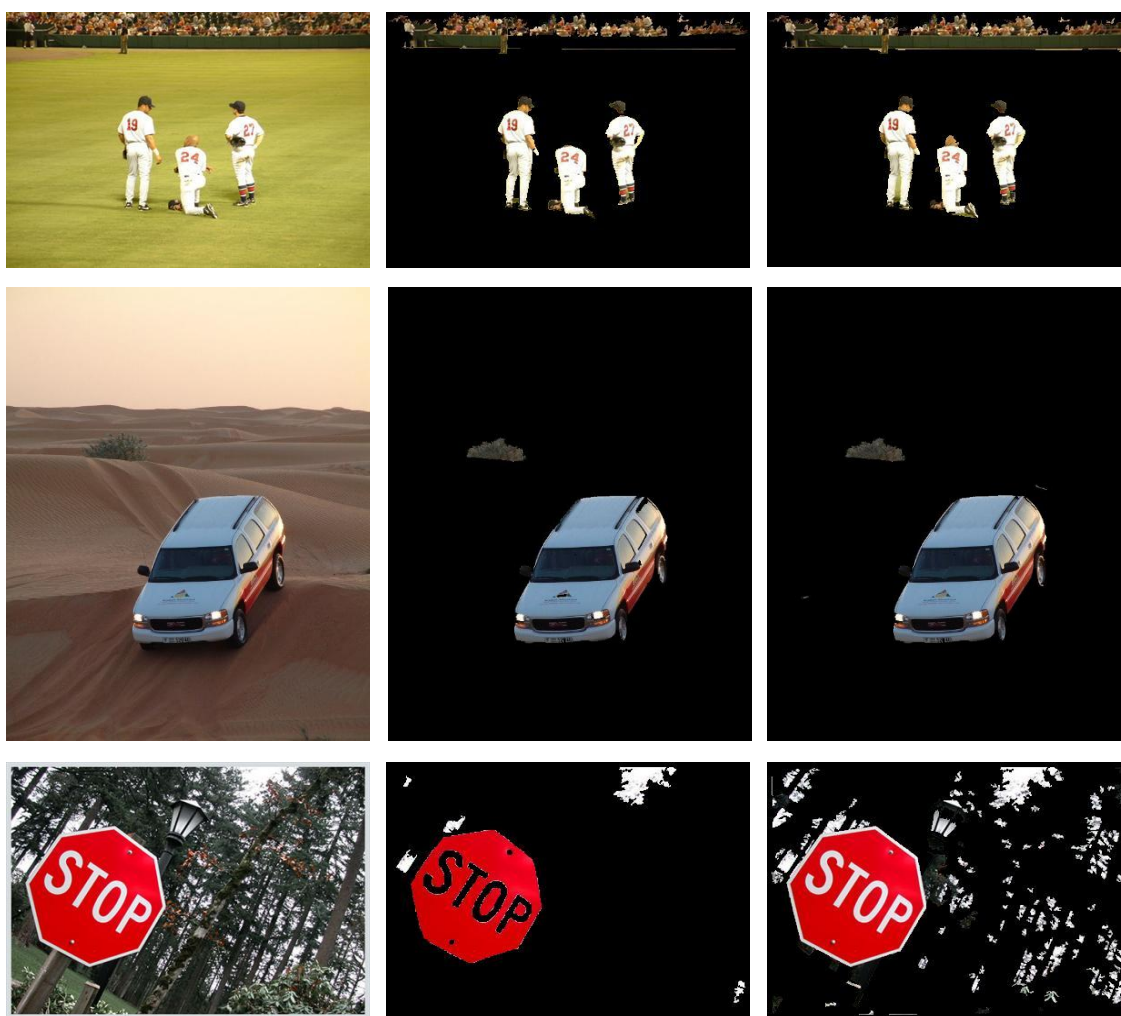


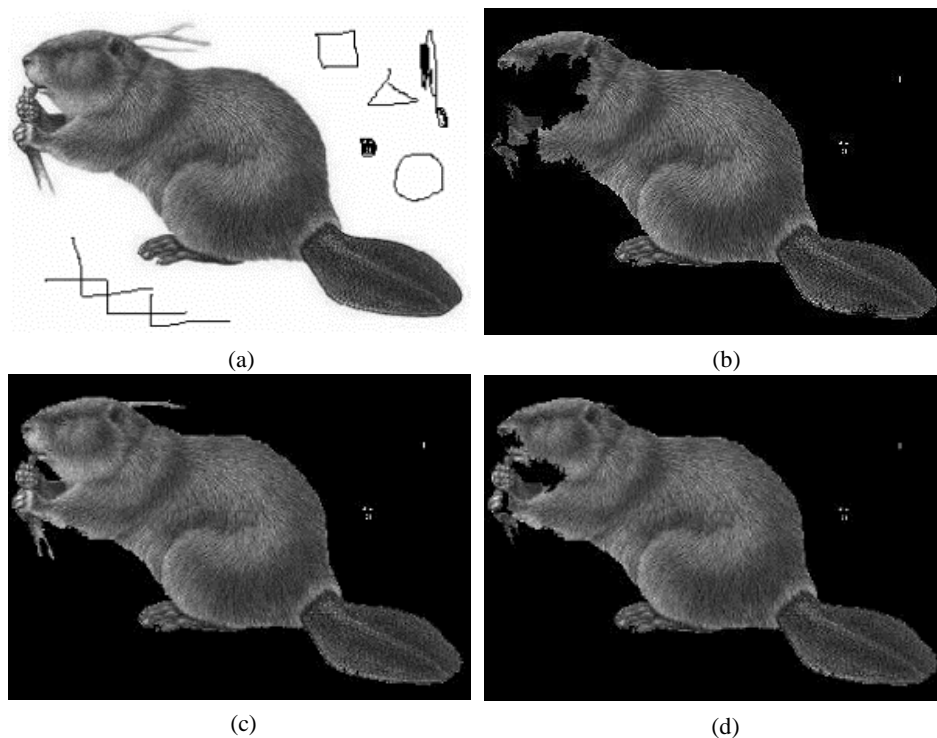
图 4-8 改进前后的图像分割方法对比

上图中第一列是实验图像，第二列是使用作者的方法得到的分割结果，第三列是改进之后的图像分割结果。可以看到填充背景后的分段保留了更多细节。与原始方法相比，修改方法的结果更加完整。对于第一行，本算法中远处的观众席和跪姿运动员的头部得到了保留；对于第二行图像，本算法额外保留了汽车车尾部分的车窗；对于第三行图像，本算法相比于原始算法保留了全部的指示牌和路灯，还有树林中露出的天空部分。

利用人工涂鸦的图像可以说明分割后特征点检测的优势。图 4-9 是实验结果，分



别对涂鸦图像直接分割，以白色纯色填充和以灰度均值填充，然后进行特征点检测。其中第一幅涂鸦图像是本人在 Rob Hess 提供的实验图像上的人工改动。



(a) 人工创建涂鸦背景的图像 (b) 直接分割的效果

(c) 用白色像素填充背景后的分割效果 (d) 分割后效果，背景填充了平均值

图 4-9 三种分割方法得到的分割图像对比

可以看出，几种分割方法都很明显地去除了图像背景中的随机噪声的干扰，从主体目标的完整性上看，原始的分割方法效果最差，丢失了很多的目标信息。表现最好的是使用纯白色填充扩展的背景范围，但是考虑到实际中的图像背景较复杂，很少会出现理想化的纯色背景，所以更具有普适性的方法是使用原图的均值进行填充。下表是针对分割图在特征点检测方面的统计结果。

表 4-2 涂鸦图像使用分割方法后的特征点检测和配准效果

	(a)	(b)	(c)	(d)
<b>N</b>	144	116	124	117
<b>RMSE</b>	16.483	1.476	12.651	6.460
<b>Time/ms</b>	1849	1615	2258	1622

从表中可以看出,相比于原始的涂鸦图像,借助于显著性图像的分割可以消除背景中无关信息的干扰,表现在特征点数量上就是减少了特征点数量。同时,由均方根误差可以看出,在准确度上也得到了改善。

#### 4.6. 本章小结

本章的主要内容是将显著性检测与 SIFT 特征点检测与配准进行了结合。首先介绍了显著性区域检测研究的重要性,简单介绍了几种显著性区域的检测方法。然后主要介绍了其中的将要使用的 IG 显著性区域检测方法,并对其进行了仿真实验,验证了该算法可以高效而准确地得到显著性图像。然后提出了在显著性图像基础上特征点检测的和匹配的算法,实验表明,该方法提取出了更多高质量的特征点,有利于充分拟合出变换模型。此外,对基于显著性区域的分割算法进行了改进,从而在检测特征点的时候可以去除背景中杂质点的干扰。





## 第5章 图像配准在图像压缩中的应用

### 5.1. 引言

正如前文中提到的，图像配准是其他数字图像处理领域中重要的步骤，比如图像拼接和图像融合，主要步骤都是图像配准，通过图像配准建立两幅或者多幅图像在空间上的联系。图像拼接应用在无人机航拍、手机摄影等领域，可以获得大尺寸大视野的图像；图像融合则是将不同类型图像叠加，将信息互补，融合之后的图像可以提供更多的信息。

图像融合和图像拼接都是在图像上做加法，目的是在空域或者其他维度增加图像的信息。导师周诤等人在发明专利“一种基于图像集的数据压缩方法与系统”中提出了一种图像压缩算法<sup>[68]</sup>，该算法利用待压缩图像与图像数据集中的参考图像之间的相似性，通过图像间运算消除更多的冗余信息，得到更加稀疏的矩阵。该算法在图像运算前需要对图像进行配准，本文提出的基于显著性图像的配准算法提升了算法的准确度，可以将其作为预处理步骤应用在图像压缩算法中。

### 5.2. 图像集压缩中的图像配准算法框架

随着数字化的发展，我们不再需要昂贵的底片就可以拍摄得到图像，拍摄成本的降低也导致了拍摄的随意性，比如在我们的手机相册中往往会出现很多同一场景下的图像，这些图像之间可能是由于拍摄的角度不同产生了些许差异，很明显，这些图像之间是合适于做图像配准的。两幅图像的重合度或者说相似性可以利用差异值哈希算法来衡量，先分别计算两幅图像各自的哈希值，然后通过两个哈希值之间的汉明距离表示两幅图像的相似度。满足条件的图像就可以进行图像配准，配准后的图像与参考图像就满足尺寸相同的特点，而且这两幅图像之间有相当一部分是重合的，配准的性能越好，重合度也就越高。利用图像减法可以去除冗余的重合部分，只记录差异的部分，从而减少信息的存储量，同时在需要原有图像时，也可以复原图像。但是需要注意的是，像素之间的差值可能为正也可能为负，为了将得到的差值转换为一幅新的差值图像，需要取差值的绝对值，同时记录每个差值的正负性，形成一个新的掩膜图像，便于日后对于原图的复原。对于差值图像，如果配准后的图像与参考图像在像素级别重合度越高，差值图像中像素灰度值为0的部分会显著增多，即新的矩阵会变成一个稀疏矩阵，有利于做图像压缩，进一步减少图像占用的空间。

本章提出的图像集压缩中的图像配准算法框架如图5-1所示。该配准算法输入的两幅图像中，一幅来自图像集，作为参考图像；另一幅图像是待压缩图像，可看作参

考图像变形之后得到的，作为待配准图像。算法可分为三部分：第一部分首先通过图像哈希算法计算得到两幅图像各自的哈希值，作为“图像指纹”，比较两个哈希值之间的汉明距离，如果汉明距离小于设定的阈值就可以认为两幅图像之间存在着某种变换关系，即关于单应矩阵的方程组是有解的，且解是有意义的。汉明距离满足条件之后就可以计算得到两幅图像各自对应的显著性图像，通过在显著性图像上提取特征点完成图像配准，这是框架的第二部分。第三部分是对配准之后的图像与参考图像做差值。配准后的图像与原图之间的差值会是稀疏矩阵，有利于进一步压缩。配准的效果越好，两幅图像的相似度越高，两幅图像对应像素点做减法得到的图像对应的矩阵会是稀疏矩阵，即图像中存在着大量像素值为 0 或者接近于 0 的像素点。而稀疏矩阵因为其独特的性质合适于做图像压缩，保证无损压缩的条件下，可以得到更大的压缩比。

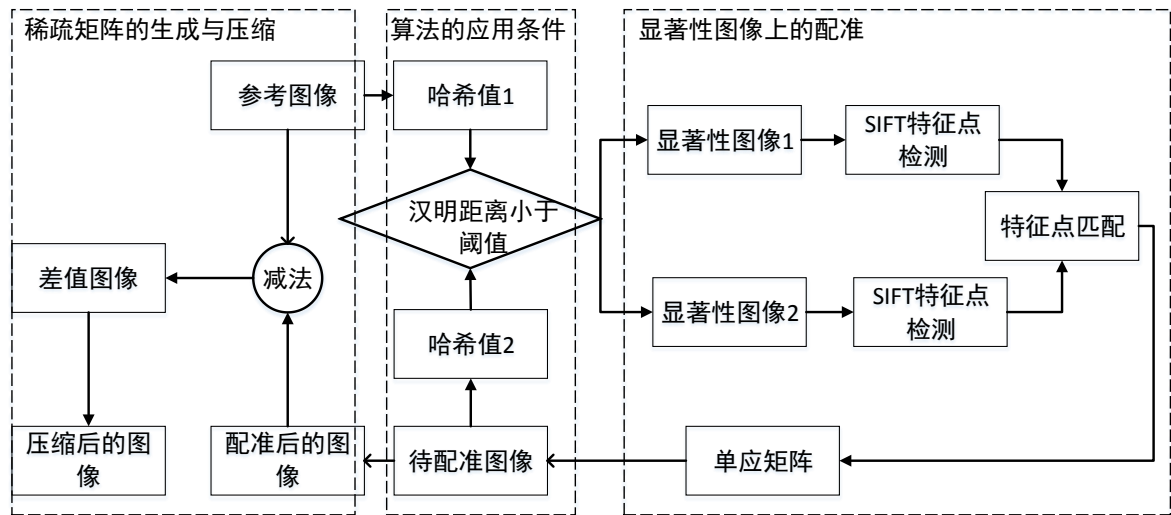


图 5-1 基于图像配准的图像压缩算法流程

### 5.3. 基于哈希算法选择可压缩的图像对

本章所提出的图像压缩算法的基础是两幅图像之间具有空域上的信息冗余，所以算法关键的第一步是找到包含冗余信息的两幅或者多幅图像。满足这一条件的图像之间可以通过图像配准拟合出二者之间的变换关系而对于在内容信息上不相关的图像则不适用采用本章的压缩算法。可以通过图像的哈希算法筛选出适用于本算法的图像，按照图 5-1 中的后续步骤进行压缩，对于不适用于本压缩算法的图像则予以略过，维持其现状而不进行多余的处理。

哈希算法是一种对图像建立“图像指纹”的算法，被广泛应用于目标检测、图像检索等领域<sup>[68]</sup>。哈希算法又叫散列算法，本质上就是一种映射关系，通过约定的映射函数将原文本映射为较短的固定长度的二进制值，相当于将图像转换到了低维的特征

空间中。加密哈希算法是其中一种哈希算法，包括 MD5（Message Digest Algorithm MD5）、SHA1、SHA256 等。加密哈希算法不可根据哈希值反向推出原信息，正是由于加密哈希的这一不可逆性，所以被广泛应用于门户网站加密，文本加密，口令加密，数字签名等领域。

加密哈希还有一个特点就是文件一旦发生哪怕微小的改变，哈希值都会发生巨大的变化，而我们希望相似图像具有相同或者相似的哈希值，这时候我们就需要采用其他的哈希算法。最常见的方案是采用感知哈希算法（Perceptual Hash Algorithm）。感知哈希算法首先对图像进行缩放，在低频上对图像进行感知，所以计算出的指纹序列对图像形变的包容度更高，对一些简单的图像变换存在一定的鲁棒性。同时，感知哈希算法计算简单，可以比较简单地得到图像的指纹。感知哈希算法是一类算法的总称，包括平均值哈希、差异值哈希。

平均值哈希方法计算简单：首先对图像进行尺寸上的归一化，计算得到归一化后图像像素灰度值的平均值；比较每个像素与整体灰度图像的灰度均值，条件判断后就可以得到 0、1 的二进制字符串。平均哈希算法思路简单，但准确度不高，因为它只记录了每个像素与均值的大小情况，像素之间的大小关系不明确，丢失了过多的信息。

$$h = \begin{cases} 1, f(x, y) > \frac{1}{M \times N} \sum f(x, y) \\ 0, else \end{cases} \quad (5-1)$$

平均值哈希会受到伽马校正或者直方图均衡的影响，而差异值哈希是一种更健壮的算法。差异值哈希与平均值哈希大体相同，差异值哈希比较的是同一行相邻的两个像素的灰度值，得到 0、1 比特。

$$h = \begin{cases} 1, f(x, y) > f(x+1, y) \\ 0, else \end{cases} \quad (5-2)$$

算法第一步是归一化，目的是得到相同长度的哈希值作为图像指纹，方便之后的比较和搜索。在平均值哈希算法中，为了得到 01 字符串，每个像素的比较对象是全局的平均值，所以 01 比特是与缩略图的像素一一对应的。当归一化之后的缩略图大小是  $8 \times 8$  时，按行拼接之后得到长度为 16 的十六进制字符串，转换成二进制就是长度为 64 的二进制字符串。在差异值哈希算法中，比较的对象是同一行相邻列的两个像素，所以为了得到相同长度的字符串，缩略图采用的大小是  $8 \times 9$  的大小。除了归一化最终的哈希值长度，缩放图像的另外一个目的是对原图模糊处理，避免过多图像细节和噪声的影响，从更加宏观和整体的角度比较图像，同时较少计算量。

下表是利用差异值哈希算法在图像集中查找得到适用于本算法的步骤。哈希值之间的汉明距离小于阈值的图像被认为是适于图像配准的图像，可以利用两幅图像之间

的重复信息实现图像压缩。

---

**算法表 2** 在图像数据集中查找相似图像的算法

---

**输入：** 图像数据集，其中存在着包含相同场景的图像；汉明距离的阈值。

**算法：**

1. 图像归一化到 $8 \times 9$ 的大小。
  2. 如果原图是三通道的真彩图像，将缩略图转换为灰度图像。
  3. 对每幅灰度缩略图，比较每一行的相邻像素，左边的像素灰度值大于右边相邻的，则将对应该位置记为 1，否则为 0。得到一幅二值图像。
  4. 将二值图像中的 01 像素值按行拼接，得到二进制字符串。
  5. 按照搜索策略，比较图像集中图像对应哈希值之间的汉明距离，如果小于阈值，则认为二者相似度较高，符合本算法的执行条件。
- 

下面是两个图像进行差异值哈希算法的示意图。两幅图像(a)、(c)尺寸不同，在计算差异值哈希值时会被归一化为相同的大小的缩略图，通过比较同一行相邻的像素值大小得到相同长度的二进制的字符串。

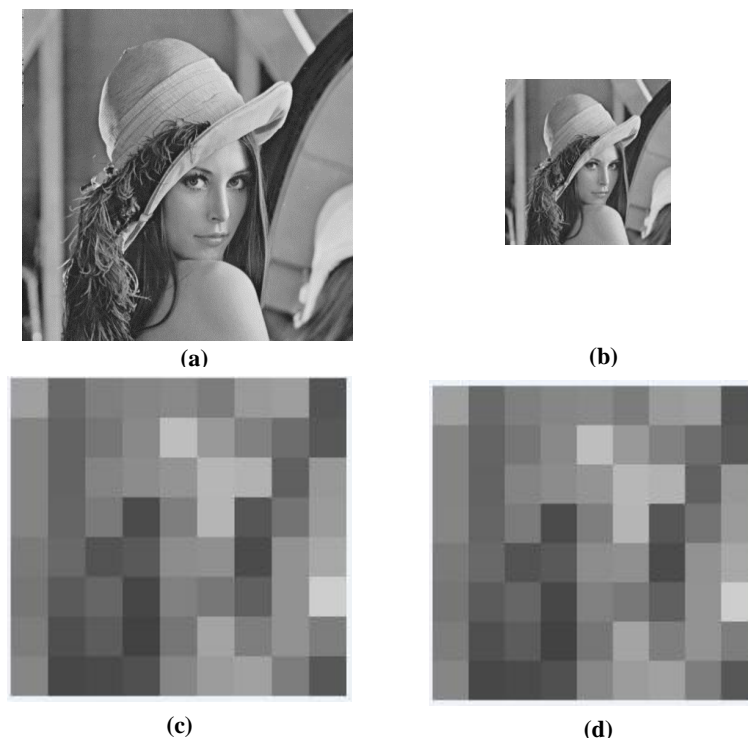


图 5-1 差异值哈希算法示意图

在获得了图像的指纹序列之后，还需要确定指纹之间的相似性测度，这里考虑到指纹序列是二进制的字符串，所以选择的是汉明距离。在信息论和密码学中，汉明距离是一种常用的距离和相似度的衡量方式，比较两个相同长度的二进制字符串对应字符，记录不相同的字符的总数作为汉明距离的值。在计算机中，汉明距离更加具有高效性，通过简单的异或运算就可以得到汉明距离的值。图 5-1 中的两幅图像的哈希值用十六进制表示为'898d82a446eca583'和'898d82a446eca583'，求其二者的汉明距离为 0。显然，二者的相似度很高，可以认为是同一幅图像。

此外，搜索策略也是重要的内容，直接关系到算法的效率。可以使用 KD 树或者 VP 树等优化算法进行加速，这里使用最简单的穷举策略进行搜索。

#### 5.4. 基于图像间冗余信息的图像压缩

图像在计算机中以矩阵（阵列）的形式表示。图像间的阵列操作一般以像素为基础，即矩阵中的元素为运算单元。阵列操作又可以分为算术操作和逻辑操作。逻辑操作一般是集合间的交、并、补，可以衍生出一些形态学的操作。

算术操作一般分为加减乘除几种类型：

$$\begin{aligned} s(x, y) &= f(x, y) + g(x, y) \\ d(x, y) &= f(x, y) - g(x, y) \\ p(x, y) &= f(x, y) \times g(x, y) \\ v(x, y) &= f(x, y) \div g(x, y) \end{aligned} \quad (5-3)$$

其中，图像之间的加法可以用在对带有噪声的图像的降噪中。对于包含加性噪声的图像，对二者进行像素间的加法，然后求其均值可以一定程度上达到降噪的效果。这是因为噪声在不同图像间的分布是随机的，而噪声一般是均值为 0 的高斯白噪声，所以可以通过求均值减少图像中的噪声。可以证明，随着图像数量的增加，可以减小均值图像的方差。这种降噪方法常常使用在天文领域。相加的连续形式是积分，所以甚至可以使用 CCD 等传感器实现感知图像的叠加，达到对图像噪声的消除。

图像减法是另外一种基础的图像间的运算。减法操作的对象是在像素级别上相似度较高的图像，有时候甚至无法通过人眼观察出二者的区别，通过图像间的减法可以突出二者之间的差异性。图像减法可以应用在医学图像的造影中，经过造影剂作用的活体图像与模板图像的减法结果在图像增强之后可以得到期望的显示特定信息的图像。对于相似度很高的两幅图像，经过图像减法，阵列中相同或者相似的元素将互相抵消，在差值图像中对应的点就会是接近 0 值的点。图像乘法和除法则可以应用在图

像三维阴影的校正和感兴趣区域（Region Of Interest, ROI）中。

几种图像之间的算术操作不仅要求图像之间在大小上是相同的，还都要求实验图像是已配准对齐的。在我们的这个应用场景下，配准之后的图像与参考图像在像素级别上具有高度的相似性，且随着配准精确度的提高，配准后图像与参考图像的相似度越高，二者经过图像减法之后的差值图像将具有整体图像偏暗的特点，在灰度直方图上表现为灰度值 0 附近的直方图较高。这是因为在差值图像中，实验图像的差异性表现为灰度值的大小，图像像素的灰度值越大，则说明实验图像在对应位置的差异越大。也可以将差异性理解为是去除冗余信息之后的数据。

下面的(a)、(c)图像来自比利时鲁汶大学（Katholieke Universiteit Leuven）。该学校为了研究仿射不变性的特征，建立了专门的数据集。图 5-2 像来自其中的 bikes 数据集，两幅图像(a)(b)之间不仅有拍摄角度的变化（注意摩托车车胎距离地面的距离），还有模糊变化。(c)是(a)、(b)做图像减法得到的差值图像。在差值图像中，差异信息表现为像素点的灰度值，灰度值越大，则表明此处的差异信息越大。图(c)整体偏暗，说明大部分像素点的灰度值较低，同时说明经过图像减法去除了较多的冗余信息。



图 5-2 图像减法(a)-(b)=(c)

我们还可以从直方图的角度看出图像减法前后的像素值的分布情况。图 5-3 中第一行是图 5-2 中对应图像在三个通道上的直方图表示，第二行是其对应的灰度直方图的表示。

可以看到相似图像在经过图像减法之后，在三个通道和整体的灰度上的直方图都被均衡在了幅值较低的区域。这说明差值图像对应的矩阵具有稀疏性，而稀疏的矩阵可以得到更简单的表示，同样是无损压缩就可以得到更大的压缩比。

稀疏化的图像具有编码冗余，具有更大的压缩空间。事实上，可以把图像看作是一种携带有信息的信号表示，稀疏图像则是一种稀疏信号。利用信号的稀疏性可以实现对信号的压缩感知（Compressed Sensing, CS），可以在低于奈奎斯特采样频率时获得对信号的无失真或者低失真的重建<sup>[70]</sup>。压缩感知的关键就是将信号进行稀疏表示。天津大学的王浩等利用多级离散小波变换得到了图像的均匀化稀疏表示<sup>[71]</sup>。压缩感



知可以用式 (5-4) 表示:

$$y = \phi x \quad (5-4)$$

其中  $x$  表示原始信号, 要求在某个变换域中具有稀疏性。  $\phi$  表示测量矩阵, 也称为随机投影矩阵, 描述对原始信号的采集过程, 为了得到更有代表性的有效信息, 应该具有一定的不相关性。  $y$  表示降维之后的数据, 通过重建算法可以无失真或者低失真地恢复原始信号  $x$ 。

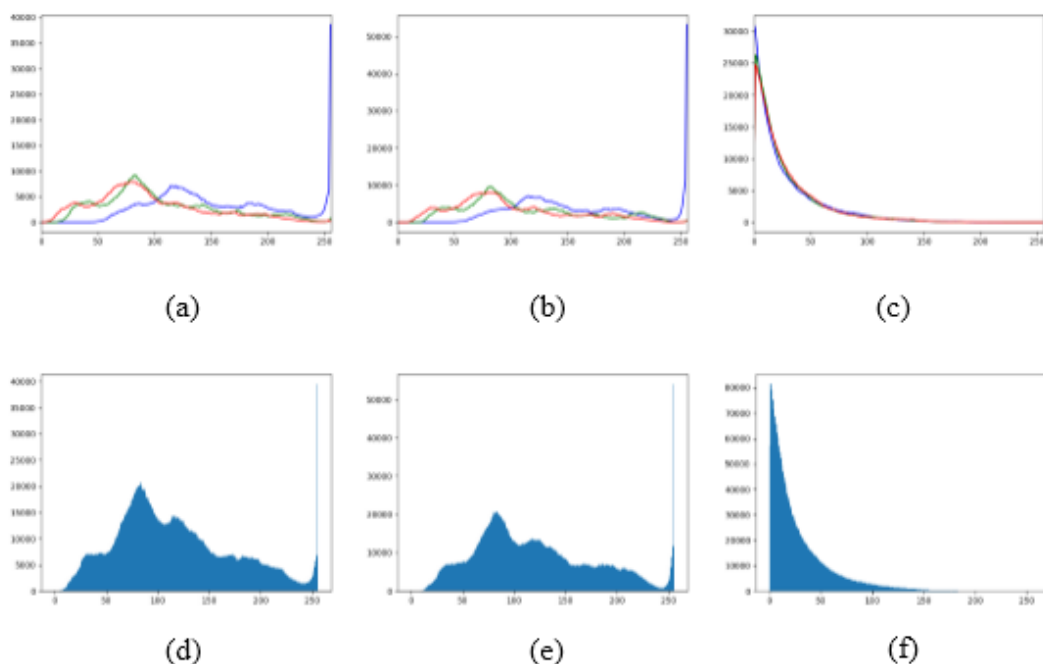


图 5-3 原图和差值图像的彩色和灰度直方图

压缩算法是一种减少数据量的算法, 在声音, 视频等多媒体介质和传输中应用广泛。图像压缩是一种在尽量保证图像信息不受损失前提下, 减少图像数据量的算法。图像压缩一般建立在编码冗余、像素间冗余和心理视觉冗余上。编码冗余是指编码技术没有使码元达到理论上可以表示的最大信息量, 最佳编码应该使得每个像素所需的平均比特数最少; 像素间冗余是由像素间相关性引起的, 包括空间和时间(视频序列中)相关性, 即不同像素表示的信息之间有重合; 心理视觉冗余是指由于人眼的局限性, 无法察觉一些失真和变化, 那么这一部分就可以视作是冗余的。

图像压缩系统一般包括两部分: 编码器和解码器, 分别完成压缩和解压缩这两个互补的操作。图像压缩是数字图像处理技术中用处最广泛和商业上最成功的技术之一, 商业上的成功反过来推动了技术的演变。形成了一些著名的图像压缩上的工业和商业标准(也被称为图像格式、容器), 规定了数据的排列方式和压缩类型及压缩过程。常见的标准有 JPEG、JPEG2000、BMP、GIF、PNG 等。其根据是否可以完全恢复原

图像又可分为无损压缩、有损压缩。

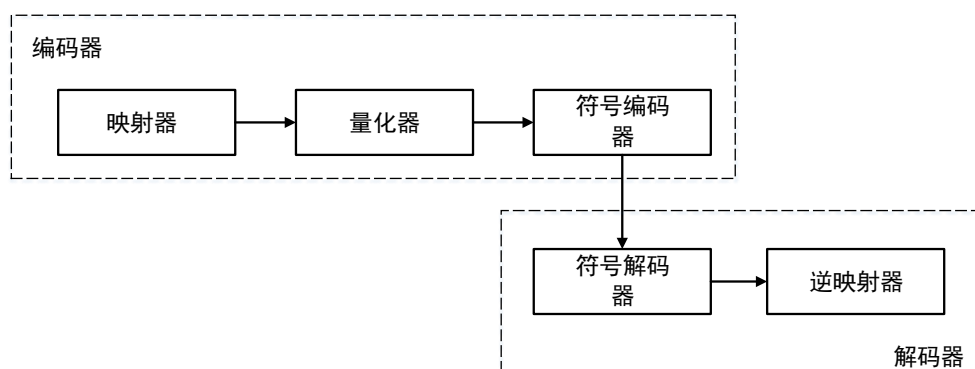


图 5-4 图像压缩流程

常用的图像压缩算法有哈夫曼编码，LZW 编码、行程编码、块变换编码、预测编码、小波编码等。哈夫曼编码是一种变长编码，根据信源出现的概率大小，使用不同长度的码字进行编码。同样属于变长编码的还有 Golomb 编码，比霍夫曼编码还要简单。LZW 编码则致力于通过消除空间冗余来实现图像压缩。空间冗余正如之前提到的，是指一些元素及其组合会出现重复。LZW 顺序地分析像素，通过对出现的字符建立字典，可以把之后重复出现的像素进行简单表示。LZW 是 PNG、GIF、PDF、TIFF 等压缩格式的主要算法。对于大概率连续出现重复元素的信息可以使用行程编码。行程编码记录了第一个出现的元素及其之后连续出现的相同元素的个数，所以更适合于二值图像，因为二值图像的每个像素只有两个取值。行程编码是 BMP 等文件格式的主要算法，但行程编码的缺点不能进一步在行程编码之后的数据中继续使用行程编码。

哈夫曼编码以哈夫曼树为基础。式 (5-5) 表示的是带权路径总长度  $WPL$ 。式中  $W_i$  表示每个叶子结点的权重， $L_i$  表示每个叶子结点到根结点的路径长度，叶子结点的总数为  $n$ 。二叉树中的叶结点的权重值与叶节点到根结点的乘积的和作为树的带权路径总长度，可以证明哈夫曼树中带权路径总长度是最短的，在这个意义上哈夫曼树被称为最优二叉树。

$$WPL = (W_1 * L_1 + W_2 * L_2 + \dots + W_n * L_n) = \sum_{i=1}^n W_i * L_i \quad (5-5)$$

哈夫曼编码已成为 JPEG 等压缩标准中常用的编码算法。本章的压缩算法也以含编码为主要压缩方式，下表展示了哈夫曼编码的基本流程。



**算法表 3** 哈夫曼编码步骤

**输入：**信源符号（图像中的像素点）及其出现的概率统计

**算法：**

1. 将信源符号按照概率大小升序排列，最为初始的叶子结点。
2. 选择其中概率最小的两个信源符号，构成一个新的二叉树。保证概率较小的符号处于左二叉树，较大概率的符号处于右二叉树，并将左右分别用 0、1 标记。
3. 将步骤 2 中形成的根结点作为新的叶子结点，其概率为两个叶子结点的概率之和。
4. 重复步骤 1-3，直至不再产生新的二叉树。
5. 从最后的根结点向每个初始的叶子结点回溯，根据之前标记的区分左右子树的标记可以得到每个字符的二进制编码。

哈夫曼编码的缺陷是必须在编码前获得信源符号的先验概率统计。所以这里将哈夫曼编码与一种更为通用的数据流压缩方法 LZ77 结合起来。LZ77 算法<sup>[72]</sup>是一种基于字典的无损压缩算法，由以色列学者 Lempel 和 Ziv 在 1977 年提出，被广泛应用在各自压缩工具中。LZ77 的主要思想是使用滑动窗口找到数据中与字典短语最大匹配的字符串，利用匹配的长度和二者的相对距离与原始信息进行映射，达到压缩的目的。

将哈夫曼编码和 LZ77 编码结合起来可以实现 deflate 压缩算法。deflate 压缩算法具有三种压缩模型，可以根据数据的特点进行选择。

## 5.5. 实验结果与分析

这里的实验环境和上文保持一致，涉及到图像配准和哈希算法的部分使用了 Python 3.6 语言编程，软件环境为 Pycharm Community 2017。特征点的检测和匹配算法使用了第四章中基于显著性图像的 SIFT 算法。变换模型选取了包含仿射变换的透视投影模型。具体的矩阵形式是单应矩阵。图像的减法使用 Visual Studio 2013 编程实现。图像压缩使用 zlib 函数库，这是一个免费高效、通用且跨平台的数据压缩库，可以提供无损的数据压缩。压缩参数使用 zlib 模块的缺省值。实验中使用了 Python 中的 PIL 模块来读取图像，将图像转换为字符串来计算图像占据的字节大小。

实验图像继续使用鲁汶大学特征检测项目组提供的实拍图像。实验分为两组，分别是 bike 图像组和 graf 图像组。

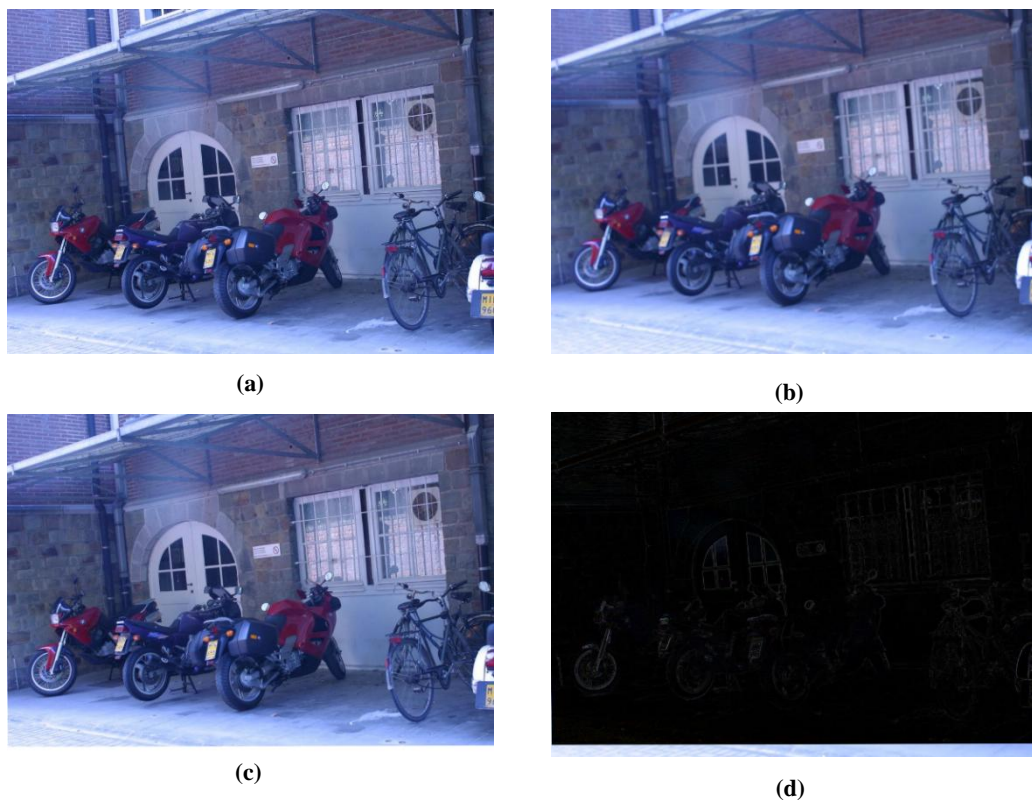


图 5-5 两种减法得到的图像灰度直方图 (bike 图)

图 5-5 中的(a)的 64 比特指纹用十六进制表示为'bcd13d11519bdabe', (b)的指纹为'dcd19393cb9b49cf'。可以很容易计算得到二者的汉明距离  $d = 11$ 。这意味着两幅图像的 64 比特的二进制字符串中只有 11 位比特是不同的, 即在缩略图上的比较结果是大部分相邻像素的大小关系都是一致的。可以说明二者的相似性程度较高, 可以使用基于配准的方法实现图像压缩。

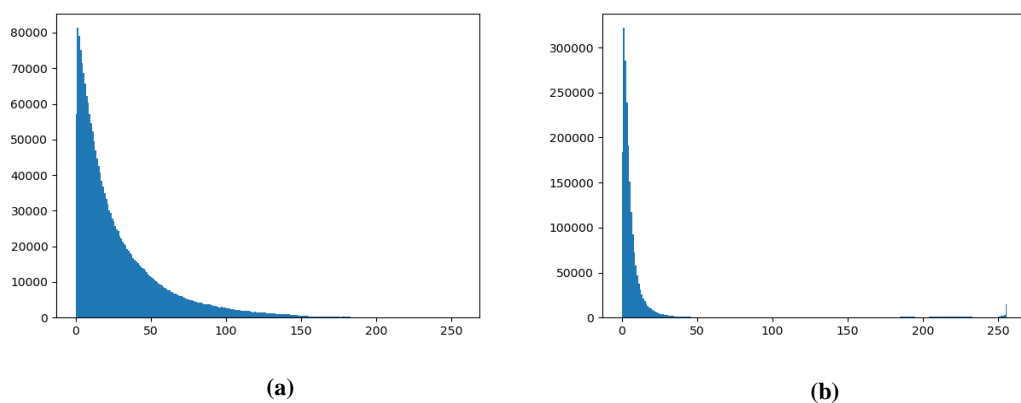


图 5-6 配准后的图像及其与参考图像的差值 (bike 图)

在配准过程中，将图 5-5 中的(a)看作待配准图像，(b)看作参考图像，得到配准后的图像(c)。(d)表示参考图与待配准图像的差值。可以将两幅实验图像直接做减法得到图像的直方图与配准后图像做减法得到的直方图做比较。图 5-6 中，(a)是两幅实验图像直接相减得到的直方图，(b)是配准后图像与参考图像做差值得到图像的灰度直方图。

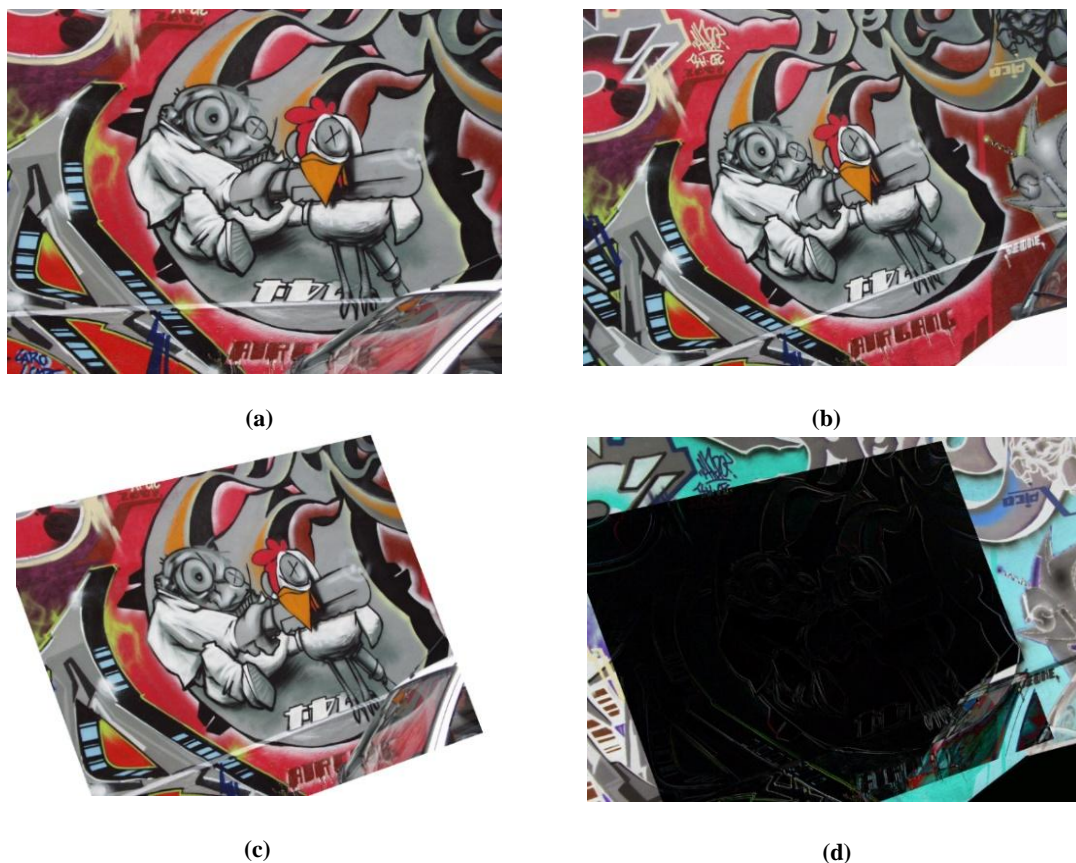


图 5-7 配准后的图像及其与参考图像的差值（graf 图）

图 5-7 是使用 graf 图像得到的实验结果。两个图像哈希值之间的汉明距离是 15，而哈希值是长度为 64 的二进制字符串，说明有 80%的内容是相近的，所以可以使用本章的算法对图 5-7 中的(b)进行压缩。图 5-7 中的(c)是(a)作为待配准图像，(b)作为参考图像得到的配准后图像；图(d)是图(c)与图(b)做图像减法得到的差值图像，为了便于展示，这里取差值的绝对值作为对应像素的灰度值。

下面的表 5-1 记录了两组实验中更多的数据，记录了原始图像的大小，直接对原始图像进行压缩得到的图像大小，和以本章算法进行压缩之后得到的图像大小。图像占据存储空间的大小以字节为单位。

表 5-1 图像配准和减法前后的压缩情况比较

数据	原始图像大小/Byte	压缩原始图像/Byte	压缩差值图像/Byte
Bike	2100000	1261957	1047757
Graf	1734000	1389097	1149562

可以看到,在保证无损压缩的前提下,本章的算法可以得到更大的压缩比。在 Bike 图像组合 Graf 图像组下都将压缩后的图像大小减少了 200000 字节以上。对于 bike 图像组,直接依靠单幅图像的压缩比为 1.664,利用与之相似的待配准图像可以将压缩比提高到 2.004,压缩比提高了 0.34。对于 Graf 图像,直接进行无损压缩,压缩比为 1.248,利用本章的压缩算法对配准之后的图像与参考图像进行图像相减,对于差异值再进行压缩,可以将压缩比提高至 1.508,压缩比提高了 20.8%。可以看到,利用差异信息的图像压缩可以得到更大的压缩比。

## 5.6. 本章小结

第四章提出了基于显著性图像的图像配准算法,可以提高配准的准确度。本章在第四章的基础上,提出了一种可以运用在图像数据集上的,具有更大压缩比的图像压缩算法。算法中首先使用图像哈希算法中的差异值哈希算法,在图像数据集中判断、找出合适于图像配准的图像组。配准后的图像与参考图像具有像素级别的相似度,本章利用图像减法获得了二者之间的差异信息,利用差值图像的稀疏性,可以获得对于参考图像更大的压缩比。通过实验对比了直接对单幅图像的无损压缩效果和本章压缩算法的无损压缩算法,实验表明本章的算法大大提高了图像的压缩比,节省了存储空间。

## 第6章 总结与展望

### 6.1. 工作总结

图像配准在图像处理中是极具应用价值的一项技术,关系到后续的各种图像处理如图像融合,图像拼接,影响着医疗、遥感、军事等领域。图像配准一般指两幅图像之间的操作,一幅被称为待配准图像,另外一幅是参考图像。两幅图像之间的内容是同一个场景,但是存在着因为视角变化、光照变化等产生的形变。图像配准的目的就是找到一个恰当的模型来描述这种形变,从而将变形的图像校正到参考图像的形态。图像配准算法大致可以分为以灰度信息为基础的配准方法和基于局部特征的方法。本文的研究重点在基于特征点的图像配准,从两幅图像中具有旋转、平移、尺度不变性的特征点之间拟合出两幅图像之间的空间变换关系。本文研究了基于特征点配准算法的关键步骤和相关的技术,针对其不足之处做出了改进。

本文主要工作可以概括如下:

(1) 从图像配准的背景和意义出发,从数学的角度严谨地认识了图像配准的原理,对各个环节有了充分的认识。按照文献,将图像配准分成三类,针对每一类介绍了其技术路线和国内外研究现状。

(2) 研究了与图像相关的基本技术。包括图像的成像模型、图像坐标系、描述图像变换的模型和图配准的评价标准。

(3) 介绍了几种流行的基于特征点的配准算法如 SIFT、SURF、KAZE,并研究了其关键的步骤。几种算法一脉相承又各具特色,编写代码对几种算法进行了比较。实验了几种算法在各种变换如旋转、平移、噪声下的配准性能。

(4) 基于 SIFT 算法提出了几种改进的算法。针对 SIFT 计算量较大的缺点,提出了使用下采样减小图像金字塔的规模,降低了图像配准计算复杂度的算法。针对 SIFT 没有利用彩色图像中色彩信息的缺点,提出了基于显著性检测的图像配准方法,弥补了 SIFT 算法没有利用图像灰度信息的特点,可以检测出的特征点在数量和质量上都有了提升,拟合出的变换模型质量也更高。对基于显著性检测的图像分割方法做出了改进,分割得到的主体目标更加完整。

(5) 针对图像数据集中可能存在图像内容相似的图像,即存在适用于图像配准的图像族,提出了应用图像配准减小图像集的方法。方法首先计算图像的哈希值找出合适于图像配准的图像,记录图像与配准后图像的差异信息就可以舍弃该图像,从而减小占据的存储空间。

## 6.2. 工作展望

本文针对 SIFT 配准算法在实时性和准确度上进行了研究，并将其中与显著性区域检测相结合的配准算法应用在了图像集的压缩中。但是由于时间和精力有限，还有一些工作需要继续进行。本文的研究课题可以从以下几个方面深入研究：

（1）根据第二章中对多种算法的仿真可以得出，不同算法在不同类型的图像下的表现各有优劣。所以，除了研究图像配准的实时性和准确度，如何通过判断图像的类型进而选择更合适的配准算法是相当有必要的。

（2）除却对算法本身的改进，如何将算法在硬件上实现，借助如 GPU、FPGA 等实现算法的加速，有待在后续的工作中进一步研究。

（3）随着增强现实的进一步普及，三维重建已经越来越重要，图像配准也在向着立体视觉的方面发展。如何将传统二维平面图像中的配准技术与立体视觉结合起来并吸收一些深度学习的技术值得深入性的研究。

基于特征点的图像配准算法及其技术会搭着大数据等技术的东风，越来越深入地进入到我们的生活中，为我们提供更大的便利。

## 参考文献

- [1] Wiesel, Torsten N. "The postnatal development of the visual cortex and the influence of environment." Stockholm: Nobel Foundation (1981).
- [2] Rafael C. Gonzalez, Richard E. Woods (阮秋琦, 阮宇智等译). 数字图像处理.第二版.北京:电子工业出版社,2003.pp:4-560.
- [3] 陈晓曦,王延杰,刘恋.小波阈值去噪法的深入研究[J].激光与红外,2012,42(01):105-110.
- [4] 易三莉,张桂芳,贺建峰,李思洁.基于最大类间方差的最大熵图像分割[J].计算机工程与科学,2018,40(10):1874-1881.
- [5] T. F. Cootes. C. J. Taylor. Statistical models of appearance for computer vision: World Wide Web Publication, 2004
- [6] Zitova B, Flusser J. Image registration methods :a survey[J].Image and Vision Computing. 2003,21(11),pp:977-1000.
- [7] 卞贤掌,费海平,李世强.基于语义分割的增强现实图像配准技术[J].电子技术与软件工程,2018(23):79-82.
- [8] 付浩威. 基于全变分约束的医学图像配准研究[D].哈尔滨工业大学,2018.
- [9] 耿庆田,于繁华,王宇婷,赵宏伟,赵东.基于 SIFT 的车标识别算法[J].吉林大学学报(理学版),2018,56(03):639-644.
- [10] 陶筱娇.SIFT 算法下的多表情人脸识别[J].信息与电脑(理论版),2018(23):127-128+131.
- [11] Lisa Gottesfeld Brown . A Survey of Image Registration Techniques . ACM Comput,ng Surveys[J] , 1992 , 24(4):325~376 .
- [12] Xie S , Tu Z . Holistically-Nested Edge Detection[J]. International Journal of Computer Vision, 2015, 125(1-3):3-18.
- [13] 谢凤英 赵丹培. VISUAL C++数字图像处理(附光盘)[M]. 电子工业出版社, 2008.
- [14] 燕磊, 孙启媛, 朱菲菲. 基于 Oriented FAST 和 MAD 的特征匹配图像自动配准[J]. 计算机与数字工程, 2017(09):179-182.
- [15] 郭龙源, 夏永泉, 杨静宇. 一种改进的彩色图像匹配算法[J]. 计算机工程与应用, 2007, 43(27):98-99.
- [16] 张维琪, 樊斐. 自适应 SSDA 图像匹配并行算法设计与实现[J]. 计算机工程与应用, 2014, 50(20):64-67.
- [17] P.A. Viola, Alignment by maximization of mutual information, Ph.D. thesis, Massachusetts Institute of Technology,Boston,MA,USA,1995
- [18] 曹源, 武艳超, 基于粒子群优化算法和烟花算法的图像配准[J].Internet Herth.2017:152-153
- [19] 伍君, 基于小波变换的图像配准方法研究, 湖南大学硕士学位论文, 2005
- [20] Hu M K. Visual pattern recognition by moment invariants[J]. Information Theory, IRE TRANSACTIONS on, 1962, 8(2): 179-187.
- [21] Khotanzad A, Hong Y H. Invariant image recognition by Zernike moments[J]. Pattern Analysis and



- Machine Intelligence, IEEE TRANSACTIONS on, 1990, 12(5): 489-497.
- [22] Hough P V C. Method and means for recognizing complex patterns: U.S. Patent 3,069,654[P]. 1962-12-18.
- [23] Lin M C, Canny J F. A fast algorithm for incremental distance calculation[C] // Robotics and Automation, 1991. Proceedings., 1991 IEEE International Conference on. IEEE, 1991: 1008-1014.
- [24] Troglio, G.; Le Moigne, J.; Benediktsson. Automatic Extraction of Ellipsoidal Features for Planetary Image Registration [J]. Geoscience and Remote Sensing Letters, IEEE , 2012 , 9 (1):95-99 .
- [25] HansP M. Towards automatic visual obstacle avoidance[C]//Proceedings of the 5th international joint conference on Artificial intelligence. 1977: 584-584.
- [26] Harris C, Stephens M. A combined corner and edge detector[C]//Alvey vision conference. 1988, 15: 50.
- [27] Lindeberg T. Edge detection and ridge detection with automatic scale selection[J]. International Journal of Computer Vision, 1998, 30(2): 117-156.
- [28] Lowe D G. Object recognition from local scale-invariant features[C]//Computer vision, 1999. The proceedings of the seventh IEEE international conference on. Ieee, 1999, 2: 1150-1157.
- [29] Lowe D G. Local feature view clustering for 3D object recognition[C]//Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. IEEE, 2001, 1: I-682-I-688 vol. 1.
- [30] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International journal of computer vision, 2004, 60(2): 91-110.
- [31] S. Wang, H. You and K. Fu, "BFSIFT: A Novel Method to Find Feature Matches for SAR Image Registration," in IEEE Geoscience and Remote Sensing Letters, vol. 9, no. 4, pp. 649-653, July 2012.
- [32] Bay H, Tuytelaars T, Van Gool L. SURF: Speeded up robust features[M]//Computer Vision–ECCV 2006. Springer Berlin Heidelberg, 2006: 404-417.
- [33] Ke Y, Sukthankar R. PCA-SIFT: A more distinctive representation for local image descriptors[C]//Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, 2004, 2: II-506-II-513 Vol. 2.
- [34] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 10, pp. 1615-1630, Oct. 2005.
- [35] Morel J M, Yu G. ASIFT: A new framework for fully affine invariant image comparison[J]. SIAM Journal on Imaging Sciences, 2009, 2(2): 438-469.
- [36] Rublee E, Rabaud V, Konolige K G, Bradski J R. ORB: an Efficient Alternative to SIFT or SURF[C]//IEEE International Conference on Computer Vision .Barcelona, Spain, 2011: 2564-257
- [37] Abdel-Hakim A E, Farag A A. CSIFT: A SIFT Descriptor with Color Invariant Characteristics [C]//IEEE. Conference on Computer Vision and Pattern Recognition. New York, 2006: 1978-1983
- [38] Amintoosi M, Fathy M, Mozayani N, A Fast Image Registration Approach Based on SIFT Key-Points Applied to Super-Resolution[J]. The Imaging Science Journal, 2012, 60(4): 185-201
- [39] Vinividyadharan, Subasurendran, et al. Automatic Image Registration Using SIFT-NCC [J]. Ijca Special Issue on Advanced Computing and Communication Technologies for Hpc



- Applications,2012,ACCTHPCA(4):29-32
- [40] Alcantarilla P F, Bartoli A, Davison A J. KAZE features[M]//Computer Vision–ECCV 2012. Springer Berlin Heidelberg, 2012: 214-227.
- [41] 李晖晖, 郑平, 杨宁, 胡秀华. 基于 SIFT 特征和角度相对距离的图像配准算法. [A]. 西北工业大学学报.2017.4.35(2):281-285
- [42] Pan J P, Hao J M, Zhao J P. Improved algorithm based on SURF for image registration[J]. Remote Sensing for Land and Resources, 2017, 29 ( 1 ) :110 -115.
- [43] 侯坚,张明,一种蚁群优化的改进 SIFT 特征点的图像配准算法[J].现代计算机 图形图像.2016.7
- [44] An Yong, Zheng Hong. SIFT Matching Method Based on K Nearest Neighbor Support Feature Points[J]. 2016 IEEE International Conference on Signal and Image Processing:64-68
- [45] Zhe Xiao, Linfang Yu, Zhen Qin, Huaigui Ren, Zhenwei Geng. A Point Matching Algorithm for Brain CT Images Based on SIFT and Gray Feature.[J] IEEE International Conference on Signal Processing 2016
- [46] Detone D, Malisiewicz T, Rabinovich A. Deep Image Homography Estimation[J]. 2016.
- [47] Calonder M , Lepetit V , Strecha C , et al. BRIEF: Binary Robust Independent Elementary Features[C]// Computer Vision - ECCV 2010, 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV. 2010.
- [48] Beis J S , Lowe D G . Shape Indexing Using Approximate Nearest-Neighbour Search in High-Dimensional Spaces[C]// cvpr. IEEE Computer Society, 1997.
- [49] Fischler M A, Bolles R C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography[J]. Readings in Computer Vision, 1987:726-740.
- [50] Leutenegger S , Chli M , Siegwart R Y . BRISK: Binary Robust invariant scalable keypoints[J]. 2011.
- [51] RostenE,Drummond T.Machine learning for high-speed cornerd etection [M] // Computer Vision–ECCV 2006. Springer Berlin Heidelberg, 2006: 430-443.
- [52] Rosten E, Porter R, Drummond T. Faster and better: A machine learning approach to corner detection[J]. Pattern Analysis and Machine Intelligence, IEEE TRANSACTIONS on, 2010, 32(1): 105-119.
- [53] Gionis A . Similarity Search in High Dimensions via Hashing[J]. Proc. VLDB, 1999, 1999.
- [54] P Perona , J Malik. Scale-Space and Edge Detection Using Anisotropic Diffusion. IEEE Computer Society , 1990 , 12 (7) :629-639
- [55] Perona, P., Malik, J.: Scale-space and edge detection using annisotropic diffusion. IEEE Trans. Pattern Anal. Machine Intell.12 (1990) 1651–1686
- [56] J.Weicket; B.M. T. HRomeny;M.A. Viergver. Efficient.Efficient and reliable schemes for nonlinear diffusion filtering. IEEE TRANSACTIONS on Image Processing.1998
- [57] Alcantarilla P F, Jes úN, Bartoli A. Fast explicit diffusion for accelerated features in nonlinear scale spaces[C]//Electronic Proceedings of the British Machine Vision Conference. Bristol, UK, 2013:1-11.
- [58] Yang X , Cheng K T . Local Difference Binary for Ultrafast and Distinctive Feature Description[J].

- IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(1):188-194.
- [59] 张晨光,周詮,回征.基于 SIFT 特征点检测的低复杂度图像配准算法[J].扬州大学学报(自然科学版),2018,21(04):52-56.
- [60] Lindeberg T. Scale-space theory: A basic tool for analyzing structures at different scales[J]. Journal of applied statistics, 1994, 21(1-2): 225-270.
- [61] Lindeberg T, Gårding J. Shape-adapted smoothing in estimation of 3-D shape cues from affine deformations of local 2-D brightness structure[J]. Image and vision computing, 1997, 15(6): 415-434.
- [62] Wang X , Yu K , Wu S , et al. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks[J]. 2018.
- [63] Itti L , Koch C , Niebur E . A Model of Saliency-Based Visual Attention for Rapid Scene Analysis[J]. 1998.
- [64] Hou X, Zhang L. Saliency Detection: A Spectral Residual Approach[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2007:1-8.
- [65] Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection[C]// Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009:1597-1604.
- [66] Dorin Comaniciu, Visvanathan Ramesh, Peter Meer. Real-Time Tracking of Non-Rigid Objects Using Mean Shift. CVPR, 2000.
- [67] 苏灵松.复杂背景下 MeanShift 结合 Kalman 滤波的车辆跟踪算法[J].电子测量技术,2018,41(06):71-75.
- [68] 周詮,李晓博,方海,黎军,呼延焱,张怡,李静玲.一种基于图像集的数据压缩方法与系统:中国, CN201210251974[P].2015-02-11
- [69] 孔颀,孙权森,纪则轩,刘亚洲.基于仿射不变离散哈希的遥感图像快速目标检测新方法[J].南京大学学报(自然科学),2019,55(01):49-60.
- [70] Candès, E, Wakin M . People hearing without listening: An introduction to compressive sampling[J]. IEEE Signal Processing Magazine, 2008, 25(2):21 - 30.
- [71] 王浩,梁煜,张为.一种均匀化稀疏表示的图像压缩感知算法[J].西安交通大学学报,2019,53(02):136-141.
- [72] Ziv J . A universal algorithm for data compression[J]. IEEE Trans.inf.theory, 1977, 23(3):337-343.

## 致 谢

白昼渐长，春江水暖之际，我的硕士研究生生涯和学生生涯也即将结束，人生的下一站是另外一个熟悉又陌生的城市。回首在西安读书的两年半时间，是我成长最快的两年半，在这里我得以浅尝科研的苦涩与甜蜜，推开了新的世界的门，度过了一段难忘的时光。

在这里首先要衷心感谢的是我的导师，周詮研究员。周老师是一个严厉与和善共存的人，处处为学生着想。感谢周老师对我的教诲让我如沐春风，也感谢老师在工作选择的问题上为我的耐心分析。周老师值得我终身去学习和追赶。

同样需要感谢的还有我的师兄师姐和同学们。回征师兄和呼延焱师兄在合作的项目中和后续的科研中都对我知无不言，解决了我的一些棘手的问题，同样还要感谢刘娟妮师姐的指导。感谢我的同学赵瑞、孟正科、龚静文、何雯等同学，从西电到老区再到新区，吃火锅、泡温泉、打雪仗、踢足球，回想起来，历历在目。

最后需要特别感谢的是我的家人。可怜天下父母心，我一直是家人的骄傲，希望以后也能一直让他们为我感到自豪。感谢我弟弟的成长，和我负担家庭的重任。值得特别感谢的还有我的女朋友张晶，一个爱笑爱哭的女孩。张晶为我的生活增添了许多色彩，谢谢她对我的激励和包容，希望我们能够择一良木，早日安居乐业。

人生路漫漫，诱惑有之，困境有之，高楼平地起和直堕深渊的事情每天都在上演，我的未来和梦想也等着我去书写。在我的梦想里，少不了每一个我爱的和爱我的人，为了梦想，我将在因上足够努力，也将在果上求缘，珍惜身边人。

张晨光

2019年3月写于西安



## 硕士学位论文原创性声明

本人郑重声明：此处所提交的硕士学位论文《基于特征点检测的图像配准算法研究》，是本人在导师指导下，在中国空间技术研究院西安分院攻读硕士学位期间独立进行研究工作所取得的成果。据本人所知，论文中除已注明部分外不包含他人已发表或撰写过的研究成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。本声明的法律结果将完全由本人承担。

签字

日期： 年 月 日

## 硕士学位论文使用授权书

《基于特征点检测的图像配准算法研究》系本人在中国空间技术研究院西安分院攻读硕士学位期间在导师指导下完成的硕士学位论文。本论文的研究成果归中国空间技术研究院西安分院所有，本论文的研究内容不得以其它单位的名义发表。本人同意院方保留并向有关部门送交论文的复印件和电子版本，允许论文被查阅和借阅。本人授权中国空间技术研究院西安分院，可以采用影印、缩印或其他复制手段保存论文，可以公布论文的全部或部分内容。

学生签名：

日期： 年 月 日

导师签名：

日期： 年 月 日



## 个人简历、在学期间发表的学术论文与研究成果

### 个人简历

1993 年 4 月 26 日出生于山西省太谷县。

2012 年 9 月进入电子科技大学通信工程专业，2016 年 7 月本科毕业并获得工学学士学位。

2016 年 9 月进入中国空间技术研究院西安分院攻读工学硕士至今。

### 发表的学术论文

- [1] 张晨光,周詮,回征.基于 SIFT 特征点检测的低复杂度图像配准算法[J].扬州大学学报(自然科学版),2018,21(04):52-56.
- [2] C.G. Zhang, Q.Zhou, and L.Huyan. A novel image registration algorithm based on salient region and SIFT[C]//The 3rd International Society for Artificial Intelligence and Robotics (已被 ISAIR2018 会议录用)
- [3] 张晨光,周詮,回征.基于 SIFT 特征点检测的一种改进的低复杂度图像配准算法.[C]//第十九届全国图象图形学学术会议

### 参与的科研项目

- [1] 国家自然科学基金面上项目:基于图像信息隐藏的卫星数据传输理论与方法(No.61372175)
- [2] 国家重点实验室稳定支持基金:基于预处理的大压缩比图像传输新体制研究(No.2018SSFNKLSMT-13)

### 参与的发明专利

- [1] 周詮,回征,张晨光,呼延焯,张茗茗,刘娟妮,魏佳圆.一种适于微纳卫星的分块数据压缩方法:中国, CN201810634416[P].2018-11-20

