



中国航天

# 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    玲

2019 年 3 月 13 日



# 基于特征点检测的图像配准算法研究

张晨光（数据传输与处理）

导师：周诤

## 摘要

图像配准是数字图像处理领域中的一项重要技术，是模式识别、图像融合、图像拼接中的关键步骤。图像配准在对地貌水文的遥感检测、医学上的病情分析、军事中的精确制导中都有着应用广泛。图像配准这个问题自从被提出，就受到了很多的关注，出现了很多种配准算法，这些配准算法又涉及到了很多其他的基础算法，如边缘检测，图像滤波，插值，图像变换等。图像配准算法大致可以分为三类，一种利用了整体的灰度信息，如互信息法等；一种的配准操作发生在变换域中；还有一种利用了图像的特征点等局部特征。特征点是指那些不随尺度变换和旋转平移而变化的稳定的点，以特征点为基础的算法具有计算量小的特点，更具有鲁棒性，对遮挡，噪声，图像形变的适应性更好。以特征点为基础的配准算法有 SIFT 算法，SURF 算法，KAZE 算法，ORB 算法等。本文将重点放在了以特征点为基础的图像配准算法，尤其是 SIFT 算法，对其进行了改进，并探索了图像配准在图像集中的应用。本文主要研究内容有：

- 介绍和分析了关于图像配准的研究现状和主要的技术路线，重点介绍了基于特征点配准的几种算法的流程，并在不同变换情况下从不同方面对不同算法进行了系统的比较，对主流算法有了清晰的认识。除了图像配准技术本身，本文还研究了与之相关的几种基础技术问题，如图像的坐标系问题，双目视觉问题，图像之间的变换关系。图像表示与变换关系的表示关系到最终的配准效果。
- SIFT 算法效果整体良好，但依然存在缺点。针对计算量过大导致配准较慢和没有利用图像色彩信息的缺点，分别提出了两方面的改进。一个思路是利用图像下采样来减小图像金字塔，从而成比例地减少运算量，达到快速匹配的效果。另外一种是将 SIFT 算法与显著性检测方法结合起来

来，在增加的计算量相当有限的情况下，检测出了更多更高质量的特点，有利于拟合出更加准确的模型，另一方面，利用显著性区域检测可以实现图像主体部分的分割，原始的分割方法在按照本文的方法改进后，可以获得更加完整的主体部分，在此基础上的特征点检测可以去除背景噪声和背景特征的干扰，这在只要求对目标部分进行配准的情况下是有用的。

- (3) 提出了一种图像配准的新的应用方法：借助图像配准描述相似图像间的关系，从而删除相似图像，减小图像数据集。这里的图像数据集特指图像中场景大概率存在相同或者相似的数据集。比如对某个建筑物或者风景的多次连续抓拍，所得图像之间的关系很可能只是视角的变化。同时保存大量的相似图像会加重存储系统的负担，所以可以使用图像配准描述出变换关系：将本地已有的图像作为待配准图像，将新增的相似图像作为参考图像，用保存变换矩阵代替保存新的图像。为了便于复原删去的图像，可以记录配准后的图像与参考图像的像素级差值情况。同时，为了方便判断两幅图像是否包含相似场景，是否需要配准，本文引入了哈希算法，分别计算两幅图像的哈希值，当它们的哈希值的汉明距离小于某个阈值时就认为可以使用图像配准。

关键词：图像配准；特征点；SIFT；显著性检测；哈希算法；差值图像

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

Zhang Chenguang (Information and Communication Engineering)

Directed by Zhou Quan

### Abstract

Image registration is an important technology in the field of digital image processing. It is a key step in pattern recognition, image fusion and image stitching. It is applied widely in satellite monitoring, medical detection, and precision guidance in military. Since the problem of image registration was raised, it has received a lot of attention. A variety of algorithms have been proposed, involving many other basic

algorithms such as edge detection, image filtering, interpolation, and image transformation. The algorithms for image registration can be roughly divided into three categories. The first one is the algorithm that uses the overall gray information, such as the mutual information method, and the second one is operated in transform domain. The third one make use of local features of the image, such as the feature points. Feature points are stable and do not change with scale transformation and rotation translation. Representative algorithms are SIFT algorithm, SURF algorithm, KAZE algorithm, ORB algorithm and so on. The feature point-based algorithm has the characteristics of small computational complexity, and is more robust, and has better adaptability to occlusion, noise, and image deformation. This paper focuses on the image registration algorithm based on feature points, especially the SIFT algorithm. This paper improves it and explores the application of image registration in image set. The main research contents of this paper are:

- (1) Introduce and analyze the research status and main technical routes of image registration, especially the flow of several algorithms based on feature points. And systematically compare different algorithms from different aspects in different situations. In addition to the image registration technology itself, this paper also studies several basic related technologies, such as the coordinate system of the image, the binocular vision problem, and the transformation relationship between the images. The representation of the image and the transformation relationship are related to the final registration effect.
- (2) The SIFT algorithm works well overall, but it still has shortcomings. In view of the shortcomings of excessive calculation which result in slower registration and making no use of image color information, two improvements have been proposed. The first idea is to use image downsampling to reduce the image pyramid, thereby proportionally reducing the amount of computation and achieving a fast matching effect. The second is to combine the SIFT algorithm with the saliency detection method. More high-quality feature points are

detected with few calculation increased, and we can obtain a more accurate model of transform. The saliency region detection can also realize the segmentation of the main part of the image. In this paper, the original method is improved, and a more complete body part can be obtained. On this basis, image registration can remove the interference of background noise and background features. It is useful in cases where only the main part is required to be registered.

- (3) A new application method of image registration is proposed: a method for reducing the scale of image dataset based on image registration. The image data set here refers to a data set which has same or similar scenes with large probability such as multiple shots or snapshots of a certain building or landscape. Save these similar images will increase the burden on the storage system. Since the relationship between images in set are likely to be change of view, we can use image registration to describe the transformation relationship: use the existing image as the image to be registered and use the new similar image as the reference image. We will save the transformation matrix instead of saving the new image. In order to restore the deleted image, the difference between the registered image and the reference image can be recorded. At the same time, there are images with completely different scenes in the data set and these are not suitable for image registration. In order to determine whether the two images contain the similar scenes and whether registration is required, this paper introduces a hash algorithm to calculate the hash values of the two images respectively. When the Hamming distance of their hash values is less than a certain threshold, image registration can be used.

# 目录

<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	4
1.3.1. 基于灰度信息法	5
1.3.2. 基于变换域	7
1.3.3. 基于特征法	8
1.4. 本文的组织结构	14
<b>第 2 章 图像配准相关技术</b>	<b>16</b>
2.1. 数字成像基础	16
2.1.1. 针孔相机模型	16
2.1.2. 图像坐标系	17
2.1.3. 齐次坐标	21
2.2. 图像变换	22
2.2.1. 刚体变换	22
2.2.2. 仿射变换	22
2.2.3. 投影变换	23
2.2.4. 非线性变换	24
2.2.5. 小结	24
2.3. 双目视觉	24
2.3.1. 本质矩阵	25
2.3.2. 基础矩阵	26
2.3.3. 单应矩阵	26
2.4. 图像配准评价标准	27
2.4.1. 准确率	27

2.4.2 配准精度 .....	27
2.4.3 峰值信噪比.....	28
2.4.4 相似度检测 SSIM .....	28
2.4.5 其他评价标准.....	29
2.5. 小结 .....	29
<b>第 3 章 基于特征点的图像配准 .....</b>	<b>30</b>
3.1. 配准流程.....	30
3.1.1. 特征点提取.....	30
3.1.2. 描述子生成.....	30
3.1.3. 特征点匹配.....	30
3.1.4. 变换模型估计.....	31
3.1.5. 像素点重映射.....	32
3.2. 常用算法介绍 .....	32
3.2.1. SIFT .....	32
3.2.2. SURF .....	33
3.2.3. KAZE.....	34
3.2.4. ORB .....	36
3.3. 算法对比.....	37
3.4. 小结 .....	41
<b>第 4 章 基于 SIFT 算法的新算法 .....</b>	<b>42</b>
4.1. 低复杂度的 SIFT .....	42
4.1.1. 理论思想 .....	42
4.1.2. 实验结果 .....	42
4.2. SIFT 与显著性检测.....	43
4.2.1. 显著性图上的配准.....	45
4.2.2. 分割图上的特征点.....	46
4.3. 小结.....	48



第 5 章 配准算法的应用 .....	49
5.1. 相关工作.....	49
5.1.1 哈希算法 .....	49
5.1.2. 图像压缩 .....	50
5.2. 理论思想.....	50
5.3. 实验结果.....	51
5.4. 小结.....	53
第 6 章 总结与展望 .....	54
参考文献.....	56
致  谢.....	61
硕士学位论文原创性声明 .....	62



## 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，看电影，都离不开视觉的贡献。事实上，人和大部分主要通过视觉获得外界的大小、颜色等对机体生存具有重要意义的信息，视觉获得信息的比例远远超过了听觉、触觉等感官。生理学的研究也在不断地为我们揭示视觉的奥秘。D.H.Hubel 和 T.N.Wiesel 凭借在视觉系统的信息加工中的贡献获得了 1981 年的诺贝尔奖<sup>[1][2]</sup>。我们看到的其实是视网膜上面所成的像，外界发出或者反射的光经晶状体等的投影就是所谓的图像。在这个过程中，图像由光信号变为电信号，刺激神经系统，大脑做出相应的反应。

所谓数字图像，就是图像中像素点的坐标和灰度值都是离散数值的一个二维函数，坐标值作为自变量，灰度值作为因变量。数字图像处理起源于 20 世纪 20 年代，当时通过海底电缆从英国的伦敦到美国的纽约传输了第一张数字照片<sup>[3]</sup>。数字图像处理的目的是通过计算机进行计算，得出图像的一些属性值，或者使得图像更适于人眼观看或进一步实行其余的图像处理。传统的数字图像处理包括图像去噪<sup>[4]</sup>、图像增强、图像分割<sup>[5]</sup>、图像压缩<sup>[6]</sup>等。而最近大热的计算机视觉可以理解为是高级的数字图像处理，它大多借助深度学习，终极目标是实现图像理解<sup>[7]</sup>，但是也可以实现传统的图像处理，张等借助残差网络实现了图像去噪，图像解压缩等<sup>[8]</sup>。在数字图像处理之中，图像配准<sup>[8][9]</sup>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需要进行图像配准<sup>[10][11]</sup>。图像配准的输入是相同场景在不同时间，不同视角，通过相同或者不同的传感器得到的两幅图像，拟合出二者之间的变换关系，并将其中一幅图像（待配准图像或变形图像）变换到与另外一幅图像（参考图像）相同状态的技术。

这项技术在遥感，医用，制图，计算机视觉领域被广泛地应用。如增强现实技术中，借助图像配准实现视觉融合，增强对环境的感知<sup>[12]</sup>；在医学分析中，有时候单靠单一医学图像很难准确识别病灶，这就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配准中的非线性不适定问题进行了研究<sup>[13]</sup>；在遥感分析中也是同样的道理，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，同时不同时期的地形地貌水文可能发生改变，需要利用多幅图像进行对比或者拼接，这都需要图像配准技术；在模式识别领域，如车牌识别<sup>[14]</sup>、人脸识别<sup>[15]</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：在样本图像与模板库中的图像进行配准，如果配准成功就说明识别成功。

## 1.2. 图像配准技术概述

图像配准在数学上指是两个二维函数之间的空间变换和灰度变换。我们定义两幅具有偏移关系（包括平移，缩放，旋转）的图像分别为参考图像和偏移图像，并利用二维数组  $f_1(x, y)$  和  $f_2(x, y)$  表示图像相应位置处的灰度值。则两幅图像在数学上有如下变换关系：

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

其中  $h$  表示二维空间坐标变换， $g$  表示灰度或辐射变换，描述因为传感器类型的不同或辐射变形所引起的变换。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只表示位置坐标的变换。灰度或辐射变换可以归为图像预处理部分。

图像配准在目标检测、模型重建、运动估计、特征匹配、肿瘤检测、病变定位、血管造影、地质勘探、航空侦察方面都有广泛的应用。每一种配准方法通常都针对某个具体问题而优化设计的，唯一的共性就是配准问题最终都要在变换空间中寻找一种最有效的变换，这种变换能够使两幅图像之间在某种意义上达到匹配。1992 年，L.G.Brown 在其论文<sup>[16]</sup>中对近三十年来的图像的配准技术进行了系统的总结，他将配准技术分为四个组成部分：特征空间、搜索空间、搜索策略和相似性度量。下面分别简单对这四个部分做一个介绍：

### （1）特征空间

特征空间是与输入空间对应的，又区别于输入空间。对输入的图像进行特征提取，提取出的多维向量构成特征空间。图像的处理大多建立在特征空间之上，这是因为相比于输入空间，特征空间可以更好地表示图像的特定信息，特征对灰度变化、图像变形以及遮挡等有很好的适应能力，同时数据量更少，降低了维度，因此能够大大减小图像处理过程的运算量。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，一般使用神经网络进行特征的提取得到特征图。

### （2）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。这里的空间变换模型根据变换范围可以分成三类：全局、局部、位移场。全局变换指图像中所有的像素可以用同一个矩阵表示；局部则表示图像的不同区域由不同的参数表示：对关键点位置进行参数变换，其他位置进行插值；位移场变换更进一步，对每一个像素点进行独立的参数变换。根据变换方式可以分成线性变换和非线性变换：线性变换包括刚体变换，仿射变换和投影变换，非线性变换一般用多项式函数表示。

### （3）相似性度量

相似性指的是配准（变换）之后的结果与参考图像的相似程度，根据对结果的评估从而对搜索进行调整。不同的特征空间使用的相似性度量一般不同，而相似性度量的值会影响变换参数的选取，所以它们之间的关系是相互影响的。常用的相似性度量有互信息、欧式距离、联合熵等。

### （4）搜索策略

搜索的任务是找到最优的变换参数，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，但在实践中一般选取效率更高的牛顿法、梯度下降法、Powell 法、Brent 法、黄金分割法、抛物线法等。搜索策略直接决定了配准的速度，并且搜索空间和相似性度量也都对搜索策略的选取有影响。

图像配准与图像匹配联系紧密。图像匹配是在两幅图像之间搜索相似内容的过程，目的仅仅是找到相似内容的位置，不包括矫正的过程。图像配准在图像匹配的基础上，还需要拟合出相似内容之间的变换关系（相似内容之间可能存在旋转、缩放等变换），从而将待配准图像变换到参考图像相同的形态。它的目的是达到两个图像的内容在拓扑上的对应和几何上的对齐；即建立两幅图像内容的对应关系，并使相对应的内容在位置上的对齐。这是图像的一种广义匹配概念。在寻找相似内容的过程中，二者使用的方法基本相同。

根据待配准图像之间的关系，Brown 将图像配准分为多源图像配准、基于模板的图像配准、多角度图像配准、时间序列图像配准四大类<sup>[17]</sup>。多源图像配准，图像之间的区别主要是其传感器类别的不同。典型应用是多传感器图像的融合，比如医学图像领域不同成像结果如的 CT、MRI、PET、SPECT；遥感图像领域中不同波段获得的图像融合。基于模板的图像配准，是在待配准图像中寻找标准的模板图像。典型应用是识别和检测模板的位置，如地图、物体和目标等。多角度图像配准，配准图像之间的关系是传感器的角度和位置的不同，在双目视觉中可以利用这一点得到物体的景深信息，但如果后续目的是图像融合，还需要利用投影变换将图像进行配准；时间序列图像配准，不同图像的生成时间不同，相应地可能会出现光照等成像条件的不同。典型应用是利用序列图像，监视和检测目标在一段时间的变化，如医学图像中血管造影，肿瘤变化，遥感中对地貌的监控等。

### 1.3. 国内外研究现状

图像配准最早在上个世纪七十年代出现，由美国提出并运用于导弹制导，极大提高了导弹的命中率，将弹着点平均误差半径缩小到十几米。在之后图像配准被国内外研究者继续研究，并在各个领域得到了广泛应用，如模式识别、医学诊断、计算机视觉等。图像配准是很多重要问题如图像拼接、图像融合的关键步骤。

可以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。但是这是 1992 年的总结，在谢凤英<sup>[18]</sup>等主编的资料中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。不同的图像配准方法总是对应于某种适用的图像变换模型<sup>[19]</sup>。下面对这三种方法做一点介绍。

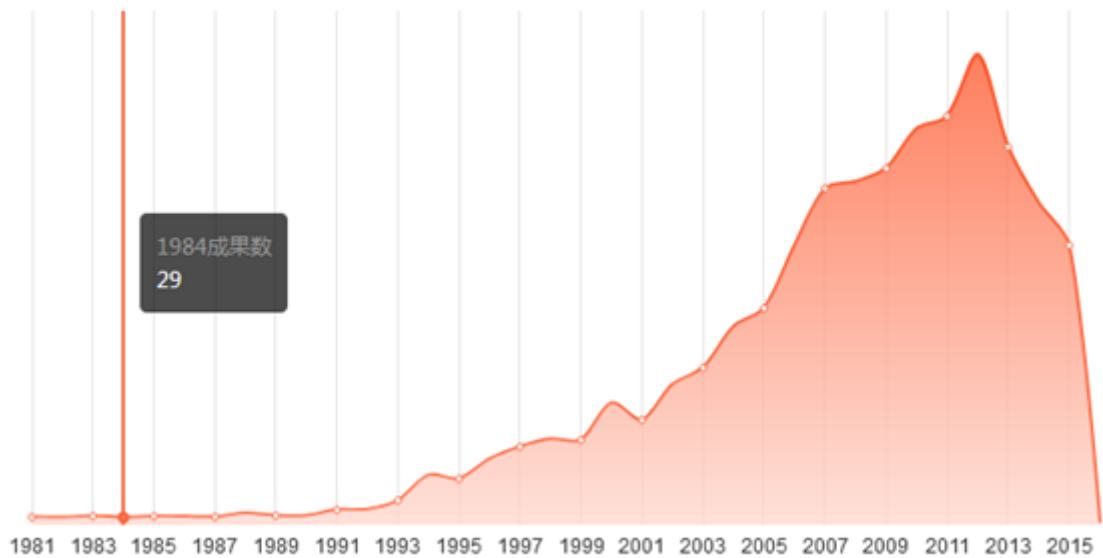


图 1-1: 百度学术统计的图像配准(Image Registration)在各个年份的成果数。

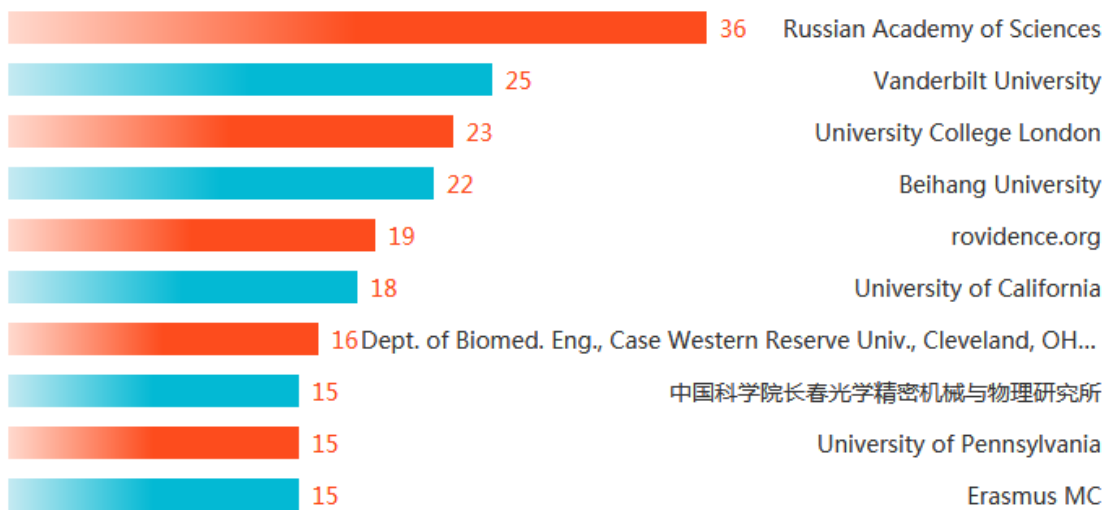


图 1-2: 百度学术统计的各个机构的研究成果数。

### 1.3.1. 基于灰度信息法

基于灰度信息的方法顾名思义就是利用两幅图像之间灰度信息的相似性进行搜索寻找最值，最终确定参考图像和待配准图像之间的变换模型的参数。这种方法的关键是找到模板与子图的（即参考图像和待配准图像）匹配位置，不同的匹配算法主要是相关性度量的准则不同。算法优点是不需要对图像进行复杂特征提取等预处理，缺点是运算量大，没有利用图像的颜色信息；不能直接用于具有非线性变换的图像配准；对噪声较

敏感。基于灰度信息的方法有互相关法；互信息法；序列相似度检测方法 SSDA<sup>[20]</sup>等。下面介绍几种基于灰度信息的配准方法。

平均绝对差算法（Mean Absolute Differences, MAD），由 Leese 在 1971 年提出，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>[21]</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）和 MAD 算法的思想几乎一样，只不过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>[22]</sup>。

而如果将 SAD 方法中的距离度量改为 L2 距离，就是误差平方和算法（Sum of Squared Differences, SSD）。如果对 SSD 中的度量公式求均值，那么就是平均误差平方和算法（Mean Square Differences, MSD）。

序贯相似性检测算法（Sequential Similarity Detection Algorithm, SSDA），由 Barnea 和 Silverman 于 1972 年提出，是一种对于之前的传统模板匹配算法的改进，比 MAD 算法可以快几十到几百倍。在 SSDA 中，首先定义了绝对误差：

$$\varepsilon(i, j, s, t) = \left| S_{i,j}(s, t) - \overline{S_{i,j}} - T(s, t) + \overline{T} \right| \quad (1-2)$$

其中  $S(x, y)$  是搜索图， $S_{i,j}$  是搜索图的子图， $T(x, y)$  表示模板。带有上划线的表示平均值。所以绝对误差就是模板与子图各自减去其自身的均值之后二者的差值。而加速的实现依赖于阈值的设定：子图与模板中随机选取的点之间的误差会累积，当子图产生的误差累加到超过阈值时，说明当前子图不能匹配，就可以及时停止当前子图的计算。根据阈值选取的不同，SSDA 可以大致分成三种：固定阈值 SSDA 算法、单调递增阈值序列的 SSDA 算法和自适应阈值序列的 SSDA 算法<sup>[23]</sup>。张维琪等利用 OpenMP 编程语言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>[24]</sup>。

针对 SSDA 受灰度值的线性变化影响大的缺陷，西北工业大学的贾凯提出了相邻像素差值矩阵的概念，在此基础上利用序贯相似度进行相关性比较，消除了灰度线性变化的影响<sup>[25]</sup>。Yijun Wang 提出了基于小波金字塔作为搜索策略的自适应序列相似性检测算法，提高了匹配速度<sup>[26]</sup>。



Hadamard 变换算法 (Sum of Absolute Transformed Difference, SATD), 也是基于绝对值求和的算法, 不同的是在 SATD 算法中对模板和子图的差值得到的矩阵先求其 hadamard 变换, 再求绝对值之和, 即 SATD 值, 作为相似性测度。SATD 值最小的子图则是最佳匹配。

归一化积相关算法 (Normalized Cross Correlation, NCC), 同样是利用子图与模板图之间的灰度信息, 相关性度量使用归一化的互相关公式。NCC 算法精度高、不易受灰度值的线性变化影响, 但是计算速度较慢。

1948 年, Shannon 首次提出互信息的概念。后经多年发展, 为解决异源医学图像配准问题, Viola<sup>[27]</sup>和 Collignon<sup>[28]</sup> 等人在 1995 年分别独立将互信息引入图像配准领域。利用图像配准位置处互信息最大的原理, 判断配准位置。这种算法的优点是对异源图像配准效果较好, 不需要对不同成像模式下图像灰度间的关系作任何假设, 也不需要图像进行分割或任何预处理, 具有自动化程度高的特点, 因此, 最近几年将互信息作为图像配准过程的相似性测度, 利用最大互信息法进行图像配准成为了图像处理领域的研究热点。但基于互信息的方法的缺点是计算量大, 要求图像间有较大的重叠区, 容易因为插值出现局部极值, 导致误配准。国防科大的马政德提出并实现了一种基于互信息的刚性图像配准并行算法<sup>[29]</sup>。北京交通大学的阮求琦教授团队和大连理工的刘秀平团队也进行了相关研究, 并且结合了局部特征研究。曹蹊渺还在图像配准结果和配准算法性能两方面评价的基础上提出了一种基于距离误差的配准结果客观评价算法<sup>[30]</sup>。曹源、武艳超以互信息作为相似性测度, 提出改进的粒子群算法与烟花算法结合的混合算法来进行图像配准<sup>[31]</sup>。粒子群的惯性权重动态变化提高了粒子的搜索范围, 并利用烟花算法的高斯变异算子增加了粒子的多样性, 防止粒子群算法陷入局部极值, 提高了配准的精确度。

### 1.3.2. 基于变换域

该方法根据变换域内图像变换的方式完成图像的配准。有人把空域和变换域的方法也认为是基于灰度信息的配准, 这里将基于变换域的配准单独看成一类。由空域到变化域的变换主要有傅里叶变换, 小波变换等。

利用傅里叶变换, 可以将一些空间变换在频域中更好地表现出来, 通过平移、旋转、

缩放在时域和频域上的对应关系，可以求得相应的变换参数。傅里叶变换可以利于 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。经傅立叶变换后相位性质表现成信噪比和最大峰值，与图像的空间域配准相比精度更高，所以傅立叶变换非常适合于存在位移变化的图像进行配准。

此外，在傅里叶变换的基础上，在 20 世纪 80 年代前后出现了小波变换。小波变换优于傅里叶变换的地方是它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）。同样是正交变换，但小波变换中的基将傅里叶变换中无限长的单一频率的正弦函数变成了有限的，会衰减的小波基。这样，可以同时应对高频和低频的变换，因此小波又被叫做数学显微镜<sup>[32]</sup>。在和图像分解和重建的塔形算法启发下，Mallat 建立了离散正交小波的一种快速算法，通过与正交镜像滤波器（QMF）的卷积可以分解或重建给定的信号<sup>[33]</sup>。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<sup>[34]</sup>。辽宁科技大学的夏召红团队针对传统边缘特征检测的不足，引入了基于小波变换的边缘特征提取方法<sup>[35]</sup>。

### 1.3.3. 基于特征法

基于特征的图像配准方法是将对整个图像的分析转化为对图像特征的分析，用于降低了分析的维度，因此能够大大减小图像处理过程的运算量。同时特征是在高层次对图像信息的提炼，所以对灰度变化、图像变形以及遮挡等有很好的鲁棒性。基于特征法的图像配准大体上分为四步：特征提取、特征描述符生成、特征匹配、去除误匹配，整体思路是根据两幅图像之间的特征进行匹配，对匹配成功的特征联立方程组，求解变换参数。局部特征一般要满足两个条件：提取的特征要求具有代表性的；提取的特征要能满足仿射、尺度不变性。

算法中常用的特征有直线、边缘、轮廓、重心、不变矩、特征点等。提取直线特征如边缘、轮廓的较为经典算法包括 Hough 特征提取<sup>[36]</sup>、Canny 算子边缘检测等<sup>[37]</sup>。基于 Hough 变换的特征提取算法，相关研究包括 H. Gonçalves;J. A. Gonçalves;L. Corte-

Real; A. C. Teodoro 等人针对遥感图像提出名为 CHAIR 的方法<sup>[38]</sup>, 该方法基于互相关和 Hough 变换, 实验表明可以达到亚像素级的精度, 同时适用于异源图像, 且对于 SAR 图像和可见光图像的配准也具有一定的潜力; 还有 Troglio G 等人提出的将分水岭分割算法和广义 Hough 变换结合的自动特征提取算法, 使用在了数据量庞大的行星图像配准分析中<sup>[39]</sup>。对于中心、不变矩特征提取, 较为经典的算法包括 Hu M K 提出的 Hu 不变矩特征<sup>[40]</sup>、基于 Zernike 多项式的 Zernike 矩特征<sup>[41]</sup>等。相关研究如 Xiaolong Dai 和 Siamak Khorram<sup>[42]</sup>提出的将链码与不变矩相结合对特征进行描述, 实现了对遥感图像的自动配准; Jan Flusser<sup>[43]</sup>利用图像中的不变矩特征, 实现了目标识别和分类; Jan Flusser<sup>[44]</sup>和 Tomáš Suk 针对对称物体识别提出一组旋转、尺度、仿射不变的不变矩。

基于全局的特征配准算法计算量大, 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。所以很多的学者将研究重点放到了对局部特征的研究上, 一系列基于局部特征的优秀算法相继被提出。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: 最大稳定极值区域 (Maximally Stable Extremal Regions, MSERs) 检测算法、基于边缘区域 (Edge-Based Regions, EBRs) 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。最大稳定极值区域 (MSERs) 检测算法基于分水岭算法, 此算法计算复杂度较低, 且运用高效的合并-查找<sup>[45]</sup> (Union-Find) 算法。Tuyltaars T, Van Gool L<sup>[46]</sup>等人提出了基于边缘区域 (EBRs) 检测算法和基于密度极值区域 (IBRs) 检测算法。EBRs 算法检测到的边缘可能发生边缘断裂, 从而造成误匹配。IBRs 算法利用图像中局部密度极值点和以极值点为中心的辐射线上的特征点构造仿射不变区域。Kadir 和 Brady 在 2001<sup>[47]</sup>年提出了显著性区域 (Salient Region) 检测算法, 在 2004 年进行了改进<sup>[48]</sup>。显著区域是图像中最容易受到关注的区域, 可以通过计算像素与其邻域像素点在底层特征上的差别来决定显著度。不同尺度下图像不同局域信息熵会表现出不同的变化规律, 由此可以进行某些局部仿射不变特征的检测。Mikolajczyk、Tuyltaars、Schmid 等人<sup>[49]</sup>对主流的仿射不变区域检测算子进行了测试, 发现 MSERs 算法综合性能最优, Hessian-Affine 算法<sup>[50][51]</sup>次之; Hessian-Affine 算法<sup>[50][51][52]</sup>和 Harris-Affine 算法可以提取较多的不变区域, 更适合复杂场景的匹配; EBRs 算法适用于包含较多交叉边缘的图像; 显著性区域算法的综合评价较低, 但在物体分类识别方面表现较好。

局部点特征是另外一种局部特征。1977 年, Moravec<sup>[53]</sup>利用灰度自相关函数检测出角点。原理十分简单: 在多个方向都有亮度变化的像素点可判定为角点。这种检测方法的缺点是对噪声敏感、不具有尺度和旋转不变性等。Harris 在 1988 年提出了 Harris 角点提取算法<sup>[54]</sup>。其基本思想是观察像素点周围小窗口内的各方向灰度变化的强度。判定方法为首先获取平均强度变化最大对应的方向, 如果它的垂直方向变化强度也很大, 那说明这一像素点就是一个角点。在具体计算中, 可以用协方差矩阵表示所有方向上的强度变化, 此协方差矩阵的两个特征值表征了最大平均强度变化以及垂直方向上的平均强度变化。事实上, 通过协方差矩阵的特征值可以检测出区域、边缘和角点: 若这两个值均较小, 那么像素点位于平坦区域; 若一个较大, 另一个较小, 则说明位于边界上; 若两个值都较大, 说明在两个方向灰度强度变化都剧烈, 属于角点。Harris 角点对旋转和灰度变换具有不变性。

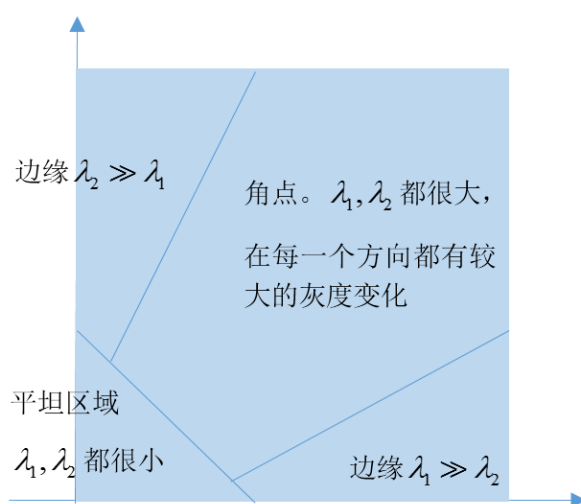


图 1-3: Harris 角点检测, 由协方差矩阵的特征值判断是否是角点

20 世纪 90 年代, Lindeberg<sup>[55-59]</sup>系统提出了图像的尺度空间理论, 从理论上证明了高斯核是构造尺度空间唯一的线性核。借助尺度空间可以提取出尺度不变特征。如 Mikolajczyk 和 Schmid<sup>[57]</sup>将 Harris 角点检测算法与高斯尺度空间相结合, 提出了 Harris-Laplacian 算子, 使 Harris 角点获得了尺度不变性。

在局部特征研究过程中, Lowe 做出了里程碑式的贡献。从 1999 年到 2004 年, Lowe 相继发表的三篇论文<sup>[61-63]</sup>, 系统地提出并完善了 SIFT (Scale Invariant Feature Transform) 算法。Lowe 利用构造高斯差分金字塔的方式, 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  $16 \times 16$  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。然而，算法仍存在一些问题，如阈值过多且难以确定，会存在一定的误匹配直接关系配准的效果。

针对有较多斑点噪声的 SAR (Synthetic Aperture Radar) 图像，基于 SIFT 出现了一种 BFSIFT 算法。由于其独特的成像方式，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。通过双边滤波，BFSIFT 将标准的高斯尺度空间用各向异性尺度空间来代替，可以抑制平坦区域的斑点噪声。在匹配策略上，BFSIFT 采用了双向匹配，减少了误匹配对。但是，双边滤波是一个耗时的过程，所以达到实时性的图像匹配。

2006 年，Bay<sup>[64]</sup>在 SIFT 算法的基础上，提出了另一个具有重要意义的局部特征，即 SURF (Speeded Up Robust Features)。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。SURF 算法在光照变化和视角变化不变性方面的性能接近 SIFT 算法，旋转和图像模糊不变性、鲁棒性优于 SIFT 算法，计算速度快 3 倍左右。在对特征点的描述方面，虽然在统计方式和描述子维数上与 SIFT 算法有所不同，但同样基于梯度直方图统计。

除 SIFT 和 SURF 算法外，比较重要的类 SIFT 算法包括 PCA-SIFT<sup>[65]</sup>、GLOH<sup>[66]</sup>、ASIFT<sup>[67]</sup>、ORB(oriented brief)<sup>[68]</sup>、CSIFT(colored SIFT)<sup>[69]</sup>等其他改进的 SIFT 算法<sup>[70][71]</sup>。PCA-SIFT 利用了主分量分析技术 (Principal Component Analysis, PCA)，其核心思想为采用 PCA 技术，在高维空间中寻找起主要作用的分量，将高维空间的样本投影到低维的特征空间，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，对于异源图像不适用。GLOH (Gradient Location-Orientation Histogram) 描述子将特征点周围矩形模板改为对数极坐标系下的圆形模板，增强了 SIFT 描述子的鲁棒性和独特性。ASIFT (Affine-SIFT)，是 2009 年提出的一种基于 SIFT 的算法，着重解决了 SIFT 等特征提取算法在图像仿射变换角度较大中配准较差的问题。Rosten 和 Drummond<sup>2006</sup> 年提出了 FAST (Features from Accelerated Segment Test) 角点检测算法<sup>[72]</sup>，并在 2008 年进行了改进<sup>[73]</sup>。FAST 算法中，如果一个点周围的圆形区域内有足够多连续点的灰度与中心点差距足够大，则认为此像素点是角

点。在一幅图像中，非角点往往是占多数，而且非角点检测要比角点检测容易得多，因此首先剔除掉非角点将大大提高角点检测速度。N=12 时，编号为 1, 5, 9, 13 的这 4 个圆周像素点中应该至少有三个像素点满足角点条件，圆心才有可能为角点。因此首先检查 1 和 9 像素点。如果  $I_1$  和  $I_9$  在  $[I_p - t, I_p + t]$  之间，则圆心肯定不是角点，否则再检查 5 和 13 像素点。如果这 4 个像素中至少有三个像素满足亮度高于  $I_p + t$  或低于  $I_p - t$ ，则进一步检查圆周上其余像素点。同时，在判定过程中采用了机器学习中的 ID3 算法，加快了判定速度，在实际应用中可以达到实时性要求。

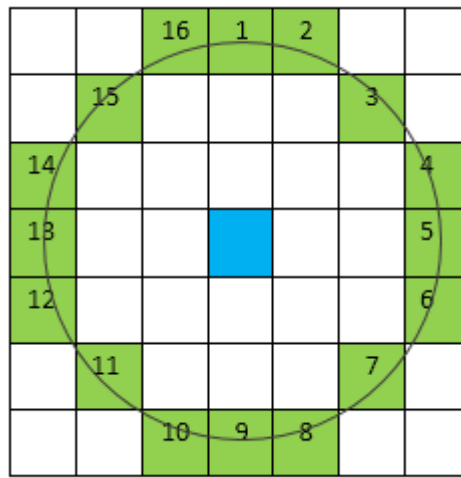


图 1-4: 选取以像素点为中心的半径为 3 的离散化的 Bresenham 圆形区域，如果在圆周上有 N 个连续的像素的亮度比中心点大或小一个阈值，则判为角点。OpenCV 中的 N 选取的是 8。

Taylor、Rosten、Drummond<sup>[74][75]</sup>等人同样利用了机器学习并利用 FAST 特征检测提出了 HIPs 算法。经过对训练图集的学习和算法优化，使得这种算法的特征匹配速度达到了微秒级。Alcantarilla、Bartoli、Davison<sup>[76]</sup>等人在 2012 年发表的论文中提出了 KAZE 特征。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。非线性的尺度分解可以解决这种问题，但新的问题是传统的基于正向欧拉法（forward Euler scheme）求解非线性扩散（Non-linear diffusion）方程时迭代收敛的步长太短，耗时长、计算复杂度高。由此，KAZE 算法的作者提出采用加性算子分裂算法 (Additive Operator Splitting, AOS) 来进行非线性扩散滤波，可以采用任意步长来构造稳定的非线性尺度空间。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。

国内的研究也紧跟国际接轨，并且做出了一些贡献。西北工业大学李晖晖，郑平，杨宁，胡秀华针对参考图像和待配准图像对之间存在的仿射变换问题，提出了一种灵活通用的、基于 SIFT 特征和角度相对距离的图像配准算法<sup>[77]</sup>。算法充分利用了图像正确匹配特征点对之间存在的角度关系，实现了特征点之间的精确匹配。潘建平等基于 SURF，引用相对距离理论，改进了图像配准算法<sup>[78]</sup>。上海海事大学的侯坚、张明提出一种蚁群优化的改进 SIFT 特征点的图像配准算法<sup>[79]</sup>。首先采用内核投影算法 Walsh-Hadamard 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>[80]</sup>。根据配对点周围的点的相似度决定配对的正确与否。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>[81]</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) 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，但仍然存在一些问题：

- (1) 首先当然是配准的精度。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，应该及时作出判断，

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，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，先介绍了图像配准的研究背景和研究意义，然后简单解释了图像配准的定义及其与图像匹配的联系。介绍了图像配准的三类方法，并分别介绍了三类方法的国内外研究现状。

第二章介绍了图像配准相关的技术，如图像变换模型，描述图像变换关系的单应矩阵、本质矩阵，图像的坐标系等。这些都是与图像配准密切相关的知识点，掌握好这些技术才能更好的认识图像配准，更好的进行研究。

第三章是本文的主要研究内容之一，主要介绍了基于特征点的图像配准算法的流程，对常用的算法进行了介绍，并且利用编程模拟了不同的图像变换下的图像配准，利用不同的评估指标对算法性能进行了比较。

第四章是本文针对 SIFT 算法提出的两种改进算法。第一种着重减少图像中特征点的数目，思路是通过下采样减少图像的金字塔规模，试验证明该方法可以有效降低运算时间。第二种是利用了显著性检测的方法，可以充分利用颜色信息，提取更多的高质量的特征点，拟合出更加准确的模型；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。虽然做出了一些微小的成果，但还是有一些不足。针对本文研究中的缺点进行了总结，对今后的工作进行了展望。

## 第 2 章 图像配准相关技术

### 2.1. 数字成像基础

#### 2.1.1 针孔相机模型

真实世界是三维立体的，但是数字图像是平面的，相机模型的出现简化表示了从三维物体到二维平面图像的成像过程。人眼看到的物体实际上都是物体反射了光或物体本身在发光，所以无论是人眼成像还是相机成像的过程实际上都是投影的过程。最基础的投影变换就是正交变换和透视变换。使用正交变换的相机是正交投影照相机（Orthographic Camera），使用透视投影的相机则被称为透视投影照相机（Perspective Camera）。透视投影相机符合我们人眼观察物体近大远小的常识，如(a)，这种模型下我们认为光源是点光源，相机被抽象成一个点；而正交投影中的光源是平行光，这样得到的图像大小和物体实际大小是相同的，在三维空间内的平行线，投影到二维空间中也是一定是平行的，如下图中的(b)。一般说来，对于制图、建模软件通常使用正交投影，这样不会因为投影而改变物体比例；而对于其他大多数应用，通常使用透视投影，因为这更接近人眼的观察效果<sup>[82]</sup>。

最常用的针孔相机模型(Pinhole camera model)的投影是透视投影，数学上大大降低了复杂性，不考虑像差和景深，并且假定等效透镜是薄透镜。所以说针孔相机模型仅仅是一种真实相机的成像过程的近似，甚至于我们可以说这是一种非常粗糙的近似，这使得这个模型对越符合针孔相机模型的真实相机近似程度越高，如网络摄像头、手机镜头、监控探头等等。

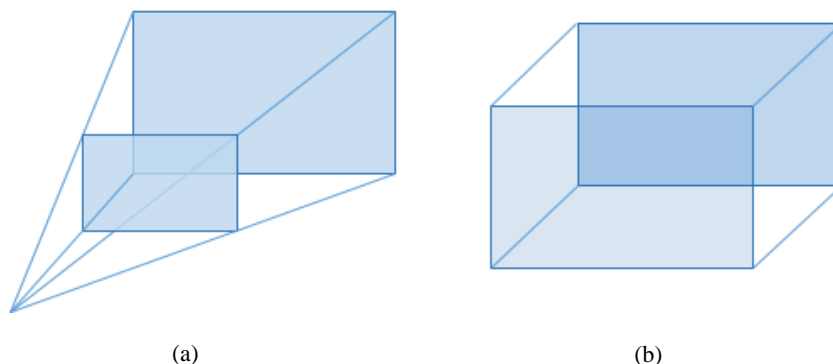


图 2-1 透视投影照相机和正交投影照相机

对于数字彩色图像的获取，还涉及到拜耳彩色滤波阵列。根据对于人眼的分析，人眼中存在红绿蓝三种感光细胞，所以我们可以把可见光分成三原色红绿蓝，根据三原色可以恢复出其他任意一种可见光。所以，我们只要分别获得每一个像素的三原色的值就可以得到真彩图像。三原色中每个通道的获得可以使用滤波片，但是滤波片的工艺难度高，造价高，而且三种需要滤波片严格对齐才能恢复真实色彩，于是，柯达公司的布莱斯·拜尔发明了一种新的滤波片排列方式，只需使用和图像像素大小相同的滤波片阵列进行一次滤波就可以得到三通道信息。下图是一种拜耳阵列：

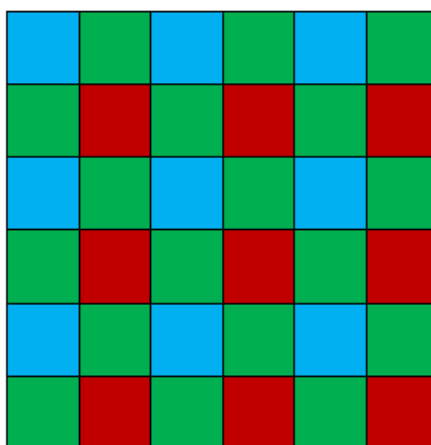


图 2-2 一种拜耳阵列

这样的结果就是有的像素位置捕获到了红色通道的信息，有的是绿色通道信息，有的是蓝色通道信息。对于每个像素位置的其他通道信息，可以通过拜耳阵列对每个三原色进行插值来恢复得到。这里使用的插值方法和常见的插值方法类似，如最近邻插值，双线性插值。

### 2.1.2. 图像坐标系

对于数字图像，我们的操作对象一般以像素为单位。为了表示每个像素，就需要知道其位置坐标，这就需要建立图像的坐标系。图像坐标系一般有四种：世界坐标系，相机坐标系，图像物理坐标系，像素坐标系。我们常用的是像素坐标系，但是了解像素坐标系与其他坐标系之间的关系有利于我们深入认识图像处理。下面对这几个坐标系做一个介绍。

世界坐标系，也称客观坐标系，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。

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，相机的光轴为 $Z$ 轴，方向指向相机前方， $X$ 轴与 $Y$ 轴与物体坐标系平行，单位为米。相机坐标系也使用右手坐标系。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：

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + T \quad (2-1)$$

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, R:3*3, T:3*1 \quad (2-2)$$

图像物理坐标系，也称为平面坐标系，原点在传感器中心， $X$ 轴与 $Y$ 轴分别与图像的边界平行。用 $(x,y)$ 表示其坐标。图像物理坐标系依然使用物理单位如毫米，依然使用右手坐标系。相机坐标系到图像物理坐标系的转换关系：

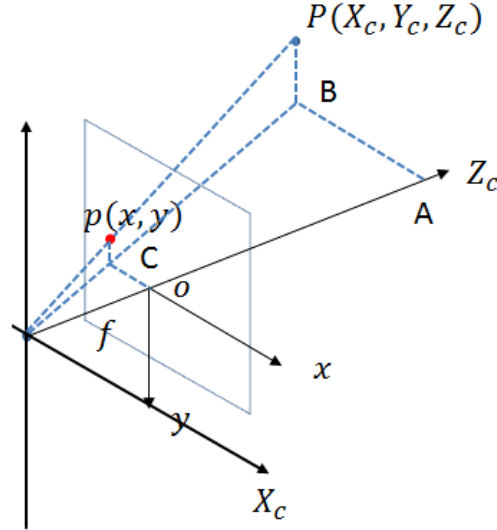


图 2-3 相机坐标系到图像物理坐标系的转换

相机坐标系到图像物理坐标系的转换可以用如下的表达式和矩阵表示。其中 $f$ 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。

$$x = f \frac{X_c}{Z_c}, y = f \frac{Y_c}{Z_c} \quad (2-3)$$

$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (2-4)$$

像素坐标系与其余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，用像素的数目来衡量。

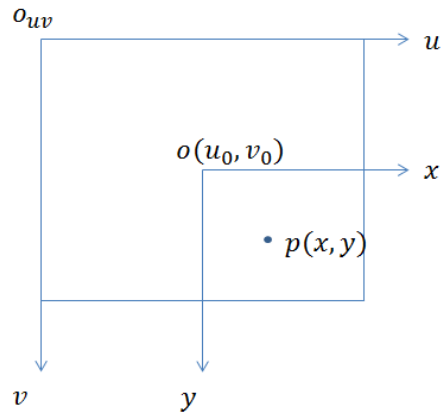


图 2-4 图像物理坐标系到像素坐标系的转换有两个要素

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  $dx, dy$ 。

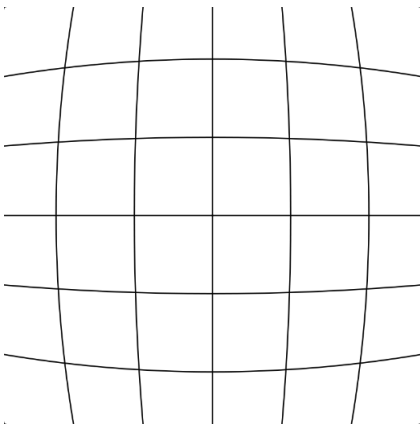
$$\begin{cases} u = \frac{x}{dx} + u_0 \\ v = \frac{y}{dy} + v_0 \end{cases} \quad (2-5)$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2-6)$$

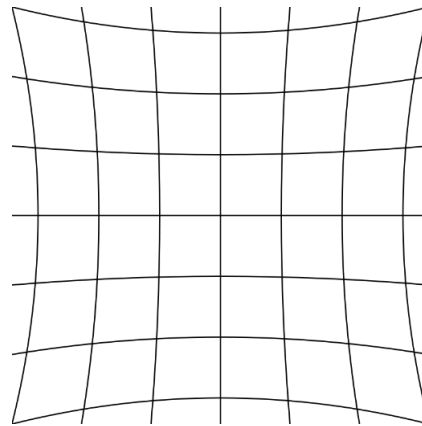
所以，我们可以一路从世界坐标转换到像素坐标，这也是数字图像的数字化生成过程。

$$\begin{aligned}
 Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \\
 &= \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}
 \end{aligned} \tag{2-7}$$

第二行等式中的第一个矩阵是相机内参，包含  $x$  轴和  $y$  轴方向上的归一化焦距和图像中心与左上角的相对位置信息，第二个矩阵是相机外参，包括旋转向量和平移向量。相机的内参和外参可以通过张正友标定进行获取。另外，畸变也属于相机的内参，也可以在标定过程中解决。畸变主要是由镜头造成的，可以分为径向畸变和切向畸变。切向畸变产生的原因是感光元平面和透镜不平行。径向畸变产生的原因是光线在透镜中心和远离中心的地方的弯曲程度不同。径向畸变可以分成桶形畸变（Barrel distortion）和枕形畸变（Pincushion distortion）。桶形畸变中，物体得到不同程度的放大，偏离光轴越远，放大程度越小，就好像是图像被重新映射在了一个桶形的面或者球面上。最常见的桶形畸变发生在广角鱼镜头（Fisheye lenses）中，该镜头将宽度无限的物体映射到了有限的图像中。枕形畸变中，偏离光轴的部分放大效果越显著，光线向光心弯曲，就好像是枕头中被下压的部分。



(a) 桶形畸变



(b) 枕形畸变

图 2-5 两种径向畸变

### 2.1.3. 齐次坐标

在平常数学运算中，我们一般使用笛卡尔坐标，但是在计算机图形学中使用的更多的是齐次坐标（Homogeneous Coordinates）。对齐次坐标每一个分量同等进行缩放，都表示的是笛卡尔坐标系中的同一个点，这就是尺度不变性（Scale Invariant），也就是齐次。齐次坐标可以用  $n+1$  维的坐标表示笛卡尔坐标系中  $n$  维的坐标。齐次坐标  $(\hat{x}, \hat{y}, \omega)$  对应笛卡尔坐标  $\left(\frac{\hat{x}}{\omega}, \frac{\hat{y}}{\omega}\right)$ ， $\omega$  是缩放系数。一般的，取  $\omega=1$ ，得到归一化的齐次坐标。 $\tilde{x}$  是齐次坐标， $\bar{x}$  被称为增广矢量（augmented vector）。

$$\tilde{x} = (\hat{x}, \hat{y}, \omega) = \omega(x, y, 1) = \omega\bar{x} \quad (2-8)$$

齐次坐标有三大优点。

第一，齐次坐标可以表示无穷远处的点。在欧式空间（Euclidean space）或者笛卡尔空间（Cartesian space）中，平行线是无法相交的，但是在真实世界中因为更多地使用透视投影，平行线在无穷远处也可以看作是相交的。在齐次坐标中，当取  $\omega=0$ ，就可以表示笛卡尔坐标系中的无穷大，对应无穷点（points at infinity），而欧式空间中的坐标  $(\infty, \infty)$  没有意义。

第二，齐次坐标可以区分点和向量，同时也更易于进行仿射几何变换<sup>[83]</sup>，这一点在之后介绍仿射变换中将再次提到。对于在笛卡尔坐标系中的一个坐标，我们无法判断它表示的是一个点还是从原点指向该点的向量。齐次坐标因为有一个冗余的维度（缩放系数），可以作为标志位，当它取 1 时说明这个坐标表示的是坐标点，当它为 0 时表示的是向量。

第三，齐次坐标允许平移、旋转、缩放和透视投影表示为矩阵与向量相乘的运算，而使用笛卡尔坐标，平移和透视投影不能表示成矩阵相乘。仿射变换其实就是线性变换（旋转、缩放和“推移”）与平移的叠加。笛卡尔坐标系不能用乘法表示仿射变换主要是因为平移变换在笛卡尔坐标系中只能表现为矩阵加法。现在利用齐次坐标就可以将平移变换中的矩阵相加转换为矩阵相乘：

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x+t_x \\ y+t_y \\ 1 \end{bmatrix} \quad (2-9)$$

## 2.2. 图像变换

图像配准的目的是将不同时间或者不同观察点或者不同传感器拍摄的同一场景的图像进行校准，本质就是要找出参考图像与待配准图像坐标系之间的空间变换和灰度变换关系，而尤以图像之间的空间变换关系最为关键。

变换模型<sup>[84][85]</sup>是指根据待匹配图像与背景图像之间几何畸变的情况，所选择的能最佳拟合两幅图像之间变化的几何变换模型。可采用的变换模型有如下几种：刚性变换、仿射变换、透视变换和非线形变换等。

### 2.2.1. 刚体变换

刚体变换（Rigid Transform）中，图像变换前后像素点之间的相对距离都保持不变，只是图像作为整体其绝对位置进行了变化。为平移变换、旋转变换和反转（镜像）变换都属于刚体变换。刚体变换中平移变换和旋转变换是不可逆的，也就是说对图像先进行平移后旋转和先旋转后平移，产生的效果是不同的。

### 2.2.2. 仿射变换

仿射变换（Affine Transformation）可以看作是线性变换和平移变换的叠加。线性变换指图像的原点保持不变，直线之间的比例保持不变，直线在变换之后也依然是直线。线性变换包括图像绕原点的旋转，还有“推移”：保证原点不变前提下平行线之间在各自直线方向的相对移动（对于不存在平行线的图像可以视作网格线之间的相对移动）。以下图为例，(a)是原图，(b)是在(a)基础上的旋转，(c)是在(b)基础上的“推移”。



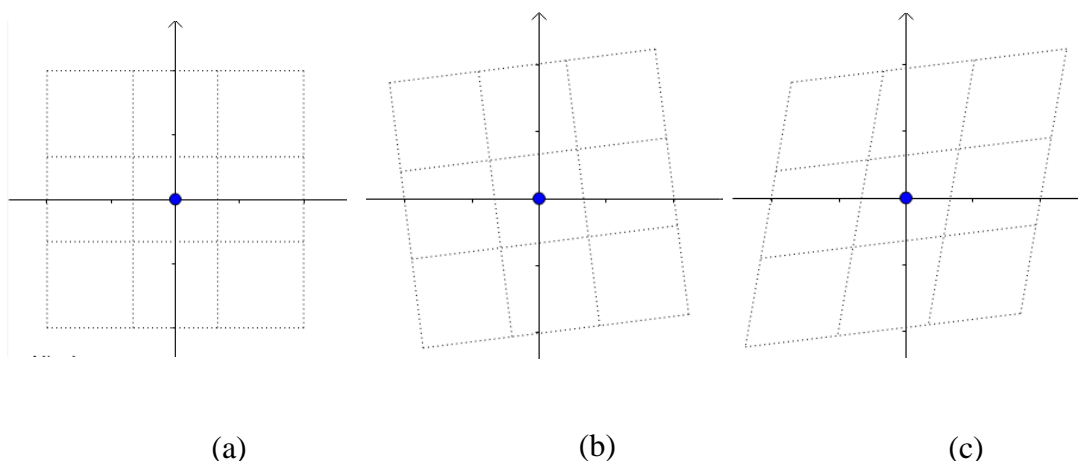


图 2-6 仿射变换中的线性变换（旋转与“推移”）

仿射变换是计算机中的一种重要的图形变换，可以用下面公式表示：

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix} + B \quad (2-10)$$

利用齐次坐标可以将仿射变换表示为齐次坐标与 2 行 3 列的矩阵相乘的形式：

$$M = [A \ B] = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \end{bmatrix}, \quad T = M \cdot [x \ y \ 1]^T. \quad \text{这时的仿射矩阵为 2 行 3 列的矩}$$

阵，其实质是将仿射转换为三维空间中的线性变换，即三维空间中的线性变换可以表示二维空间中的线性变换和平移变换，因为三维空间中的“推移”表现为平行平面之间的相对运动，从而可以表示二维平面的平移运动。

### 2.2.3. 投影变换

投影变换（Projective Transformation）是我们常见的一种变换，在光源下，三维的物体会投影变成一个二维的图像。变换后直线仍然是直线。不同于欧式几何，在投影空间中，借助于齐次坐标，平行线之间可以在无限远处相交。和相机模型一样，投影变换根据不同的投影方式又分正侧投影（Orthographic Projection）和透视投影（Perspective Projection）。

正侧投影也叫正交投影，多用于 2D 游戏中。正交投影的特点是光线为平行线，投影结果与原物体大小一样。透视投影更符合我们的常识，因为太阳的光线只是接近于平行线，实际上还是一个点光源，这就会造成投影图像近大远小的特点，离点

光源越近，其图像越大，这个特点有利于逼真地还原 3D 图像。透视投影对应的变换是透视变换，透视变换对应的矩阵是一个  $3 \times 3$  大小的矩阵。

事实上，无论是平移和旋转这样的基本变换形式，还是刚体变换、相似变换和仿射变换这样的复杂的空间变换，都是投影变换的某种特殊表示，都可以利用齐次坐标写成  $3 \times 3$  大小的投影变换矩阵的形式。

#### 2.2.4. 非线性变换

非线性变换又称为弯曲变换(Curved Transform)，直线经过非线性变换不一定再是直线。多项式变换是典型的非线性变换，非线性变换也可以使用二次、三次函数或者样条函数、指数函数表示。

#### 2.2.5. 小结

在 2.2 节中基本上按照从易到难的顺序介绍了图像的变换模型，复杂的模型可以包含简单模型中的图像变换。下表是各种图像变换之间的比较，可以看到越复杂的变换模型，可以表示越多的图像变换类型。

表 2- 1：各种变换模型比较

刚体变换	反转	旋转	平移			
仿射变换	反转	旋转	平移	缩放		
投影变换	反转	旋转	平移	缩放	投影	
非线性变换	反转	旋转	平移	缩放	投影	扭曲

图像配准中，采用不同的图像变换，图像像素点之间的重映射关系将会不同。因此对于不同类型的图像，选取恰当的图像变换模型可以适当地提高图像的配准精度。本文综合考虑了复杂性与准确性，选择了投影变换中的透视变换模型作为重映射类型。

### 2.3. 双目视觉

图像配准中的输入图像很大一类属于视角变化，是相机相对位置的变化造成

的。同时在计算机图形学中也存在一种双目视觉，使用两个相机获得图像的景深信息。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，表示了两幅来自不同相机之间的图像的像素之间的约束关系。下图是对极几何的示意图。

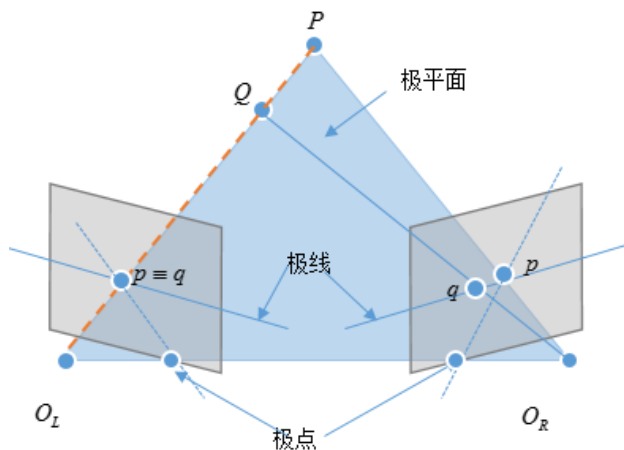


图 2-9 对极几何

上图中，三维空间中的点  $p$  与左右两个相机的光心  $O_L, O_R$  三点组成了极平面，极平面与成像平面的交线就是极线。两个光心之间的连线是基线，基线与成像平面的两个交点就是极点。

### 2.3.1. 本质矩阵

使用相机坐标系，三维空间中的点分别以两个光心为原点，可以得到两个坐标  $P_L, P_R$ 。两个坐标与光心的连线构成两个向量，连同光心之间的连线，基线，三个向量构成了极平面。根据同一平面内三个向量混合积为 0 的定理，可以得到两个三维空间点的两个相机坐标  $P_L, P_R$  之间的约束关系： $P_R^T E P_L = P_R^T R S P$ 。其中  $E$  就是本质矩阵。

根据对极几何，三维空间中的一个点在不同相机的成像平面上投影得到的像素点之间存在极线约束，对极几何利用极平面的特殊性质将对应点约束在另外一幅图像的极线上。两个相机之间的关系用旋转矩阵  $R$  和平移向量  $T$  表示，相当于引入的先验信息。由极平面内三向量共面，可以推导出两个成像平面内点的约束关系，

即本质矩阵  $E$  (Essential matrix):

$$E = RS = R \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \quad (2-11)$$

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。如果将相机坐标系转化为图像物理坐标系，得到坐标  $p_l, p_r$ ，二者之间的约束关系仍然通过本质矩阵表示： $p_r^T E p_l = 0$ 。

### 2.3.2. 基础矩阵

在本质矩阵的基础上，如果将坐标转化成利于图像处理的像素坐标  $\bar{p}_r, \bar{p}_l$ ，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：

$$\bar{p}_r^T F \bar{p}_l = 0。$$

$$F = M_r^{-T} R S M_l^{-1} \quad (2-12)$$

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。

### 2.3.3. 单应矩阵

对于图 2-9 中的空间点  $P, Q$ ，在相机坐标系或者图像像素坐标系下，本质矩阵和基础矩阵其在两个成像平面上的点约束在极线上，如直线  $pq$ 。这样可以在右侧相机生成的图像中搜寻  $P$  点对应的像素时，可以将搜索空间从二维缩小到一维。特别的，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，这时候的约束关系通过单应矩阵 (Homography matrix) 来表示。

$$H = R + T \frac{1}{d} N^T \quad (2-13)$$

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，是它不仅仅引入了两个镜头之间的旋转和平移变换，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。

## 2.4. 图像配准评价标准

### 2.4.1 准确率

准确率  $A_c$  定义如下：

$$A_c = N_c / (N_c + N_e) \quad (2-14)$$

式中， $N_c$  为最终正确的匹配点对数目， $N_e$  是错误的匹配点数目。在理想情况下，图像  $I_1$  的点  $p_i$  通过变换后映射到  $I_2$  的点为  $p_j$ ， $p_i$  和  $p_j$  重合时被视为正确匹配点对。由于各种原因导致实际无法达到完全重合的效果，如果  $p_i$  和  $p_j$  在空间和尺度上非常近，其距离在规定范围内，则认为他们就是匹配正确的特征点对。

### 2.4.2 配准精度

配准精度是用来衡量配准结果的好坏，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  $RMSE$  来衡量。

$$RMSE = \sqrt{\frac{1}{n} \|(x'_i, y'_i) - f(x_i, y_i)\|^2} \quad (2-15)$$

式中， $(x', y')$  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。 $RMSE$  数值越小，则表示精度越高，效果越好。

### 2.4.3 峰值信噪比

峰值信噪比  $PSNR$  (Peak Signal to Noise Ratio) 经常用作图像压缩等领域中信号重建质量的测量方法, 通过均方差 ( $MSE$ ) 定义。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:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i, j) - K(i, j)\|^2 \quad (2-16)$$

峰值信噪比定义为:

$$PSNR = 10 \cdot \log_{10} \left( \frac{MAX_I^2}{MSE} \right) = 20 \cdot \log_{10} \left( \frac{MAX_I}{\sqrt{MSE}} \right) \quad (2-17)$$

其中,  $MAX_I$  是表示图像像素点取值的最大值, 如果每个像素点用 8 比特量化, 那么就是  $MAX_I = 255$ 。

$PSNR$  是使用最广泛的一种图像客观评价指标。它基于对应像素点间的误差, 并未考虑人眼的特性。人眼在观察时对波长、空间频率不同的区域有差异性。

### 2.4.4 相似度检测 SSIM

$SSIM$  (Structural Similarity) 结构相似性更加符合人眼的观感受。 $\mu_x$ 、 $\mu_y$  分别表示图像  $X$  和  $Y$  的均值,  $\sigma_x$ 、 $\sigma_y$  分别表示图像  $X$  和  $Y$  的方差,  $\sigma_{xy}$  表示图像  $X$  和  $Y$  的协方差。 $SSIM$  从亮度  $l$ 、对比度  $c$ 、结构  $s$  三方面衡量了图像的相似性。

$$l(X, Y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (2-18)$$

$$c(X, Y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (2-19)$$

$$s(X, Y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (2-20)$$

$C_1$ 、 $C_2$ 、 $C_3$  为常数

$$SSIM(X,Y)=l(X,Y)*c(X,Y)*s(X,Y) \quad (2-21)$$

$SSIM$  取值范围[0,1],  $SSIM$  值越大, 表示图像之间相似度越高。

### 2.4.5 其他评价标准

配准时间、配准率、算法复杂度、算法的可移植性、算法的适用性、图像数据对算法的影响等。采用主客观结合来评价本算法的好坏, 更加具有说服力。客观评价方法为归一化互信息(NMI), 是信息论互信息的延伸, Studholme<sup>[86]</sup>等提出 NMI 评价配准结果, 其值越大表示配准效果越好。主观评价方法为图像差值法, 即参考图像减去配准后的图像得到差值图像, 差值图像中灰度值越小表示配准结果越好。

## 2.5. 小结

本章介绍了与图像配准相关的一些基本技术, 从图像的成像原理, 图像的坐标系和坐标表示, 到图像变换类型和其使用的矩阵表示。图像的变换模型将两幅图像联系起来, 图像配准的目的就是拟合出图形之间的变换模型参数, 拟合的好坏直接关系到配准的效果。图像配准与流行的双目视觉也有共通之处, 图像配准的输入图像在很大程度上就是其传感器之间存在位置的相对变化, 所以研究了对极几何和与之相关的几种矩阵。此外, 还介绍了几种图像配准的评价指标, 这在本文之后的一些实验中将会用到。

## 第3章 基于特征点的图像配准

### 3.1. 配准流程

#### 3.1.1. 特征点提取

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；每次检测出的特征点应该是一致的，即特征点要具有可重复性；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，减少不确定因素对结果造成的干扰。不同图像配准算法之间主要的不同点就是关于如何提取特征点。

#### 3.1.2. 描述子生成

相对于原始图像，特征点所含的信息较少，但特征点除了包含和普通二维平面的点一样的坐标信息，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中特征点的匹配过程。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。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 **BEIEF**，可以直接生成二进制的字符串。

#### 3.1.3. 特征点匹配

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 (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，距离最近的点认为是匹配的，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 算法来构建 KD 树来实现。

#### 3.1.4. 变换模型估计

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。

### 3.1.5. 像素点重映射

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。

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。

## 3.2. 常用算法介绍

自从 SIFT 算法提出以来，以此为基础出现了很多优秀的特征点检测和匹配算法，如 SURF、ORB、KAZE 等。各种算法具有不同的特性，这里做一个简单的介绍和对比。

### 3.2.1. 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。此算法由 David Lowe 在 1999 年所发表<sup>[61]</sup>，2004 年完善总结<sup>[63]</sup>。在 Mikolajczyk 对包括 SIFT 算子在内的十种局部描述子做了不变性对比实验，证实了 SIFT 及其扩展算法在同类描述子中具有最强的健壮性。

极值点的检测在差分高斯金字塔中进行。搜索同尺度的 9 个邻域点和上下相邻尺度对应的  $9 \times 2$  个点中的极值，再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。再借助 Hessian 矩阵和特征值的特点剔除落在边缘的不稳定的特征点。

特征点除了要保证具有尺度不变性,还需要具有旋转不变性,这一点通过为特征点分配方向信息来实现。数字图像中像素点是离散的,梯度通过差分的形式表示。梯度幅值大小  $m(x, y) = \sqrt{(L(x+1, y) - L(x, y+1))^2 + (L(x, y+1) - L(x, y-1))^2}$ , 梯度方向为  $\theta(x, y) = \alpha \tan(2(L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y)))$ 。对特征点的  $3\sigma$  邻域内所有像素的梯度信息进行直方图统计。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。为了增强鲁棒性,可以借鉴最近邻比率的方法,峰值大于主峰值 80% 的次高峰代表的梯度方向作为辅方向。

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。

### 3.2.2. SURF

虽然 SIFT 算法被认为是最有效的,也是最常用的特征点提取的算法,但如果不借助于硬件的加速, SIFT 算法很难达到实时的程度。对于需要实时运算的场合,如基于特征点的目标实时跟踪系统,处理速度要达到 8-24 帧每秒,这是 SIFT 不能满足的。SURF 借鉴了 SIFT 中使用差分高斯 (Difference of Gaussian, DoG) 近似拉普拉斯高斯 (Laplace of Gaussian, LoG) 的思想,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。实验证明, SURF 算法的运算速度要比 SIFT 算法快 3 倍左右。

SURF 算法利用 Hessian 矩阵行列式的极大值检测特征点。具体的,为了使得特征点具有尺度不变性,在计算行列式之前要使用高斯模板对图像进行滤波,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,计算图像内任何矩形区域内的像素值之和只需要三次加减法,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。

每个像素的 Hessian 矩阵：

$$H(f(x, y)) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} \quad (3-1)$$

经过高斯滤波后再进行 Hessian 的计算，其公式如下：

$$H(x, \sigma) = \begin{pmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{pmatrix} \quad (3-2)$$

SURF 构建金字塔的方式和 SIFT 不同。在 SIFT 算法中，不同组（octave）的图像大小是不一样的，下一组是上一组图像的下采样（1/4 大小）；在每一组中，图像大小一样，不同的是采用的尺度  $\sigma$  不同。对于 SURF 算法，由于采用了盒子滤波和积分图像，采用了不断增大盒子滤波模板的尺寸的间接方法来构建金字塔。图像的大小总是不变的，改变的只是高斯模糊模板的本身的大小和尺度  $\sigma$  的大小。

SURF 确定主方向和生成描述子的方法与 SIFT 不同。在 SURF 中，不计算梯度大小和方向，而是统计特征点圆形邻域内的 Haar 小波特征。 $x$  和  $y$  方向的 Haar 小波特征值反应了图像灰度在水平和垂直方向变化的情况，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。就是描述那些灰度变化特别剧烈的区域方向。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。

### 3.2.3. 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是在线性尺度空间上检测得到的。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。

2012 年，欧洲计算机视觉国际会议（European Conference on Computer Vision, ECCV）中提出了一种比 SIFT 更稳定的特征检测算法 KAZE<sup>[76]</sup>。作者提出了一种基于非线性尺度空间的特征点检测方法。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主要基于非线性扩散滤波原理，非线性扩散滤波的基本公式是：

$$\frac{\partial L}{\partial t} = \text{div}(c(x, y, t) \cdot \nabla L) \quad (3-3)$$

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。非线性滤波在图像方面的应用是由 Perona 和 Malik 于 1990 年<sup>[87]</sup>提出来的。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 提出了一个依赖图像梯度幅值的传导函数，使得扩散能够适应图像局部特征：

$$c(x, y, t) = g(|\nabla L_\sigma(x, y, t)|) \quad (3-4)$$

其中  $L_\sigma$  是原始图像在高斯平滑后的梯度。由于非线性偏微分方程没有解析解，只能通过一些数学方法近似求解，文章中使用了加性算子分裂算法（Additive Operator Splitting, AOS）进行求解<sup>[88]</sup>。采用加性算子分裂算法(Additive Operator Splitting, AOS)来进行非线性扩散滤波，采用任意步长来构造稳定的非线性尺度空间。下面是构建尺度空间的一些细节。

与 SIFT 类似，KAZE 算法也通过指数步长构建了  $O$  个组， $S$  个层的离散化尺度空间。但与 SIFT 不同，而与 SURF 相似的是各组图像都是对原图像的操作而不是对图像进行下采样。各个层之间的尺度关系如下：

$$\sigma_i(o, s) = \sigma_0 2^{o+s/S}, o \in [0 \dots O-1], s \in [0 \dots S-1], i \in [0 \dots N] \quad (3-5)$$

其中  $\sigma_0$  是基本尺度， $o$  表示组的序号， $s$  表示所在的组的层序号， $N$  表示总的层数。为了进行非线性扩散滤波，需要将尺度空间中的尺度单元转换为时间单元：

$$t_i = \frac{1}{2} \sigma_i^2, i = \{0 \dots N\} \quad (3-6)$$

然后根据 AOS 方法进行非线性扩散滤波得到每一层。KAZE 算法中特征点的检测与 SIFT 类似，而确定特征点主方向的方法与 SURF 类型，描述子的建立采用了 M-SURF 描述子。

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。AKAZE 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 Fast Explicit Diffusion (FED) 加入到图像金字塔中，更准确的同时显著提升了速度，且易于实施。在描述子方面，AKAZE 使用了更高效的 Modified Local Difference Binary (M-LDB)，从非线性尺度空间提取了梯度 (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。

### 3.2.4. ORB

2011 年，OpenCV 实验室提出了 ORB 算法[68]。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。ORB 算法真正实现了实时性的特征点检测和匹配，作者在安卓智能手机上进行了验证。

ORB 的特征点检测和特征点的描述子分别借鉴了速度最快的 FAST 算法和 BRIEF 算法，但是在具体使用时又都针对他们的缺点进行了改善。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。这种改进的 FAST 方法称为 oFAST。

即便根据特征点主方向将特征点进行旋转，因为 BRIEF 描述子没有旋转不变性，描述子在旋转前后会产生变化。ORB 基于 BRIEF 提出了两种描述子，一种是 steer BRIEF，另外一种是 rBRIEF。tests 是 BEIEF 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 **tests** 之间相关性低的 **tests**。

在寻找最近邻点匹配时，**ORB** 使用了局部哈希（Locality Sensitive Hashing, LSH）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。

### 3.3. 算法对比

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。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，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。

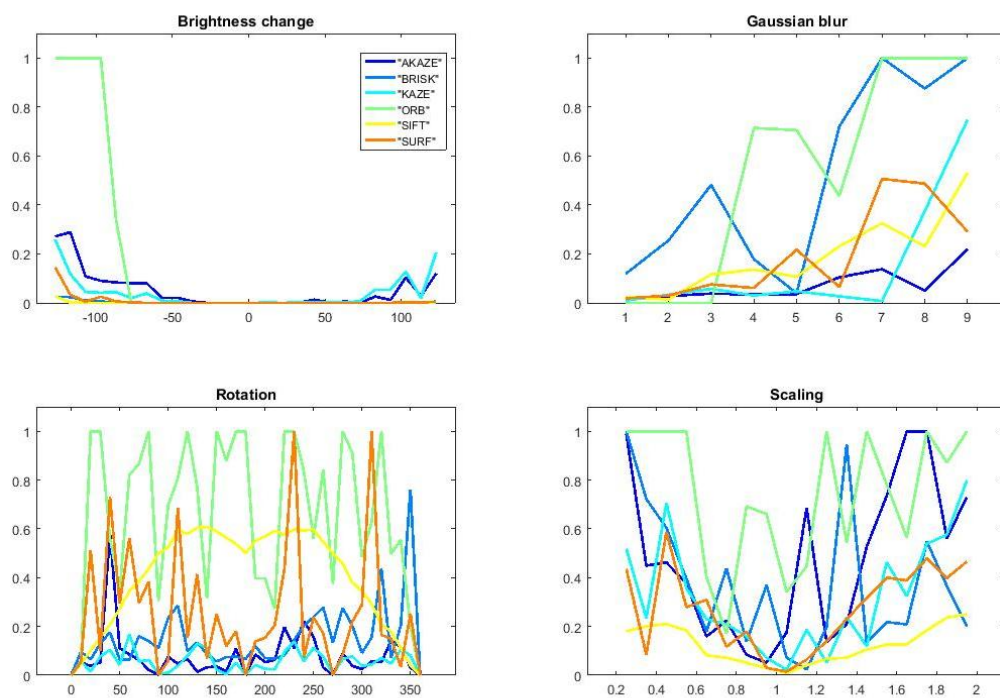
矩阵误差描述的是估计出的变换矩阵与真实的变换矩阵之间的差距。事实上，在编程实现中，根据匹配点对一开始求解出的变换矩阵与真实的变换矩阵理论上应该是互逆的，即二者乘积是单位矩阵。实际中乘积与单位矩阵的差值往往不为 0，求这个差值的无穷范数作为评价误差的指标。矩阵差值越小越好。

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。

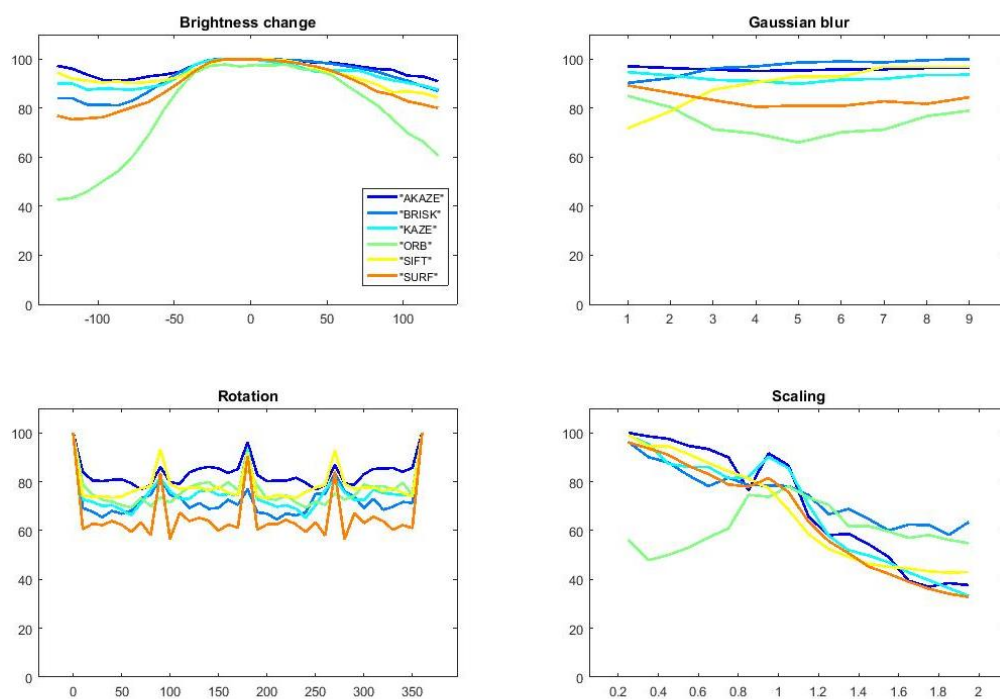
平均距离指的是将待配准图像按照求出的矩阵的逆变换回去后，特征点与原图对应点的距离的均值。平均距离越小，说明配准的精度越高。

峰值信噪比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。

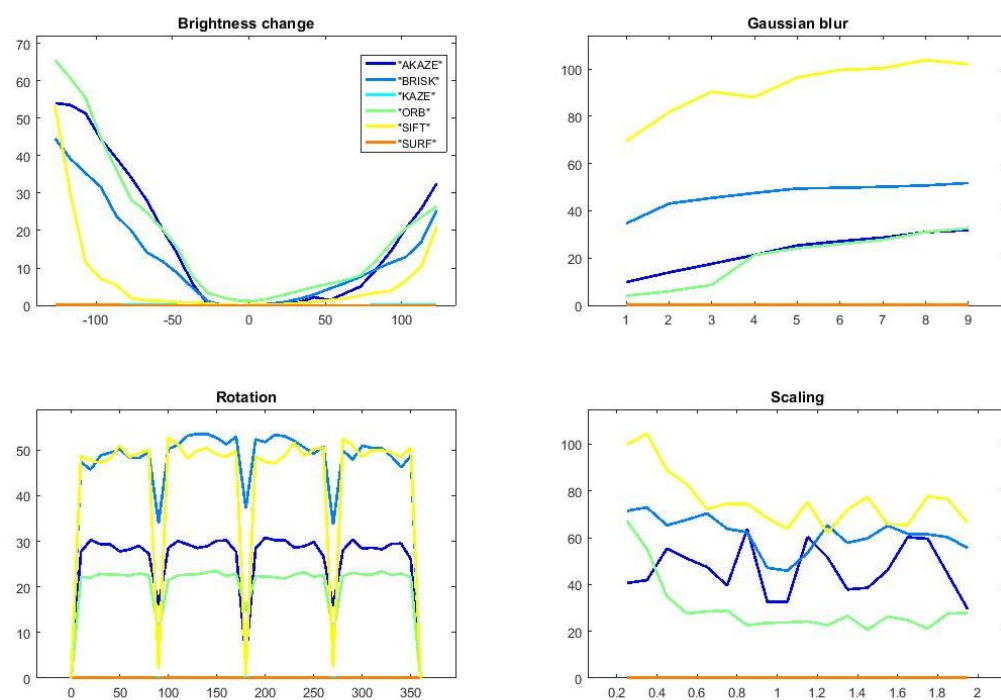


(a) HomographyError

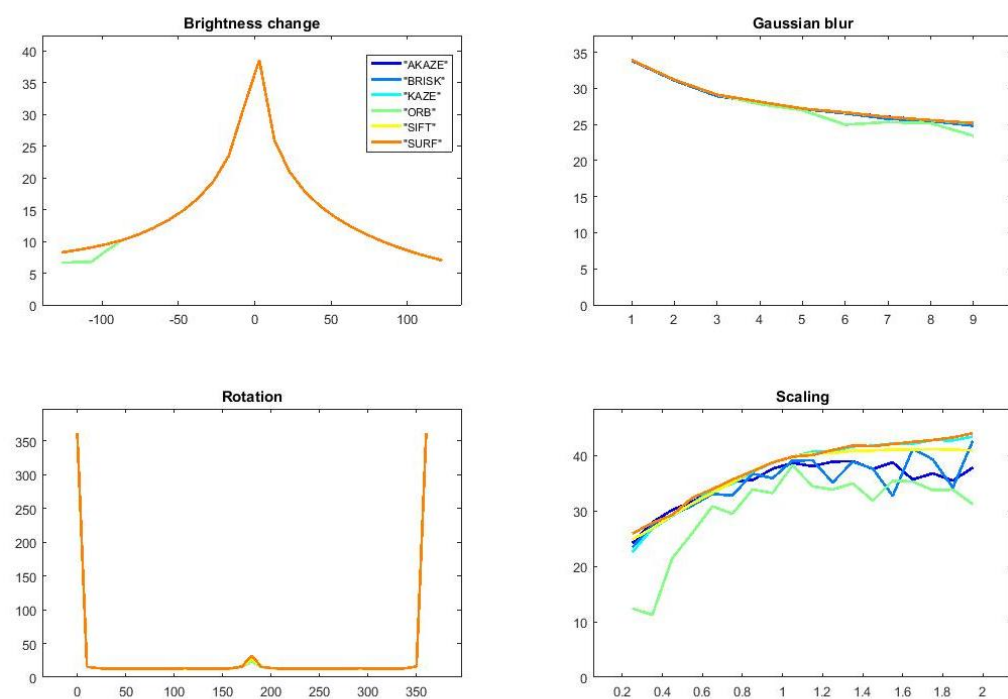


(b) correctMatchesPercent

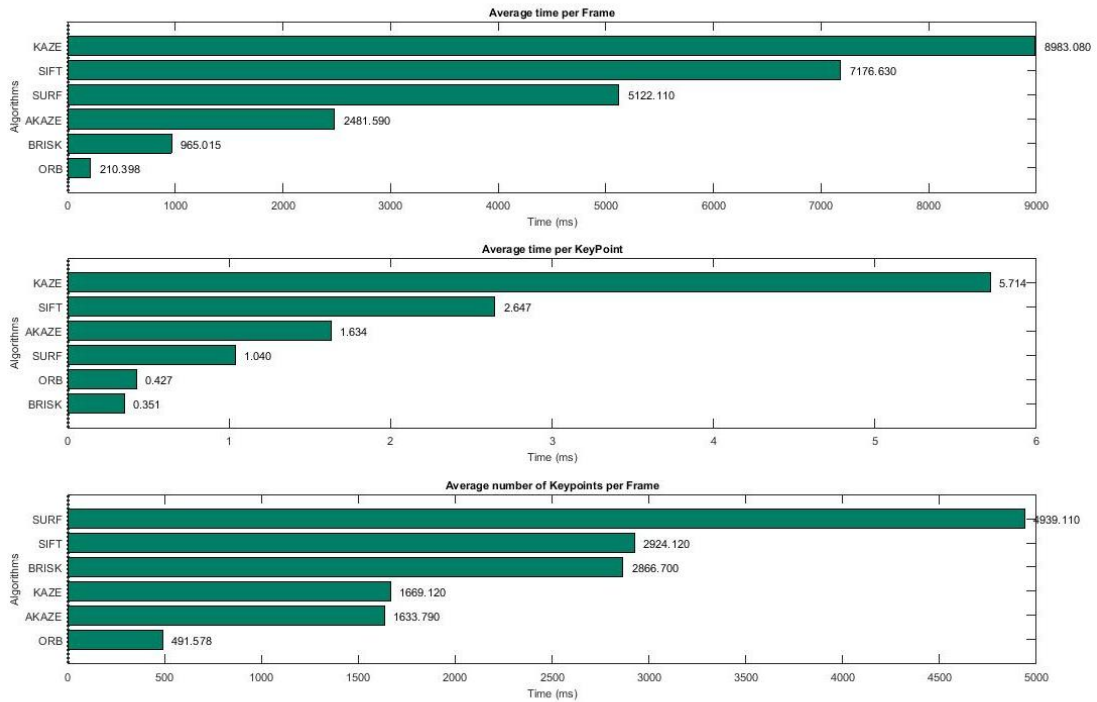




(c) Meandistance



(d) PSNR



(e) speed

图 3-1 几种主流特征点配准算法的对比

图 3-1 展示了几种算法在不同图像变换条件下的表现。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。检测处的特征点数目方面，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。

### 3.4. 小结

本章首先介绍了基于特征点配准的算法的基本流程。在 SIFT 算法被提出之后，图像配准的基本流程就基本被确定了，但是不同算法针对不同的环节作出了改进。本章介绍了主流的几种基于特征点的配准算法，如 SIFT、SURF、KAZE、AKAZE、ORB。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，对比了不同变换下不同算法的表现，使用了不同的指标对表现情况进行衡量。

## 第 4 章 基于 SIFT 算法的新算法

### 4.1. 低复杂度的 SIFT

#### 4.1.1. 理论思想

图像金字塔的构建是使用不同尺度因子的高斯卷积核对图像模糊来达到尺度不变性，卷积的计算量较大，计算量与图像的大小成正比。对于图像金字塔，计算量与金字塔的大小成正比。若能缩小金字塔的大小，就可以成比例地减少计算量。因为 SIFT 算法在构建金字塔的过程中使用了下采样，这里提前对实验图像进行下采样的预处理，考虑到下采样的插值方法和分别对待配准图像下采样还是同时对 2 幅图像下采样的情况进行了讨论。

若是单独对待配准图像进行下采样的预处理，在拟合得到投影变换模型之后，可以直接将变换作用在预处理之后的图像上，从而得到配准之后的图像；若是对待配准图像和参考图像都进行了下采样的预处理，则需要根据下采样的比例系数对拟合出的变换模型进行校准，得到适用于原图的变换关系。设任意 1 幅图像放大后的目标图像的宽为  $D_w$ ，高为  $D_h$ ，原图像宽为  $S_w$ ，高为  $S_h$ ，则目标像素点  $(D_x, D_y)$  在原图像上对应的位置为：  $S_x = D_x S_h / D_h$ ，  $S_y = D_y S_w / D_w$ 。当下采样比例为  $s$  时，待

配准图像到参考图像的变换矩阵  $\mathbf{H}'_p = \begin{pmatrix} a_{11} & a_{12} & a_{13}/s \\ a_{21} & a_{22} & a_{23}/s \\ sa_{31} & sa_{32} & a_{33} \end{pmatrix}$ 。因此，由中间变量  $\mathbf{H}_p$  进行

修正可得到最终待配准图像与参考图像之间的变换矩阵  $\mathbf{H}'_p$ ，利用  $\mathbf{H}'_p$  和插值（参考 Rob Hess 源码采用线性插值）可以实现配准<sup>[89]</sup>。

#### 4.1.2. 实验结果

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。

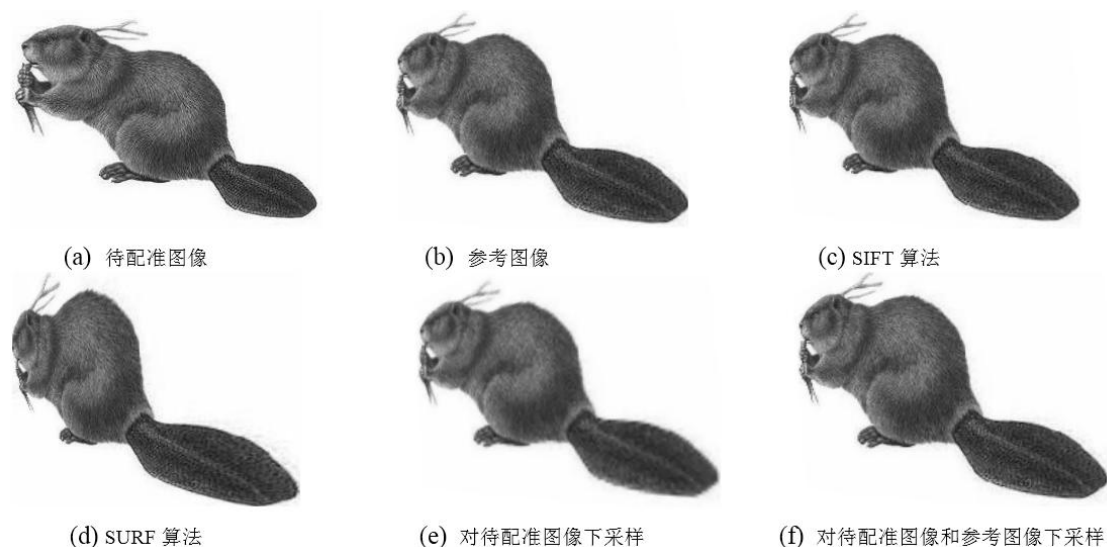


图 4-1 下采样缩小图像金字塔之后的图像配准

直观上可以看出 SURF 算法较差。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,同时比较使用不同插值方法时的效果,包括最近邻插值、双线性插值、区域插值、双立方插值方法,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 4-1 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			
			最近邻 插值	双线性 插值	区域 插值	双立方 插值	最近邻 插值	最近邻 插值	区域 插值	双立方 插值
$t/s$	0.858	1.421	1.029	1.083	1.084	1.100	0.513	0.537	0.557	0.507
$R/dB$	20.48	38.72	31.09	36.82	35.10	34.66	27.17	31.21	32.51	31.61

从表中可以看到,在只对待配准图像预处理时,使用双线性插值方法综合效果最优,牺牲 2dB 信噪比的条件下可以缩短 23.7%左右的运行时间;当对两幅实验图像均做预处理时,使用区域插值和双立方插值方法具有最好的表现,在保证峰值信噪比在 31dB 以上时还将运行时间缩短了 64.8%,两个指标都好于 SURF 算法。

## 4.2. SIFT与显著性检测

SIFT 算法的关键是检测出具有不变性,且可以代表图像的关键信息的特征点,

依靠两幅图像之间的特征点进行匹配，拟合出某个策略下的最优变换模型，所以特征点的质量和数量都很重要。

显著性区域检测方法模仿了人眼优先锁定感兴趣重要区域的特点，可以应用在目标检测和识别中，一定程度上减少搜索的工作量。将显著性检测方法和 SIFT 算法结合起来，前者作为后者的预处理，可以在一定程度上促使 SIFT 算法检测出更加稳定，高质量的特征点，从而有利于后续的变换模型估计和最终的配准效果。

显著性区域检测方法有多种，比如 X Hou 和 L Zhang 通过分析输入图像的对数谱来消除背景<sup>[90]</sup>，Yun Zhai 提出了一种关于时空视频的注意点的技术，而 Achanta 的方法输出具有明确界定的显著性物体的全分辨率显著性图<sup>[91]</sup>。SIFT 还有一个缺陷是算法的使用对象是灰度图像，即在图像处理的过程中丢失了图像的彩色信息。Frequency-tuned 显著性检测方法使用了 CIRLab 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的值表现为 CIRLab 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(用  $5 \times 5$  的独立二值核),  $\|\cdot\|$  是二范数。FT 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 CIE 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。

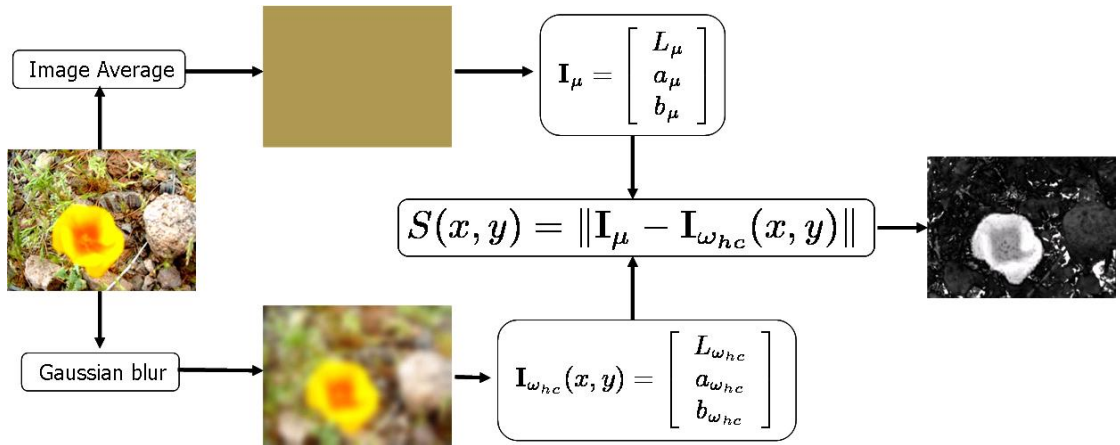


图 4-1 FT 显著性检测方法图示

### 4.2.1. 显著性图上的配准

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。

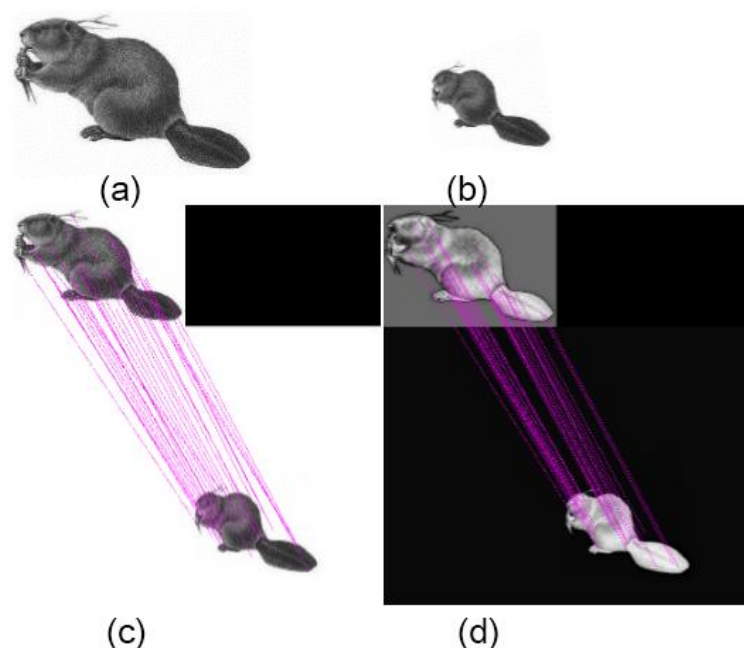


图 4-1 显著性图上的特征点检测与匹配

下表是图 4-1 中特征点数目、运行时间和配准效果的统计结果。这里统计了配准后图像与参考图像之间的峰值信噪比和特征点之间的均方根误差 RMSE。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	353	167	2140	43.920	0.556

从实验结果可以看出，基于显著性图的 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，代价是减速。

#### 4.2.2. 分割图上的特征点

基于FT的显著性检测方法还可以实现目标物体的分割。分割原理如下图所示：

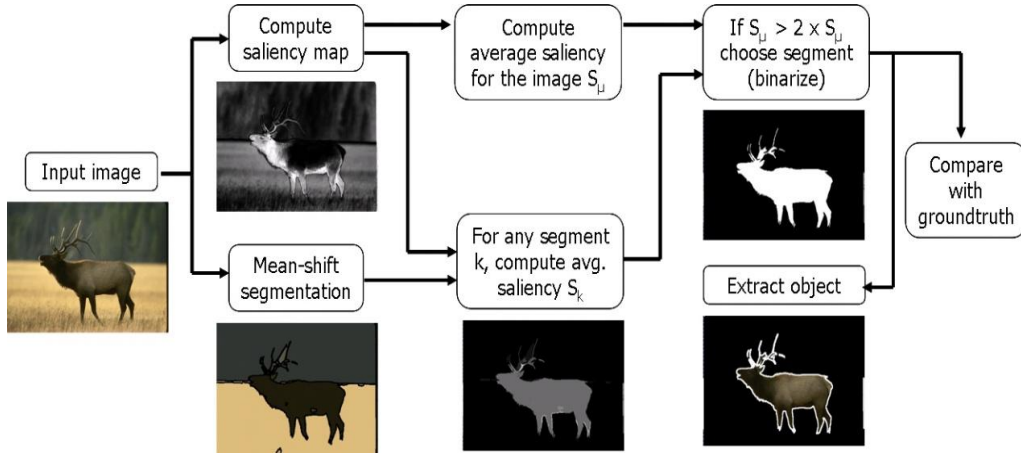


图 4-2 基于显著性检测的图像分割方法

当我们的配准目标也是图像中的待检测分割目标时，可以将这一图像分割方法应用在图像配准中，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。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。下面是实验结果：





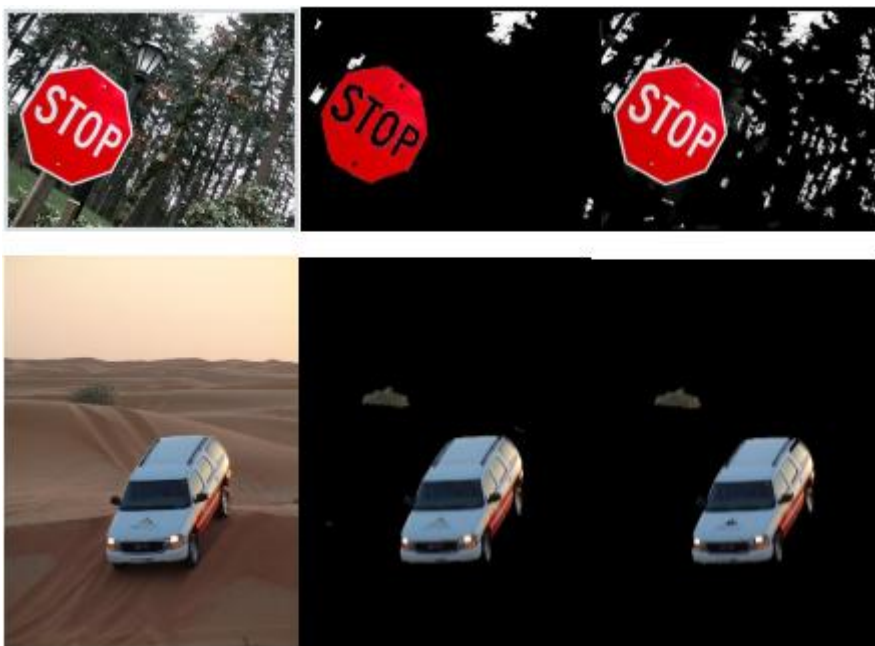
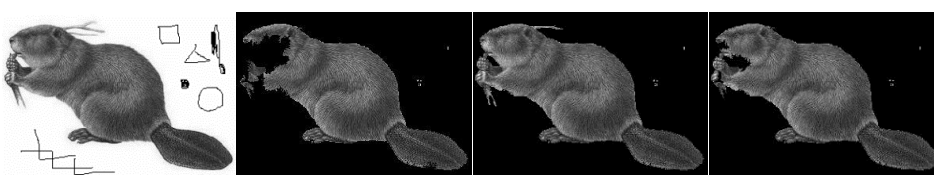


图 4-3 改进前后的图像分割方法

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。

利用人工涂鸦的图像可以说明分割后特征点检测的优势。下图是实验结果，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 4-4 三种分割方法的性能。

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。

### 4.3. 小结

本章介绍了两种基于 SIFT 算法的改进算法。第一种针对了 SIFT 计算量大的特点,利用下采样作为预处理从而达到了缩小图像金字塔的效果。本文对实行下采样的图像和使用的插值方式进行了讨论,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。第二种改进算法针对的是 SIFT 算法没有充分利用图像颜色信息的特点,将 SIFT 与显著性区域检测的方法结合起来,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志点的干扰。

## 第5章 配准算法的应用

### 5.1. 相关工作

#### 5.1.1 哈希算法

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。

加密哈希算法有 MD5,SHA1,SHA256 等，加密哈希有一个特点就是文件一旦发生改变，即便变化很小，哈希值都会发生巨大的变化。这个特点在防止用户修改文件时很有帮助，这也是加密哈希名称的由来。算法还有一个特点，就是不可逆，不可逆根据哈希值反向推出原信息，这也有利于加密，网站保存我们的密码就不会直接保存原始密码，而是保存其哈希值，这样即使网站被黑，依然不会泄露密码。

但同时很多时候我们希望相似图像或者相似文本具有相同的哈希值，这时候我们就需要采用其他的哈希算法。比较简单、易用的解决方案是采用感知哈希算法（Perceptual Hash Algorithm）。感知哈希算法是一类算法的总称，包括 aHash（平均值哈希）、pHash（感知哈希）、dHash（差异值哈希）。顾名思义，感知哈希不是以严格的方式计算 Hash 值，而是以更加相对的方式计算哈希值，因为“相似”与否本来就是相对的而不是绝对意义上的。

平均值哈希方法计算简单，只需在归一化图像尺寸后，比较每个像素与整体灰度图像的灰度均值，从而就可以得到 0、1 的二进制字符串，作为原图像的哈希值。但是平均哈希算法的准确度不高，因为它只记录了每个像素与均值的大小情况。这里采用了一种称为差异值哈希的算法，在保证快速的同时，具有良好的精确性。

差异值哈希 aHash 与平均值哈希 aHash 大体相同，不同的是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。不同图像的哈希值应该是相同的长度，这里采用的都是长度为 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。因为灰度值比较的参考值不同，平均值哈希算法中将图像归一化为 8x8 的大小，差异值哈希算法将图像的宽归一化为 9，将高归一化为 8，这样最后按行拼接得到的二进制字符串长度都是 64。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。

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。相比于加密哈希, 差异值哈希方法对图像微小变化的鲁棒性更强, 相似性越高的图像其哈希值之间的汉明距离越小。

### 5.1.2. 图像压缩

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。

## 5.2. 理论思想

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。由于二者之间存在较多的重复信息, 记录两

幅相似图像之间的变换关系,就可以得到配准后的图像,所以我们只需要记录配准后的图像与原图的差异信息就可以恢复出原图。而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。

两幅图像之间的减法指的是两幅大小相同的图像按照对应像素做减法运算,可以简单地得到两幅图像之间像素级别的差异信息。两幅尺寸大小相同的图像做减法,本质上是矩阵的减法,只需保证矩阵的大小一致。在图像配准中,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。而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。

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的重合度越高,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。

### 5.3. 实验结果

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。本文主要进行了两组实验。第一组的实验图像为 KAZE 算法作者提供的“box”图像,第二组为牛津大学特征检测项目组提供的“graffiti”涂鸦图像。两组图像中发生的变换主要是视角变换,并且在第一组的图像中还出现了遮挡的情况。实验中使用了 OpenCV 库中实现的 AKAZE 算法来进行图像的特征点检测,最后使用透视变换完成图像配准。

下面图 5-1 和 5-2 分别是两组实验图像的实验结果。(a)、(b)分别是待配准图像和场景中的参考图像。(c)是配准之后的结果。(d)、(e)分别是参考图像和配准后图像做减法之后,记录差值正负属性的掩膜图像和记录差值的绝对值的差值图像。图 f)是通过差值图像对参考图像,即认为可以舍弃的冗余图像的复原结果。

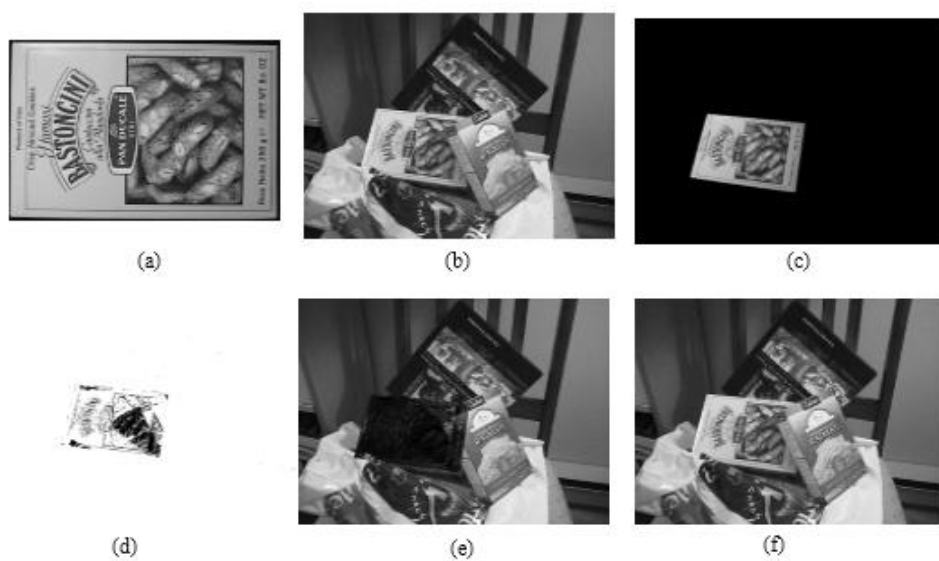


图 5-1 配准后图像与参考图像的差值和复原结果 (box 图)

在上图中的(e)中可以看到, 在经过图像减法之后的图像出现了较多的接近于 0 的值, 这意味着与图像相对应的矩阵变得更加稀疏。当图像之间包含的内容重复度更高时, 如图 5-2 中的示例, 这一点表现得更加明显。

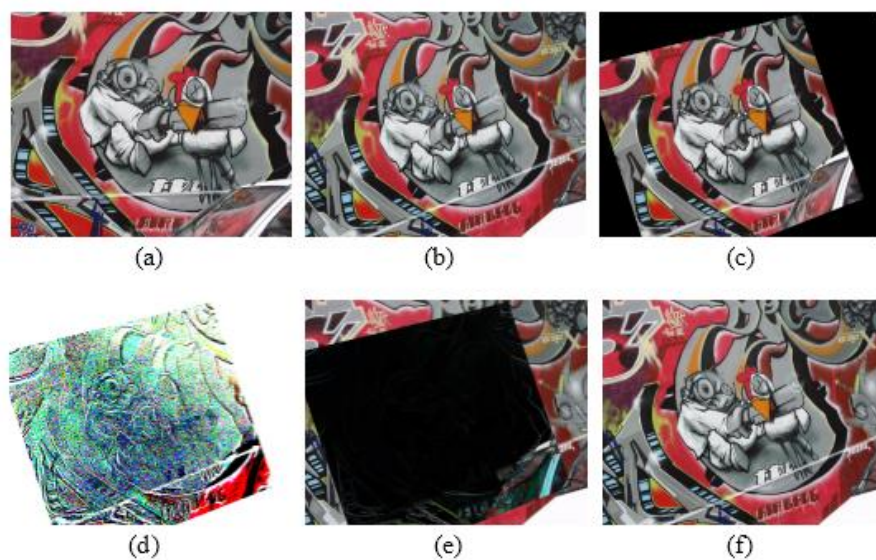


图 5-2 配准后图像与参考图像的差值和复原结果 (graffiti 图)

下表记录了两组实验中更多的数据, 如哈希值, 图像占据存储空间的大小 (以字节为单位) 等。实验中, 使用了 Python 中的 PII 模块来读取图像, 并将图像转换为字符串来计算图像占据的字节大小。使用模块 zlib 对字符串进行无损压缩。从上表中可以看出, 在保证无损压缩的情况下, 拥有更多低像素值, 更稀疏的差值图

像可以得到更大的压缩比。

表 5-1 图像配准和减法前后的压缩情况比较

	图像名称格式	汉明距离	原始图像大小	差值图像大小	原图压缩	差值压缩
Box 图像	box.jpg	16	216756	589842	----	214548
	box_in_scene.jpg		589842		215228	
涂鸦图像	graf1.ppm	15	1536000	1536000	-----	1129109
	graf2.ppm		1536000		1389097	

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，所以如果同时保存两张图像会在一定程度上造成信息冗余。对于明显存在某种变换的图像，可以使用图像配准拟合出二者之间的关系，用配准后的图像在一定程度上替换参考图像，甚至直接用拟合出的变换矩阵表示冗余的参考图像：变换矩阵与保留的待配准图像可以恢复参考图像中的主体部分，对于配准产生的误差与参考图像无法恢复的部分，借助图像减法记录这一部分信息。

#### 5.4. 小结

在向服务器上传或者向本地拷贝图像时，我们不希望无限制地允许图像集增大，对于重复性的图像我们希望保证其单一性，而不是无谓地占用内存空间，但直接比较两个图像的相似性有时过于复杂，一种可取的方法就是提前计算得到图像的哈希值，对于两个图像各自的哈希值的汉明距离为 0，就认为两个图像完全一样；当汉明距离不为 0 但较小时，认为两幅图像相似性程度较高，这时可以利用配准算法估算出二者之间的变换关系，为了便于之后可能的复原，还可以记录待删除图像（即配准过程中的参考图像）与配准之后的图像的差值图像和其对于像素点的正负性。

## 第 6 章 总结与展望

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。从图像的成像模型到图像坐标系，到描述图像变换的模型和图配准的评价标准。估算出的模型的准确性直接关系到配准的效果。
- (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 算法提出了两种改进的算法。一种是使用下采样减小图像金字塔的规模，降低了图像配准在特征点提取中的计算复杂度。一种使用了显著性检测的方法，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。将基于显著性检测的图像分割方法应用在图像配准中，并对其分割方法做出了改进，可以保留更多的主体信息。



- (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。

图像配准其实是一个很大的课题，涉及到的知识点众多。由于个人水平和时间有限，没有能做出特别满意的工作，未来还有很长的路要走。本文的待改进之处有：

- (1) 文章提出的一些思路还停留在实验阶段，各种优化工作做得不够完善，距离真正的实验还有一段距离。
- (2) 对图像配准算法本身的改进有限。现在图像算法的发展已经广泛使用了深度学习，如何将神经网络和深度学习应用在图像配准方向有待发掘。

## 参考文献

- [1] Hubel, David H., and Torsten N. Wiesel. "Receptive fields of single neurones in the cat's striate cortex." *The Journal of physiology* 148.3 (1959):574-591.
- [2] Wiesel, Torsten N. "The postnatal development of the visual cortex and the influence of environment." Stockholm: Nobel Foundation (1981).
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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/OL].西安交通大学学报,2019(02):1-6[2019-01-03].<http://kns.cnki.net/kcms/detail/61.1069.T.20181206.1718.004.html>.
- [7] T. F. Cootes. C. J. Taylor. Statistical models of appearance for computer vision: World Wide Web Publication, 2004
- [8] Zhang K , Zuo W , Chen Y , et al. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising[J]. IEEE TRANSACTIONS ON IMAGE PROCESSING.
- [9] Schaffalitzky F, Zisserman A. Multi-view Matching for Unordered Image Sets, or "How Do I Organize My Holiday Snaps?"[J]. Proc.european Conf.comput.vision May Copenhagen Denmark, 2002, 2350:414-431.
- [10] Tuytelaars T , Gool L V . Matching Widely Separated Views Based on Affine Invariant Regions[J]. International Journal of Computer Vision, 2004, 59(1):61-85.
- [11] Zitova B, Flusser J. Image registration methods :a survey[J].Image and Vision Computing. 2003,21(11),pp:977-1000.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] Lisa Gottesfeld Brown . A Survey of Image Registration Techniques . ACM Computing Surveys[J] , 1992 , 24(4):325~376 .
- [17] Gottesfeld Brown L . A survey of image registration techniques[J]. Acm Computing Surveys, 1992, 24(4):325-376.
- [18] 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.
- [19] 孙雅琳, 王菲. 基于边缘和互信息的红外与可见光图像配准[J]. 电子科技, 2010, 23(4):80-

82.

- [20] 张强. 图像匹配算法研究[D]. 西安电子科技大学, 2006.
- [21] 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.
- [22] 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.
- [23] 沈晓芳, 向建勇, 董维科. 基于图像边缘特征的 SSDA 算法[J]. 电子科技, 2009, 22(3):16-18.
- [24] 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.
- [25] 贾凯, 曲仕茹. 改进的 SSDA 图像匹配算法[J]. 测控技术, 2012, 31(10):47-50.
- [26] Wang Y, Wu D. An improved adaptive SSDA based on wavelet-pyramid[C]// American Institute of Physics Conference Series. AIP Publishing LLC, 2017.
- [27] P.A. Viola, Alignment by maximization of mutual information, Ph.D. thesis, Massachusetts Institute of Technology, Boston, MA, USA, 1995
- [28] A.Collignon, F.Maes, D.Delaere, D.Vandermeulen, P.Suetens, and G.Marchal, Automated multi-modality image registration based on information theory, Information Processing in Medical Imaging, Y. Bizais, C. Barillot, and R. Di Paola, Eds. 1995, Kluwer Academic Publishers, 1997, 187-198
- [29] 马政德, 基于互信息的图像配准并行算法研究与实现, 国防科技大学工学硕士论文, 2007
- [30] 曹蹊渺, 基于互信息的图像配准算法研究, 北京交通大学硕士学位论文, 2007
- [31] 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
- [32] 靳济芳. VC++小波变换与工程实践. 北京:人民邮电出版社., 2004, 第一章
- [33] S.Mallat. A Theory for Multiresolution Signal Decomposition: the Wavelet Representation. IEEE TRANSACTIONS on Pattern Analysis and Machine Intelligence, 1989, 11(7):674-693
- [34] 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
- [35] 夏召红, 基于小波变换的图像配准, 辽宁科技大学硕士学位论文, 2008
- [36] Hough P V C. Method and means for recognizing complex patterns: U.S. Patent 3,069,654[P]. 1962-12-18.
- [37] 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.
- [38] H. Gonçalves; J. A. Gonçalves; L. Corte-Real; A. C. Teodoro. CHAIR: automatic image registration based on correlation and Hough transform [J]. International Journal of Remote Sensing, 2012, 33 (24 ):7936-7968.
- [39] 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.

- [40] Hu M K. Visual pattern recognition by moment invariants[J]. Information Theory, IRE TRANSACTIONS on, 1962, 8(2): 179-187.
- [41] 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.
- [42] Dai X, Khorram S. A feature-based image registration algorithm using improved chain-code representation combined with invariant moments[J]. Geoscience and Remote Sensing, IEEE TRANSACTIONS on, 1999, 37(5): 2351-2362.
- [43] Flusser J. Moment invariants in image analysis[C]//proceedings of world academy of science, engineering and technology. 2006, 11(2): 196-201.
- [44] Flusser J, Suk T. Rotation moment invariants for recognition of symmetric objects[J]. Image Processing, IEEE TRANSACTIONS on, 2006, 15(12): 3784-3790.
- [45] Kristensen F, Mac Lean W J. Real-time extraction of maximally stable extremal regions on an FPGA[C]//Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on. IEEE, 2007: 165-168.
- [46] Tuytelaars T, Van Gool L. Matching widely separated views based on affine invariant regions[J]. International journal of computer vision, 2004, 59(1): 61-85.
- [47] Kadir T, Brady M. Saliency, scale and image description[J]. International Journal of Computer Vision, 2001, 45(2): 83-105.
- [48] Kadir T, Zisserman A, Brady M. An affine invariant salient region detector [M]//Computer Vision-ECCV 2004. Springer Berlin Heidelberg, 2004: 228-241.
- [49] Mikolajczyk K, Tuytelaars T, Schmid C, et al. A comparison of affine region detectors[J]. International journal of computer vision, 2005, 65(1-2): 43-72.
- [50] Mikolajczyk K, Schmid C. An affine invariant interest point detector[M]//Computer Vision—ECCV 2002. Springer Berlin Heidelberg, 2002: 128-142.
- [51] Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors[J]. International journal of computer vision, 2004, 60(1): 63-86.
- [52] Schaffalitzky F, Zisserman A. Multi-view matching for unordered image sets, or “How do I organize my holiday snaps?”[M]//Computer Vision—ECCV 2002. Springer Berlin Heidelberg, 2002: 414-431.
- [53] Hans P M. Towards automatic visual obstacle avoidance[C]//Proceedings of the 5th international joint conference on Artificial intelligence. 1977: 584-584.
- [54] Harris C, Stephens M. A combined corner and edge detector[C]//Alvey vision conference. 1988, 15: 50.
- [55] Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion[J]. Pattern Analysis and Machine Intelligence, IEEE TRANSACTIONS on, 1990, 12(7): 629-639.
- [56] 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.
- [57] 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.

- [58] Lindeberg T. Edge detection and ridge detection with automatic scale selection[J]. International Journal of Computer Vision, 1998, 30(2): 117-156.
- [59] Lindeberg T. Feature detection with automatic scale selection[J]. International journal of computer vision, 1998, 30(2): 79-116.
- [60] Mikolajczyk K, Schmid C. An affine invariant interest point detector[M]//Computer Vision—ECCV 2002. Springer Berlin Heidelberg, 2002: 128-142.
- [61] 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.
- [62] 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.
- [63] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International journal of computer vision, 2004, 60(2): 91-110.
- [64] Bay H, Tuytelaars T, Van Gool L. SURF: Speeded up robust features[M]//Computer Vision—ECCV 2006. Springer Berlin Heidelberg, 2006: 404-417.
- [65] 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.
- [66] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors[J]. Pattern Analysis and Machine Intelligence, IEEE TRANSACTIONS on, 2005, 27(10): 1615-1630.
- [67] 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.
- [68] 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
- [69] 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
- [70] 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
- [71] Vinividyadharan, Subusurendran, 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
- [72] Rosten E, Drummond T. Machine learning for high-speed corner detection [M] // Computer Vision—ECCV 2006. Springer Berlin Heidelberg, 2006: 430-443.
- [73] 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.
- [74] Taylor S, Rosten E, Drummond T. Robust feature matching in 2.3  $\mu$ s[C]//Computer Vision and

- Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on. IEEE, 2009: 15-22.
- [75] Taylor S, Drummond T. Binary histogrammed intensity patches for efficient and robust matching[J]. International journal of computer vision, 2011, 94(2): 241-265.
- [76] Alcantarilla P F, Bartoli A, Davison A J. KAZE features[M]//Computer Vision–ECCV 2012. Springer Berlin Heidelberg, 2012: 214-227.
- [77] 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
- [78] 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.
- [79] 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
- [80] 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
- [81] 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
- [82] 佚名. Three.js 开发指南[M]. 2015.
- [83] Hill F, Kelley S. 计算机图形学:OpenGL 版[M]. 清华大学出版社, 2009.
- [84] Bouchiha R, Besbes K. Automatic remote-sensing image registration using SURF[C]. 2010 The 3rd International Conference on Machine Vision (ICMV 2010), 2010: 406-410.
- [85] 阮芹彭刚李瑞. 基于特征点的图像配准与拼接技术研究[J]. 计算机与数字工程, 2011, 32(2): 141-144.
- [86] STUDHOLME C, HILL L G, HAWKES D. An overlap invariant entropy measure of 3D medical image align-ment [J] .Pattern Recognition, 1988, 32( 1 ) : 43-46.
- [87] P Perona , J Malik. Scale-Space and Edge Detection Using Anisotropic Diffusion. IEEE Computer Society , 1990 , 12 (7) :629-639
- [88] J.Weickert; B.M. T. H.Romeny;M.A. Viergver. Efficient.Efficient and reliable schemes for nonlinear diffusion filtering. IEEE TRANSACTIONS on Image Processing.1998
- [89] 张晨光, 周诠, 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.
- [90] 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.
- [91] 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.

## 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，去往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，推动了新的领域的大门，也收获了珍贵的友情。

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。回征师兄还有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。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，大家平常在生活和学习中互帮互助，那无疑是一段值得怀念的时光。

最后需要特别感谢的是我的家人。可怜天下父母心，是我的母亲劝说我攻读的研究生，而父亲还有外公外婆则一直在背后默默支持。如今他们年纪大了，遇事反而要找我们商量，我一定要让他们继续为我而骄傲。感谢我弟弟的成长与懂事，为我负担家里的大小事情。还要感谢的是我的女朋友张晶，一个爱笑爱哭的女孩。谢谢张晶为我的生活增添了许多色彩，也谢谢她对我的激励，更感谢她对我的宽容，原谅了我的种种疏忽和冲动，希望我们能早日安居乐业。

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，我也不知道等待我的会是怎样的一段故事。我只知道在我的梦想里，少不了每一个我爱和爱我的人，只有安身立命，怀揣着梦想，因上努力。

## 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] 《A novel image registration algorithm based on salient region and SIFT》以第一作者被 The 3rd International Society for Artificial Intelligence and Robotics (ISAIR 2018) 录用
- [3] 一种适于微纳卫星的分块数据压缩方法,专利号:201810634416.7。发明人:周詮、回征、张晨光、呼延琅。
- [4] 《基于 SIFT 特征点检测的一种改进的低复杂度图像配准算法》以第一作者被第十九届全国图象图形学学术会议(NCIG2018) 录用。