



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 俭

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一项非常重要的技术，它是模式识别、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，对遮挡，噪声，图像形变的适应性和鲁棒性更好。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：

（1）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，重点介绍了几种基于特征点配准的算法的流程，将图像配准算法应用在不同变换情况下的图像，从不同方面对不同算法进行了系统的比较。除了图像配准技术本身，图像表示和图像变换模型的准确性直接关系到配准的效果。所以本文在一开始研究了相机模型、图像的坐标系和坐标，双目视觉问题，图像的变换模型。

（2）SIFT 算法是特征点配准算法中的经典算法，被广泛应用。本文针对 SIFT 计算量过大导致的配准较慢的问题和没有利用图像色彩信息的缺点，分别做出了两方面的改进。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，在增加的运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，这在只要求对主体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，及时避免无意义的配准，本文引入了哈希算法，分别计算得到两幅图像的哈希值，当哈希值之间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. 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.

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

目录

第 1 章 绪论	1
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	9
第 2 章 基于特征点图像配准的技术分析	11
2.1. 数字成像基础	11
2.2. 图像变换模型	14
2.3. 配准流程	15
2.4. 算法对比实验	17
2.4.1 实验设计	18
2.4.2 实验指标	20
2.4.3 实验结果与分析	22
2.5. 本章小结	25
第 3 章 基于 SIFT 的低复杂度配准	26
3.1. 算法原理	26
3.1.1 图像高斯金字塔	26
3.1.2 插值方法	27
3.1.3 单下采样	29
3.1.4 双下采样	29
3.2. 实验结果与分析	30
3.3. 本章小结	31
第 4 章 SIFT 算法与显著性检测	32
4.1. 显著性图上的特征点检测	32
4.1.1. 显著性检测 FT 算法	32
4.1.2. 对 FT 的仿真	33
4.2. 显著性图上的配准实验	34
4.2.1. 实验理论	34
4.2.2. 实验结果与分析	35
4.3. 分割图上的特征点	36

4.3.1 实验理论	36
4.3.2 实验结果与分析	37
4.4. 本章小结.....	39
第 5 章 配准算法的一种新应用	40
5.1. 理论思想.....	40
5.1.1. 哈希算法	40
5.1.2. 图像减法	42
5.1.3. 图像压缩	44
5.2. 实验设计.....	46
5.3. 实验结果与分析.....	46
5.4. 本章小结.....	50
第 6 章 总结与展望.....	51
参考文献.....	53
致 谢.....	58
硕士学位论文原创性声明	59

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，人主要通过视觉获得外界色彩、形状等直接关系到生存和正常生活的信息，视觉获得信息量远远超过了听觉、触觉等感官。生理学的研究为我们揭示了视觉的奥秘。D.H.Hubel 和 T.N.Wiesel 凭借对视觉系统的信息加工的研究获得了 1981 年的诺贝尔奖^{[1][2]}。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，人做出各种相应的动作。在这个过程中，图像由光信号了电信号。

而数字图像实质上是一个二维函数，坐标值作为自变量，灰度值作为因变量，二者都是离散值。上个世纪 20 年代，通过海底电缆从伦敦到纽约传输了第一张数字照片^[3]，这不仅是一次数字通信，更被认为是数字图像处理的起源：对图像进行数字编码，接收端再进行解码。所以数字图像也可以看作是信源编码。数字图像处理更广泛的定义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^[4]、图像增强、图像分割^[5]、图像压缩^[6]等。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，Zhang 等借助残差网络实现了图像去噪，图像解压缩等^[8]，计算机视觉的终极目标是实现图像理解^[7]。在数字图像处理之中，图像配准^{[8][9]}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^{[10][11]}。图像配准的输入是两幅图像，二者之间可能因为成像角度、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。

这项技术在遥感、模式识别、遥感、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^[12]；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适定问题进行了研究^[13]；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，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，同样需要图像配准技术；在模式识别领域，如车牌识别^[14]、人脸识别^[15]、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

在数学，图像配准上指的是两个二维函数之间的空间变换和灰度变换。两个二维函数 $f_1(x, y)$ 和 $f_2(x, y)$ 分别指参考图像和待配准的变形图像。

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

其中 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只关心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。

1992 年，L.G.Brown 在其论文^[16]中对近三十年来的图像的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。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。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。相比于输入空间，特征空间对图像进行抽象化和数字化表示，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。

(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。

(3) 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。

（4）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，但在实践中一般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。

1992 年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。本文中参考了谢凤英^[18]等主编的资料，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。

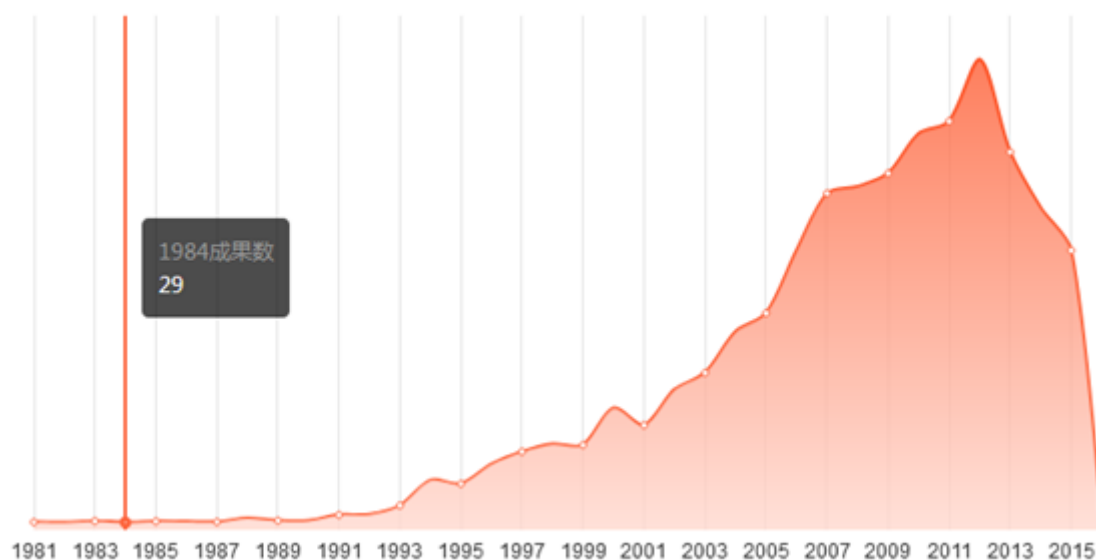


图 1-1: 百度学术统计的图像配准(Image Registration)在各个年份的成果数

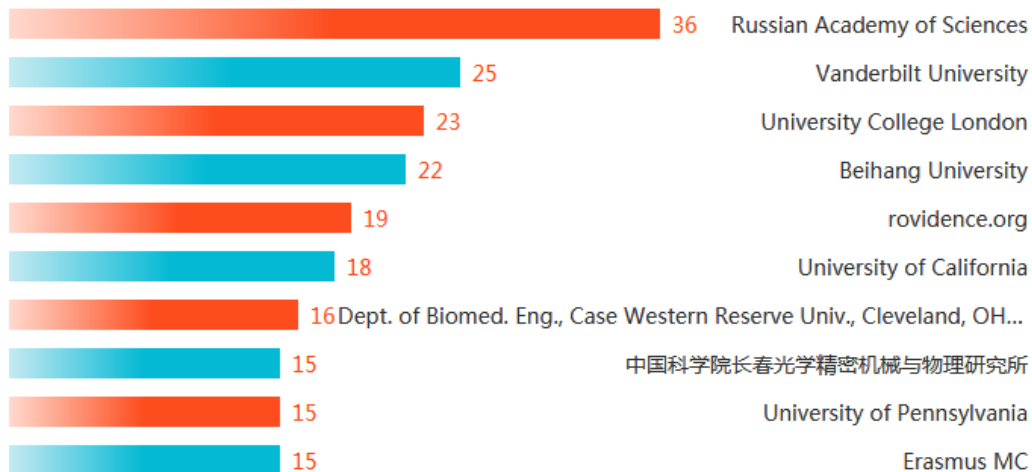


图 1-2: 百度学术统计的各个机构的研究成果数

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, 特征空间主要是图像中的像素的灰度信息。不同的匹配算法之间区别主要是相关性度量的不同, 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。平均绝对差算法 (Mean Absolute Differences, MAD), 在 1971 年被提出, 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^[19]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^[20]。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。根据阈值选取大致可以分成三种: 固定阈值、单调递增阈值和自适应阈值^[21]。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^[22]。SSDA 对灰度值的线性变化的鲁棒性较差, 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^[23]。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^[24]。1948 年, Shannon 首次提出互信息的概念。Viola^[25]和 Collignon^[26]等人在 1995 年将互信息引入图像配准, 解决异源图像的配准。国防科大的马政德互信息基础上提出并实现了一种的刚性图像配准并行算法^[27]。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将互信息与局部特征结合起来进行了研究。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^[28]。

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。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）。这是因为同样是正交变换，小波变换中的基函数是有限的，会衰减的小波基，傅里叶变换的基函数是无限长的单一频率的正弦函数。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^[29]。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^[30]。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^[31]。

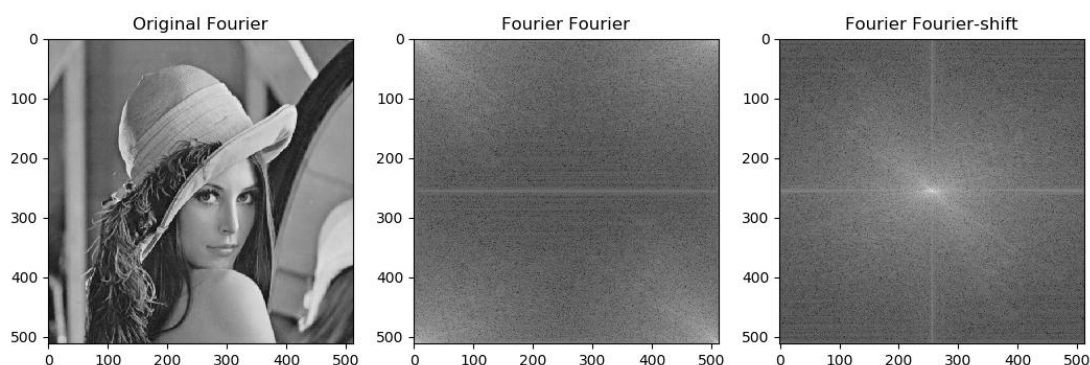


图 1-3 图像的傅里叶频谱表示

基于特征的图像配准方法将对整个图像的分析转化为提取图像特征之后对图像特征的分析，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。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^[32]、基于 Zernike 多项式的 Zernike 矩特征^[33]等。相关研究如 Xiaolong Dai 和 Siamak Khorram^[34]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^[35]利用图像中的不变矩特征，实现了目标识别和分类；Jan Flusser^[36]和 Tom á 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。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^[37]、Canny 算子边缘检测等^[38]。H. Gonçalves 等人针对遥感图像提出名为 CHAIR 的方法^[39]，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；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^[40]。

局部点特征是另外一种局部特征。1977 年，Moravec^[41]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^[42]，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。

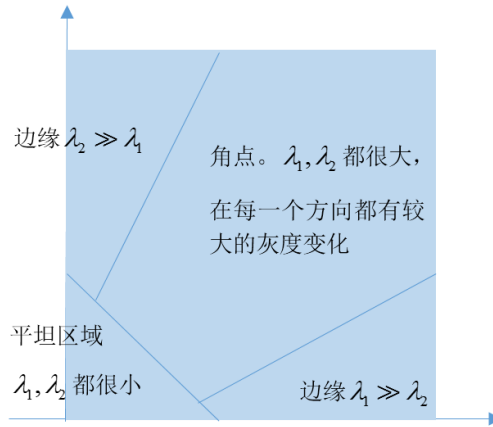


图 1-4: Harris 角点检测，由协方差矩阵的特征值判断是否是角点

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^[43-47]系统提出了图像的尺度空间理论。Mikolajczyk^[45]将 Harris 角点检测算法与高斯尺度空间结合，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，不仅完善了对特征点的检测，而且提出了一整套特征点描述和匹配的方法。Lowe 从 1999 年到 2004 年，Lowe 相继发表的三篇论文^[49-51]，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*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 **SAR** (Synthetic Aperture Radar) 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** 是一种改进算法，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，双边滤波的耗时影响了算法的实时性。

2006 年，Bay^[52]在 **SIFT** 算法的基础上，提出了一个实时性强的算法，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，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**^[53]、**GLOH**^[54]、**ASIFT**^[55]、**ORB**^[56]、**CSIFT**(colored **SIFT**)^[57]等其他改进的 **SIFT** 算法^{[58][59]}。**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，从而降低了特征描述子的维度，寻找起主要作用的分量，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 2006 年提出了 **FAST** (Features from Accelerated Segment Test) 角点检测算法^[60]，并在 2008 年进行了改进^[61]。**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。

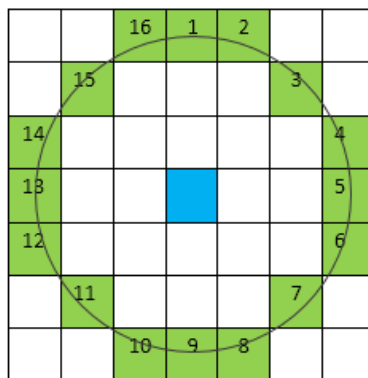


图 1-5 选取以像素点为中心的半径为 3 的离散化的 Bresenham 圆形区域，如果在圆周上有 N 个连续的像素的亮度比中心点大或小一个阈值，则判为角点。

Taylor、Rosten、Drummond^{[62][63]}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。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^[64]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,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^[65],实现了特征点之间的精确匹配。基于 SURF,潘建平等人引用相对距离理论,改进了图像配准算法^[66]。侯坚、张明利用蚁群算法对 SIFT 进行优化^[67],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^[68]。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^[69]。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^[70]。

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,但仍然存在一些亟待解决的问题:

(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下的国内外研究现状。

第二章的重点在基于特征点的图像配准方法上。首先对与之相关的基本技术进行了介绍，包括图像表示和图像的变换模型。配准的实质就是估计两幅图像之间的空间上的变换关系，具体可通过矩阵来表示。然后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，以 SIFT 算法为切入点，针对其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。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 基于特征点图像配准的技术分析

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进行变换矩阵估计的特点,受到了国内外众多学者的研究。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,并从不同方面进行比较,总结其大致的配准步骤。既然配准的核心是估计出两幅图像之间的变换模型的参数,所以在算法的对比之前有必要了解一些数字图像成像的原理,包括关于图像坐标的知识点,还有图像变换的原理,这对之后的研究大有裨益。

2.1. 数字成像基础

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,对应的相机模型分别是正交投影照相机(Orthographic Camera),和透视投影照相机(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,如下图中的(b)。一般说来,制图、建模用正交投影;在其他场景中通常使用透视投影,更接近人眼的观察效果经验^[71]。

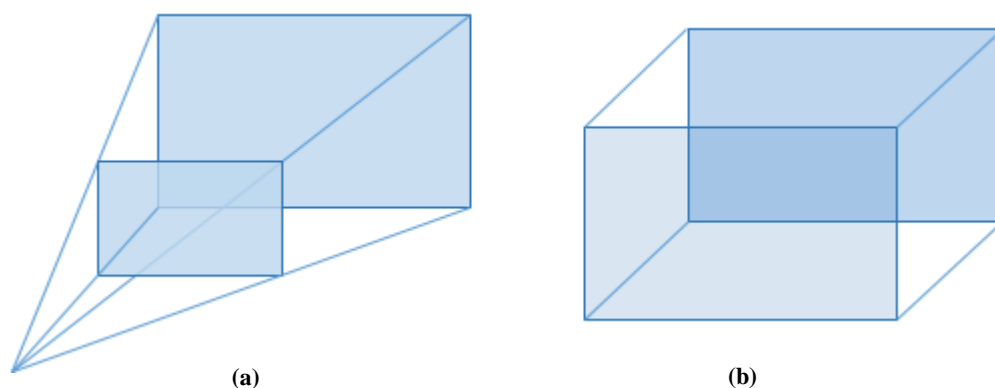


图 2-1 透视投影照相机和正交投影照相机

最常用的模型是针孔相机模型(Pinhole camera model),使用了透视投影,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。网络摄像头、

手机镜头、监控探头等都可以使用针孔相机模型近似。

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，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：

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

(3) 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可以通过相似三角形推导得到。

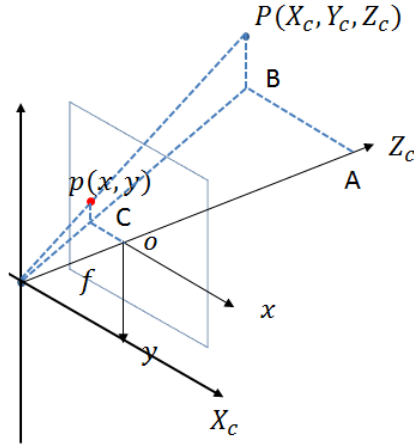


图 2-2 相机坐标系到图像物理坐标系的转换

式 3-3.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。

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

(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，用像素的数目来衡量。

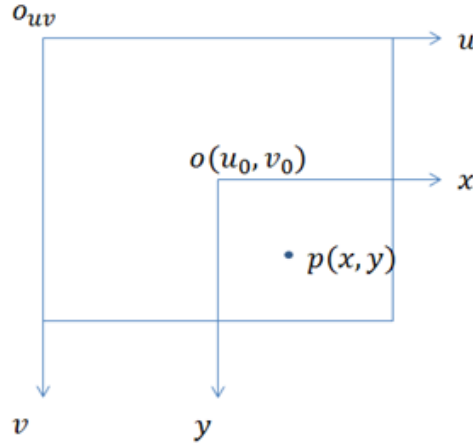


图 2-3 图像物理坐标系到像素坐标系的转换有两个要素

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 。

$$\begin{cases} u = \frac{x}{dx} + u_0 \\ v = \frac{y}{dy} + v_0 \end{cases} \quad (2-3)$$

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生成过程。

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刚才提到的，可以将式 3-1 这样包含平移的变换表示为矩阵相乘的形式：

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, R: 3 \times 3, T: 3 \times 1 \quad (2-4)$$

2.2. 图像变换模型

图像配准的本质是要估算出两幅图像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^{[73][74]}就是指描述两个图像空间变换关系的模型。现在介绍几种常用的模型。

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的顺序不同，结果也不同。

仿射变换（Affine Transformation）是线性变换和平移变换的复合变换。线性变换中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”。

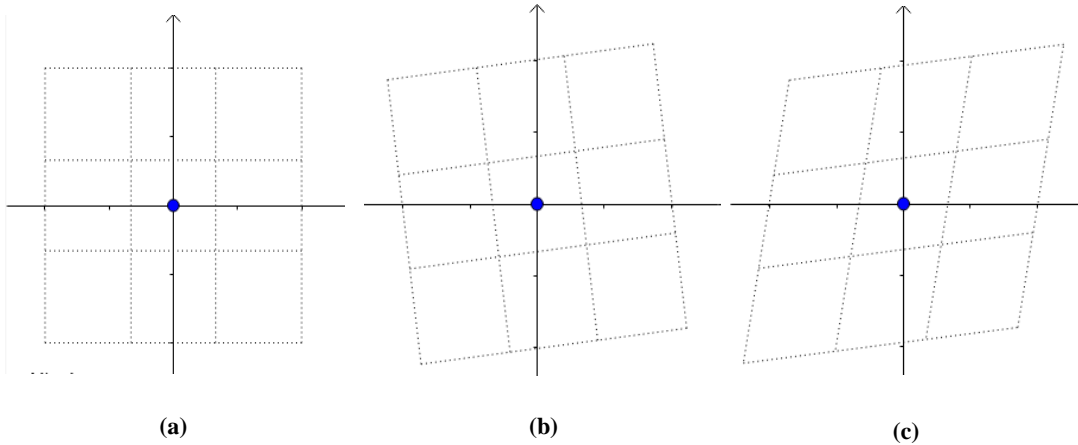


图 2-4 仿射变换中的线性变换（绕原点的旋转与“推移”）

仿射变换是计算机图形学中的一种重要变换，对于二维坐标，可以用下面公式表示：

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix} + B \quad (2-5)$$

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：

$M = [A \ B] = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \end{bmatrix}$, $T = M \cdot [x \ y \ 1]^T$ 。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。

投影变换（Projective Transformation）已经在前文中多次提到，是应用方程广泛的一种变换，在光源下，三维的物体经过投影变成一个二维的图像。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相同。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 3×3 大小的矩阵。

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。

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. 配准流程

SIFT 算法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得到匹配点对，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 2-5 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；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。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。

相对于原始图像，特征点所含的信息较少，但特征点除了包含了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。

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 算法来构建 KD 树来实现。

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 3×3 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。

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。插值方法有：最近邻插值、双线性插值、双三次插值等。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。

2.4. 算法对比实验

在 SIFT 算法提出之后针对 SIFT 的各种缺点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，所以没有哪一种算法是绝对完美的。本小结对各种算法进行了实现，在保证相同图像，相同变换的条件下对不同配准算法进行多方面的比较，以期分析得出不同算法的特性，加深对图像配准的认识，找到研究课题的突破点和着手点。

2.4.1 实验设计

本次对比实验选取了常见的几种基于特征点检测的图像配准算法，包括 SIFT、SURF、ORB、BRISK、KAZE、AKAZE。

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。对特征点的 3σ 邻域内所有像素的梯度信息进行直方图统计。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，幅度大于主峰值 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（ 4×4 ），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。

不借助于硬件的加速，SIFT 算法很难达到实时的程度。对于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。2006 年提出的 SURF（Speeded Up Robust Features）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。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 4×4 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。

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

Invariant Scalable Keypoints)。BRISK 在特征点检测上使用了 FAST 检测方法，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。

在 BRISK 算法提出的同年，OpenCV 实验室提出了 ORB 算法[56]。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。与 BRISK 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，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。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 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。

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。2012 年，提出了一种基于非线性尺度空间的特征点检测算法 KAZE^[64]。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：

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

其中 div 和 ∇ 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^[75]：

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

其中 L_σ 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^[76]。采用任意步长来构造稳定的非线性尺度空间。

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。

本次实验除了保证了算法的多样性，还考虑了图像变换类型的多样性。以牛津大学提供的图像集为基础，对其进行亮度变换，高斯模糊，旋转变换和尺度变换。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，使用区域内插。

2.4.2 实验指标

图像配准的评估指标多种多样，有的是阶段性的指标，描述具体配准步骤中的性能好坏，如最直接的每幅图像中检测得到的特征点数目，两幅图像之间达成匹配的匹配点对数。对配准效果的评价也可以从最终效果的方面考虑，比如使用峰值信噪比 *PSNR*（Peak Signal to Noise Ratio），基于像素点之间的误差，评价配准之后的图像与参考图像的相似程度。本次实验从多个方面对算法性能进行了比较。

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。另外一种衡量精度的指标是均方根误差 *RMSE*,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。

$$RMSE = \sqrt{\frac{1}{n} \sum \| (x'_i, y'_i) - f(x_i, y_i) \|^2} \quad (2-8)$$

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* 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。

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:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \| I(i, j) - K(i, j) \|^2 \quad (2-9)$$

峰值信噪比定义为:

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

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。特别的,使用在图像配准中,两幅图像之间的相同部分的多少直接影响峰值信噪比的值,所以使用范围有限,只有当配准后的图像与参考图像接近完全重合或者纵向比较不同算法的效果时适合使用峰值信噪比。

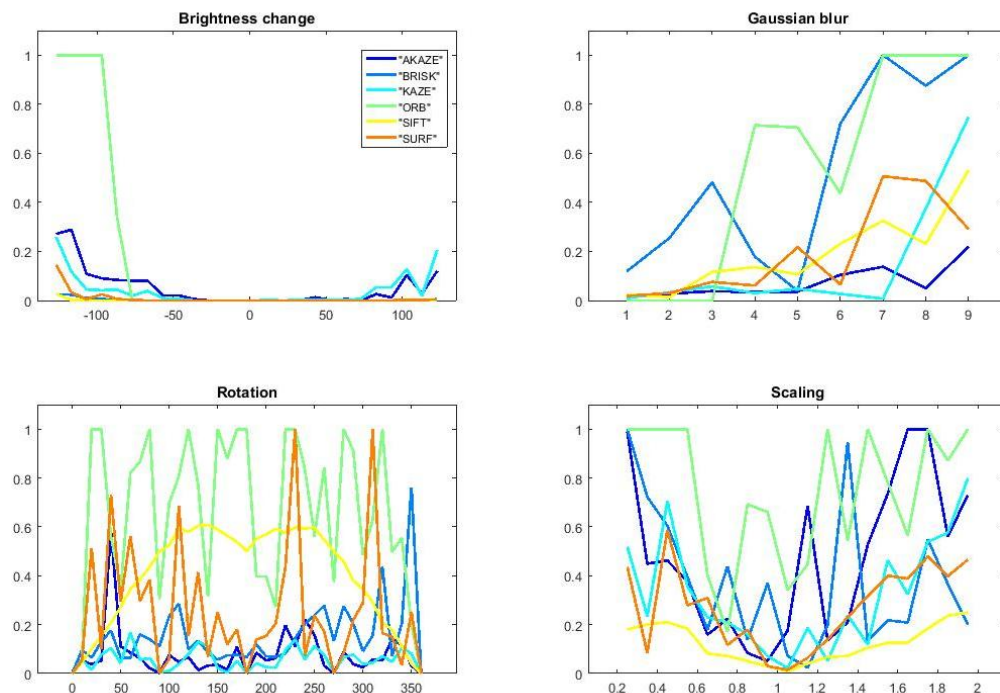
除了配准算法中特征点的准确度和匹配的成功率,以及最终的配准效果,即拟合

出的变换模型的准确度，我们关注的还有算法的实时性。移动端设备的普及使得对实时性要求高的场合越来越多，但同时又受限于计算资源的有限，所以这对算法提出了很高的要求。本次实验中统计了不同算法下每次配准的平均运行时间。配准的耗时同时与特征点数目有关，本文根据每种算法检测出特征点总数的均值，计算出了不同算法平均到每个特征点下的运行时间。

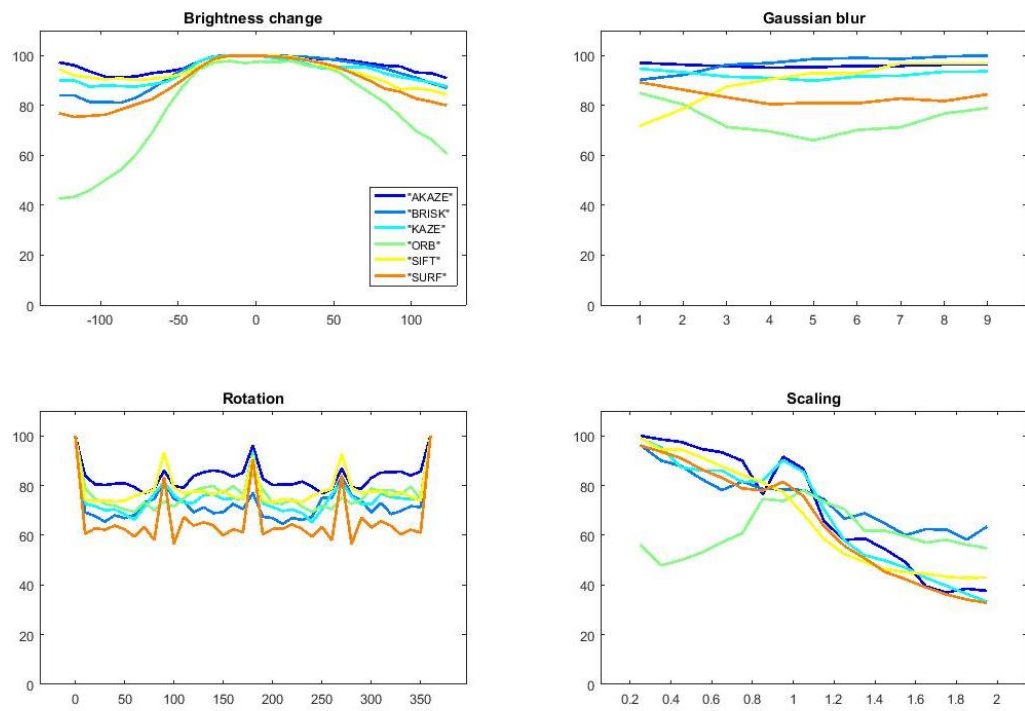
2.4.3 实验结果与分析

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总数的均值。

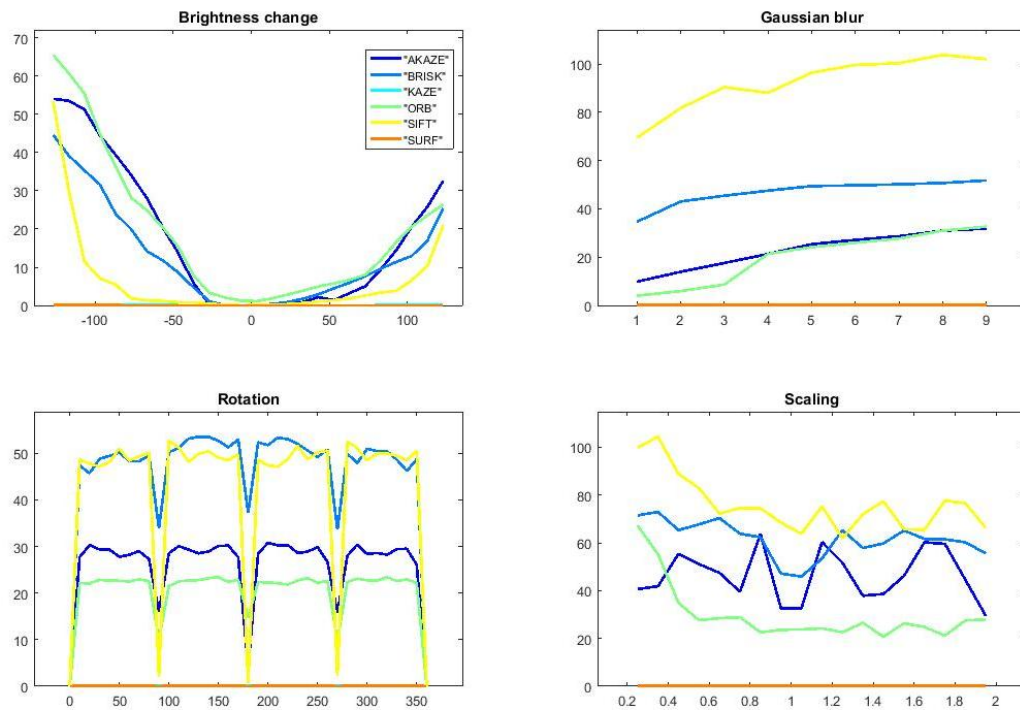
下面是根据统计数据生成的 txt 文件使用 matlab 软件画出的特性曲线。不同图像使用了不同的评价指标，每幅图像中又分成了四个子图，代表了实验模拟的四种变换。不同颜色的曲线代表不同的特征点检测和匹配算法。



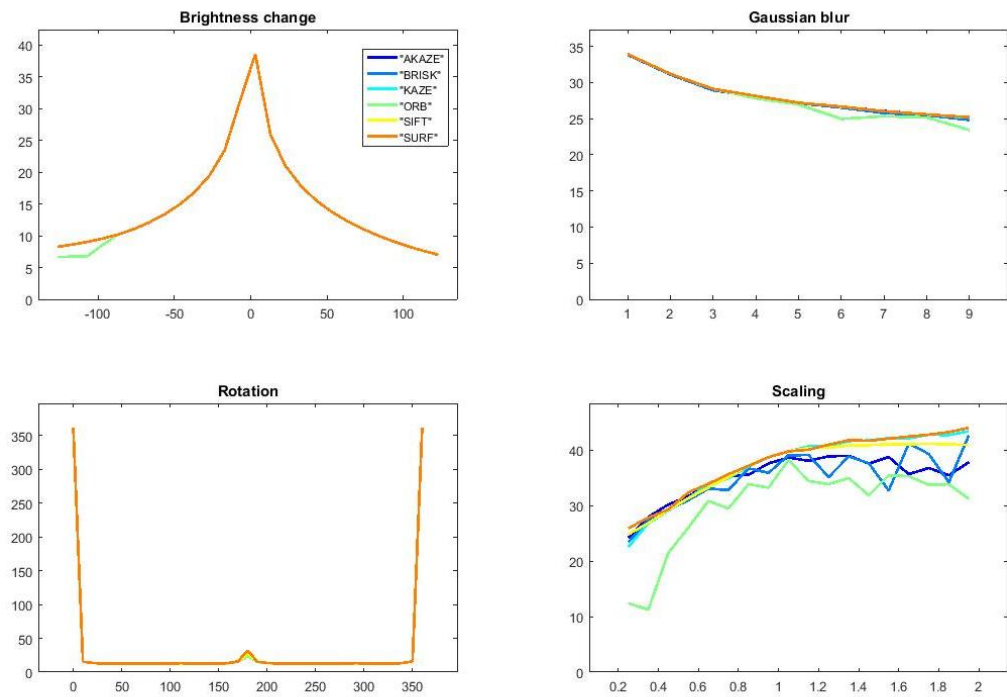
(a) HomographyError



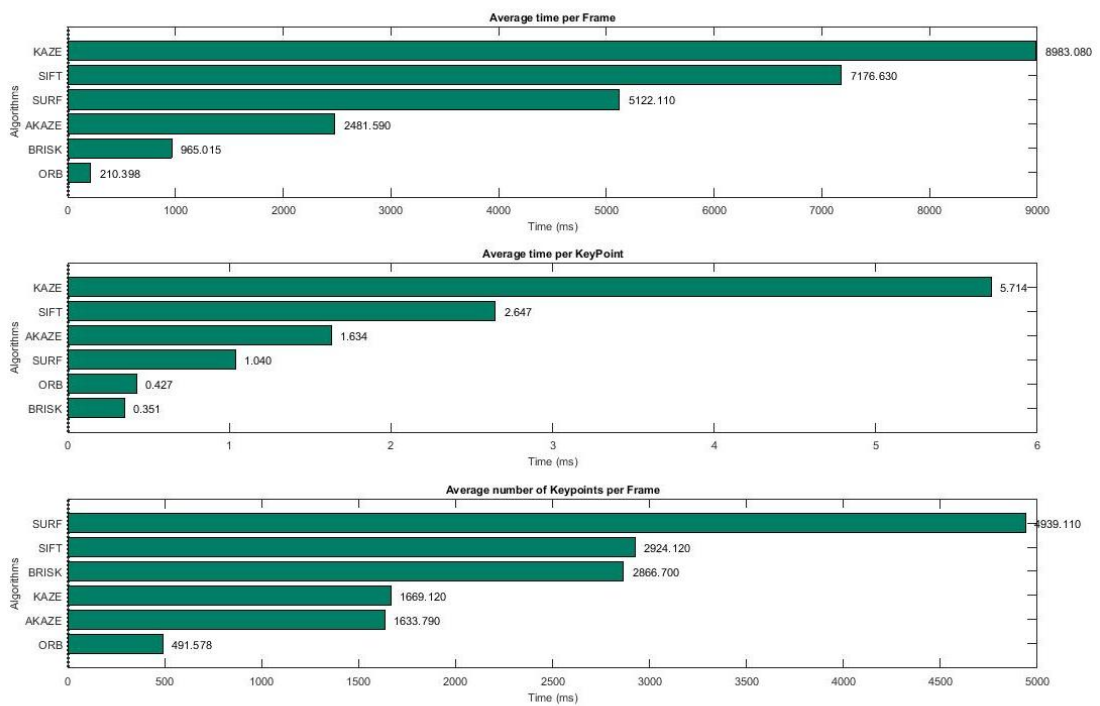
(b) correctMatchesPercent



(c) Meandistance



(d) PSNR



(e) Speed

图 2-6 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。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。本章设计了实验对其进行仿真和比较，画出了不同算法的特性曲线。实验从多角度对算法进行评价比较，如特征点之间的平均距离，峰值信噪比，平均运行时长等，有些指标在之后的研究中还会用到。

第3章 基于 SIFT 的低复杂度配准

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，可以根据自身的使用场景进行算法的选择。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做出了改进，提出了一种具有低复杂度的算法，不必改变 SIFT 本身的架构，可以加快配准的速度。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。

3.1. 算法原理

3.1.1 图像高斯金字塔

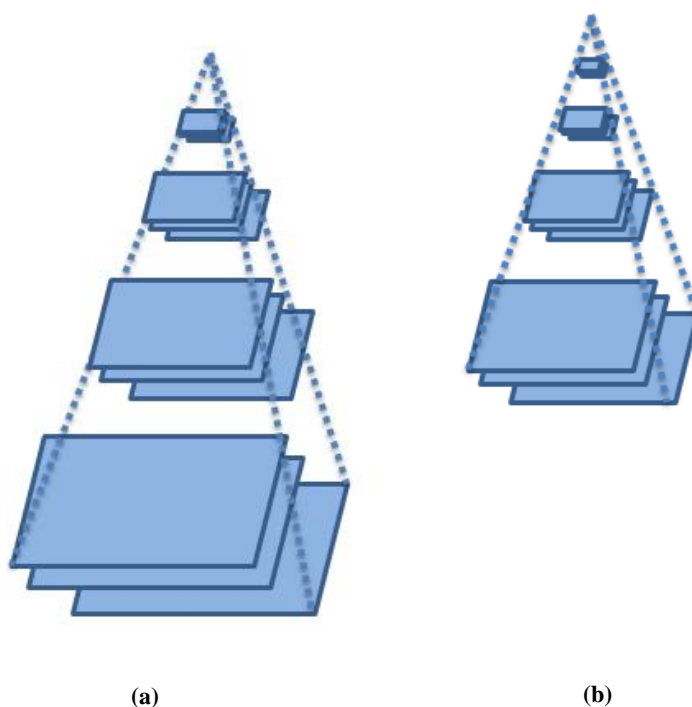


图 3-1 对实验图像采样前后的金字塔规模比较。不同的组之间
是下采样的关系，同一组的不同层之间是不同方差因子的高斯模糊

在前文中介绍透视投影时就提到近小大远小的特点，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，也是 SIFT 算法的第一步就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。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。对图像高斯滤波的原因在于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。

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。也正是由于金字塔的构建加大了整个算法中的计算量，包括高斯模板在图像上的平滑，检测特征点时在尺度空间上寻找极值。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对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，如图 3-1 所示。

这里使用 2 倍下采样对原实验图像进行预处理。因为配准中的实验图像有 2 幅，根据只对待配准图像下采样还是同时对两幅图像下采样可以分成下面两种情况。

3.1.2 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。采用最近邻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。香港中文大学的研究者在 2018 那年的 ECCV 上利用对抗生成网络 (Generative Adversarial Networks, GAN)，搭建了 ESGAN，实现了从低分辨率图像复原超分辨率图像^[77]。

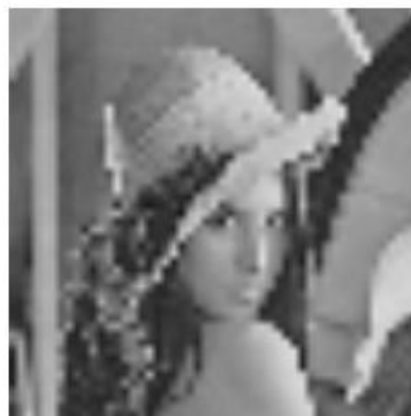
下图是将 Lena 原图进行十倍下采样之后的仿真，图像 Lena 分辨率从 512×512 大小缩小 51×51 大小 (图像长宽为离散的整数值，这里对按比例缩小之后的数值进行了向下取整)。下图 3-1 是 matlab 库中自带的实验图像，图 3-2 是使用不同插值方式得到的下采样图像。依次为区域插值 INTER_AREA、立方插值、线性插值和最近邻插值。为了便于对比，在显示时放大到了相同的大小，但保持其分辨率大小不变。



图 3-2 下采样插值实验中使用的原图 Lena.bmp



(a)



(b)



图 3-3 不同插值方法的实验图像对比

由图 3-1 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。

3.1.3 单下采样

这里的单下采样是单单对待配准图像下采样。所以可以认为是在下采样后的图像与参考图像之间的配准。使用 SIFT 算法估计出的变换模型可以直接作用在下采样图像上，得到配准后的图像。

2 倍下采样后的图像变为原图大小的 $1/4$ ，所以在构建图像高斯金字塔过程中高斯模板与像素点的卷积次数也缩小为 $1/4$ ，在同一尺度和上下相邻尺度对应的邻域中寻找极值点时的搜索空间也减少为 $1/4$ 。

3.1.4 双下采样

双下采样指的是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。

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

$$\mathbf{H}_p = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \quad (3-1)$$

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。

$$dst(x, y) = src\left(\frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}, \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}\right) \quad (3-2)$$

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:

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

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^[78]。

3.2. 实验结果与分析

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。

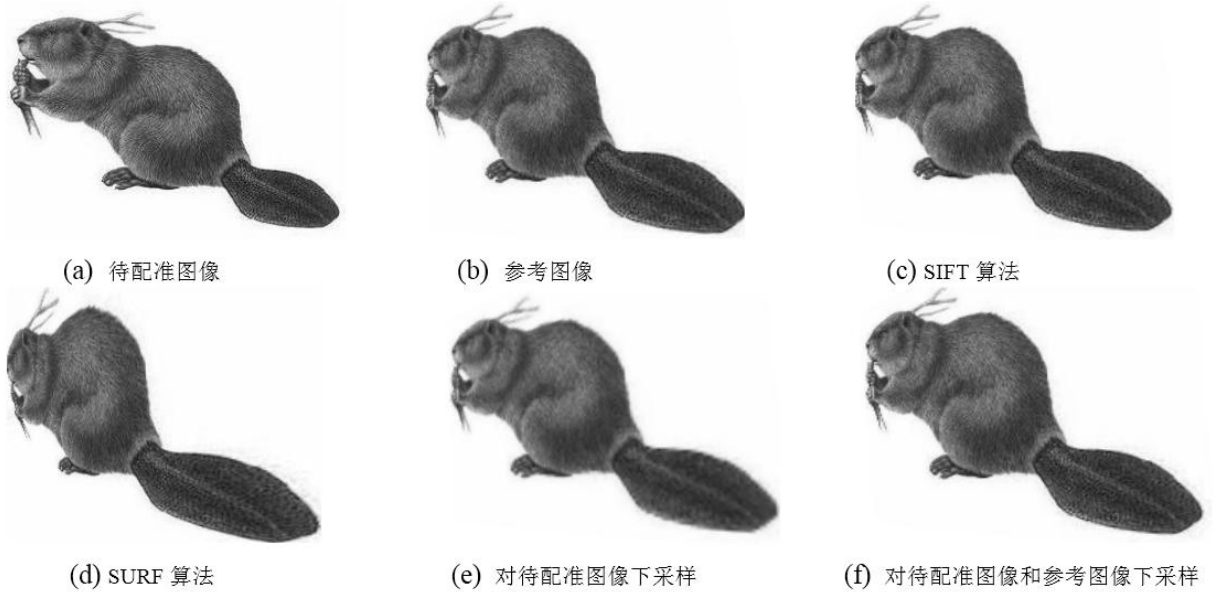


图 3-4 下采样缩小图像金字塔之后的图像配准

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 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			
			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。

3.3. 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，并在 SIFT 源码和其所提供的测试图像上进行了实验。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 SIFT 算法与显著性检测

SIFT 除了运算量较大,实时性较差的缺点,还有一个缺点是 SIFT 算法没有利用图像的颜色信息,算法全程都是基于灰度图像进行的,不论是图像金字塔的构建还是特征点的检测,不论是对彩色图像还是二值图像,都使用单通道的 8 比特量化表示。除了 SIFT 算法,其他很多算法具有同样的缺点,显然,损失了颜色信息的图像配准是不完整的。已经出现了一些增加考虑了颜色信息的配准算法,如出现在 2006 年的 CVPR 的 CSIFT^[57],在颜色不变空间中构建描述子,但是大大增加了计算量。SIFT 还存在的问题是关于特征点的数目和质量。增加特征点的数目有利于增强对真实变换矩阵的拟合能力,但同时也会加大算法的计算量。所以,对特征点数目和质量的调整是一个研究的重点。本章在 SIFT 算法基础上引入了显著性检测 FT 算法及其衍生出的对于主体部分分割的算法,并对其进行了改进,可以在一定程度上解决以上这些问题。

4.1. 显著性图上的特征点检测

4.1.1. 显著性检测 FT 算法

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,Litti^[79]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^[80],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 FT (Frequency-tuned),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^[81],有更好的精度和召回率。

CIRLab 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。FT 利用 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_{μ} 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。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 CIELab 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。

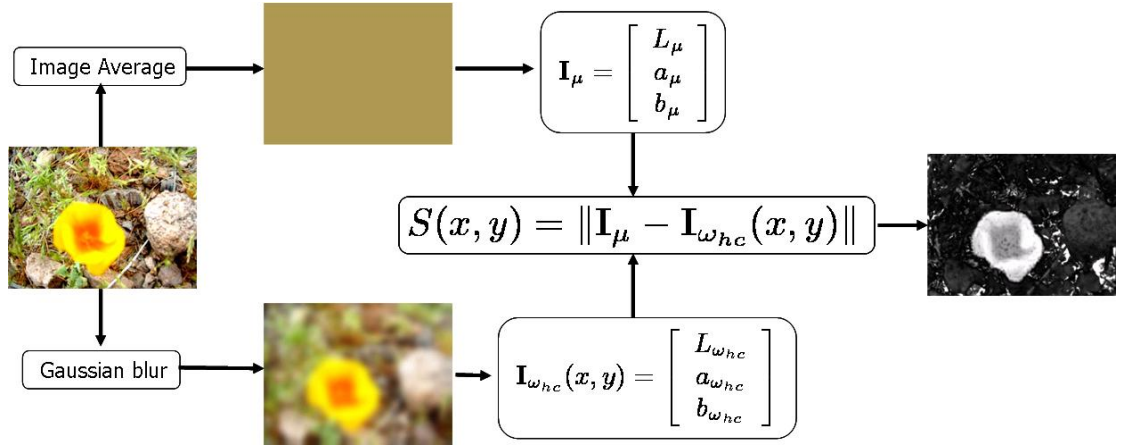


图 4-1 FT 显著性检测方法图示

4.1.2. 对 FT 的仿真

实验环境继续使用 DELL 计算机，在 Visual Studio 2013 下实现了 FT 算法对图像的显著性检测。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。下图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 4-2 FT 显著性检测结果

从图 4-2 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。

4.2. 显著性图上的配准实验

4.2.1. 实验理论

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 CIRLab 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拥有更大的显著性，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。

4.2.2. 实验结果与分析

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。

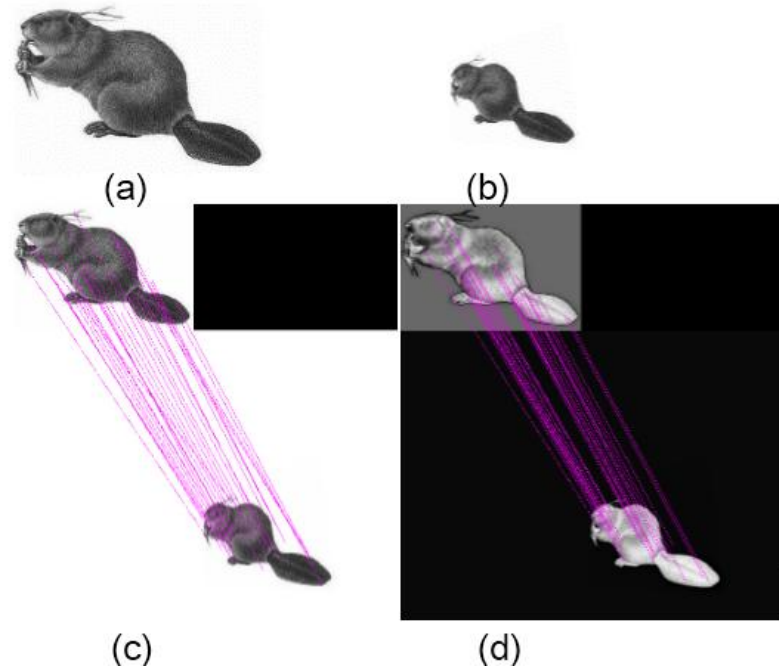


图 4-3 显著性图上的特征点检测与匹配

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，代价是减速。

4.3. 分割图上的特征点

4.3.1 实验理论

Achanta 还借助相对简单的分割技术证明了最终的显著性图像可以应用在图像分割中。整个物体中分割原理如下图所示：

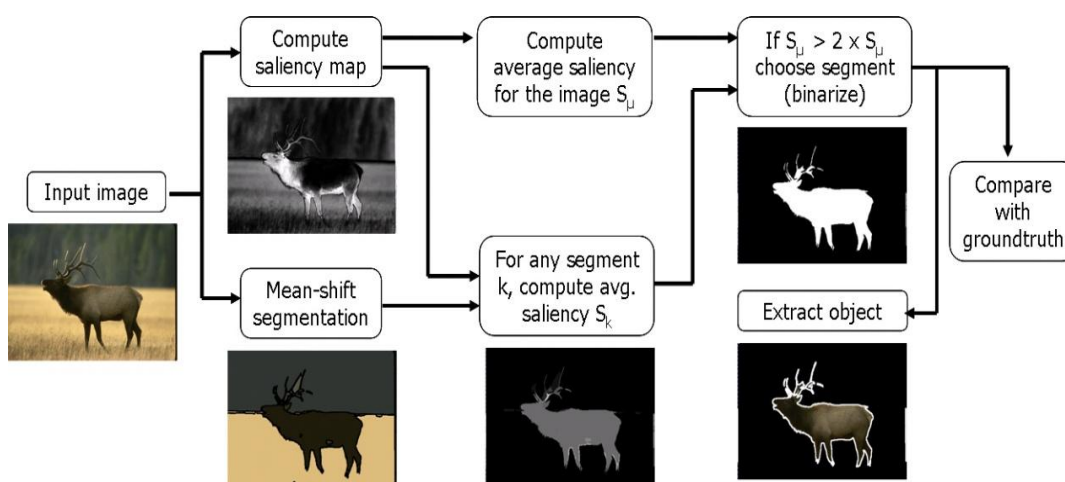


图 4-4 基于显著性检测的图像分割方法

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。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，如果我们在原始地图上施加掩码，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滤除掉背景的影响，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。

4.3.2 实验结果与分析

下面是实验结果，实验环境与实验图像与本章的上一小节相同。

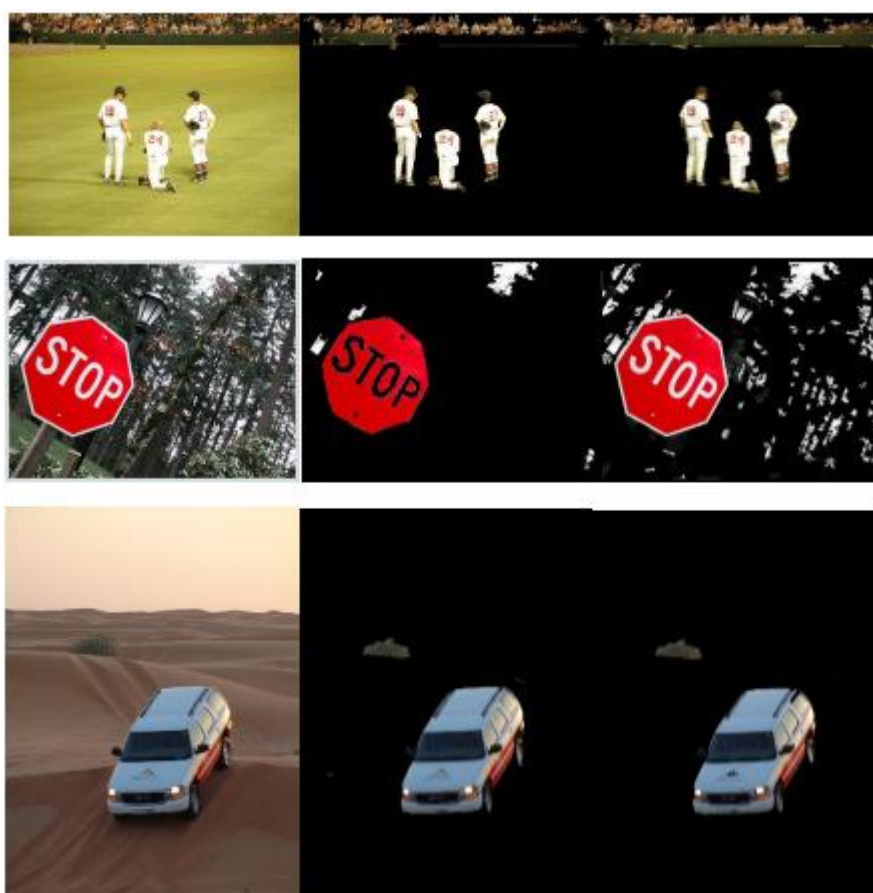
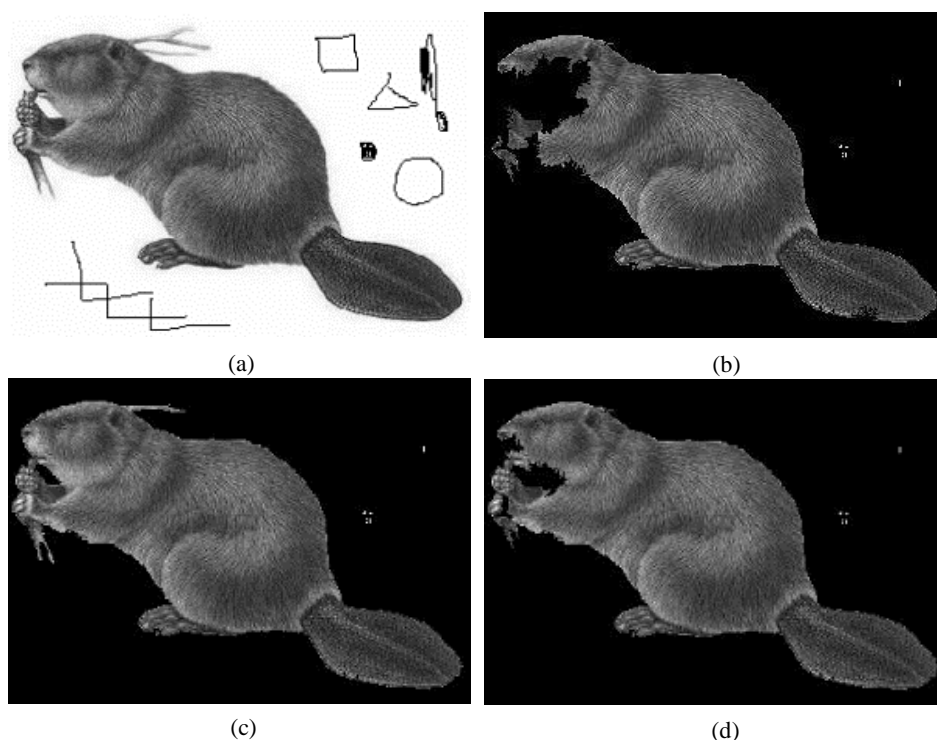


图 4-5 改进前后的图像分割方法

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 4-6 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)
N	144	116	124	117
RMSE	16.483	1.476	12.651	6.460
Time/ms	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。

4.4. 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 FT 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 配准算法的一种新应用

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。图像拼接应用在无人机航拍、手机摄影等领域，可以获得大尺寸的较大视野的图像。图像融合则是将不同类型图像叠加，将信息互补，得到可以提供更多信息的图像。图像拼接和图像融合（特别是像素级的图像融合）的主要步骤都是图像配准，建立两幅或者多幅图像之间在空间上的联系。此外，利用图像配准描述图像间关系的特性可以实现图像间冗余数据的去除。本章就提出了一种利用配准后的图像与参考图像重合度较高的特点，对二者进行像素级的运算，得到稀疏性的图像，从而可以实现压缩感知，较小图像集的规模。

5.1. 理论思想

5.1.1. 哈希算法

我们算法的整体思路是找到相似的，可以配准的图像组，对其进行配准，从而达到减少存储空间的目的。所以第一步也是很关键的一步就是如何在图像数据集中得到满足配准关系的图像。在各自搜索引擎中我们可以注意到以图搜图的功能的上线，这其实就是运用了搜索相似图像算法。和特征点匹配算法类似，以图搜图可以划分为相似度的测量和搜索策略的确定。这里首先介绍一些常用的相关算法，对图像进行编码，得到每个图像对应的指纹。

哈希算法就是一种对图像建立指纹的算法。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，相对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。加密哈希算法不可逆，不可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。所以被广泛应用在门户网站加密，文本加密，口令加密，数字签名等领域，但是这种算法却不适合应用在我们的算法中，这是因为加密哈希有一个特点就是文件一旦发生改变，即便变化是微小的变化，哈希值都会发生巨大的变化，即相似图像之间的哈希值会差异巨大。

我们希望相似图像或者相似文本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（average Hash, aHash）、差异值哈希（difference Hash, dHash）。

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，作为原图像的哈希值，因为图像经过了归一化，所以得到的哈希值的长度是一定的。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。

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。下面分别是对 512×512 大小的图像进行缩放得到缩略图和更小尺度的 256×256 大小的图像的缩略图，并通过计算得到了二者各自的哈希值。



图 5-1 差异值哈希算法示意图

差异值哈希算法可以分为下面几步：

(1) 图像归一化到 8×9 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) 比较相邻像素，左边的像素灰度值大于右边相邻的，则将对位置记为 1，否则为 0，实际上得到一幅二值图像

(4) 将二值图像中的 01 像素值按行拼接，得到二进制字符串

这里说明一下归一化的意义。归一化的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 8×8 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 8×9 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。

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，这里暂不展开。

5.1.2. 图像减法

图像在计算机中以矩阵或者说阵列的形式表示。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：

$$\begin{aligned} s(x, y) &= f(x, y) + g(x, y) \\ d(x, y) &= f(x, y) - g(x, y) \\ p(x, y) &= f(x, y) \times g(x, y) \\ v(x, y) &= f(x, y) \div g(x, y) \end{aligned} \tag{5-1}$$

其中，图像之间的加法可以用在带噪图像的降噪中。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）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。下面图像来自其中的 bikes 数据集，图像之间不仅有角度变化，还有模糊变化。(c)是(a)、(c)做图像减法的差值图像。



图 5-2 图像减法(a)-(b)=(c)

本章对实验图像使用差异值哈希算法计算了其图像指纹。(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。

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。下图中第一行是图 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。

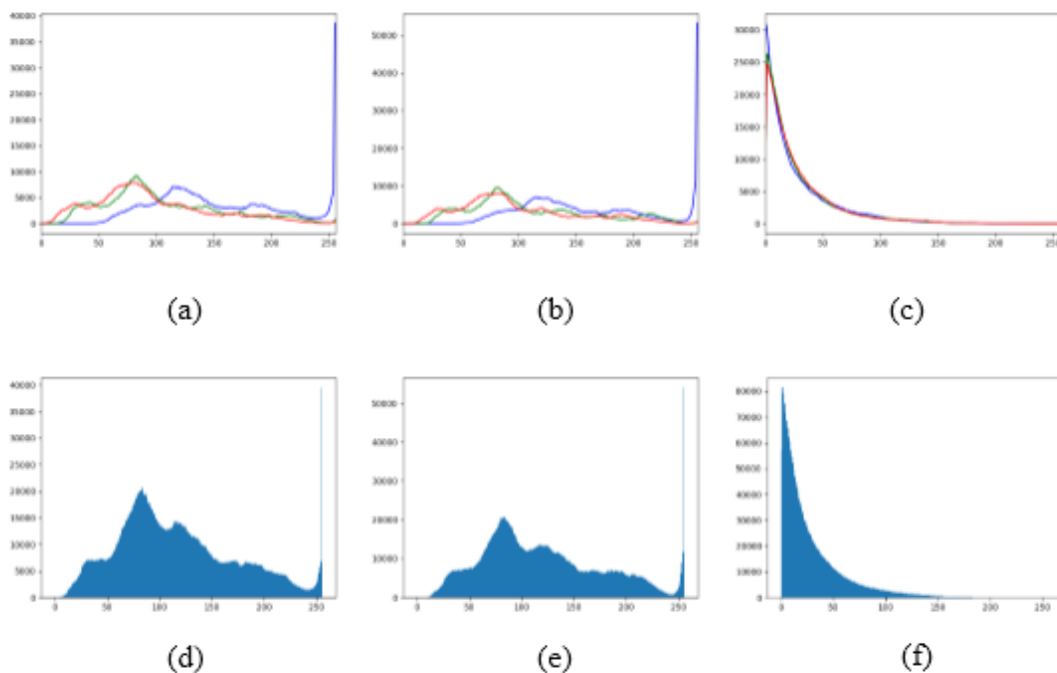


图 5-3 原图和差值图像的直方图

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。

5.1.3. 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。

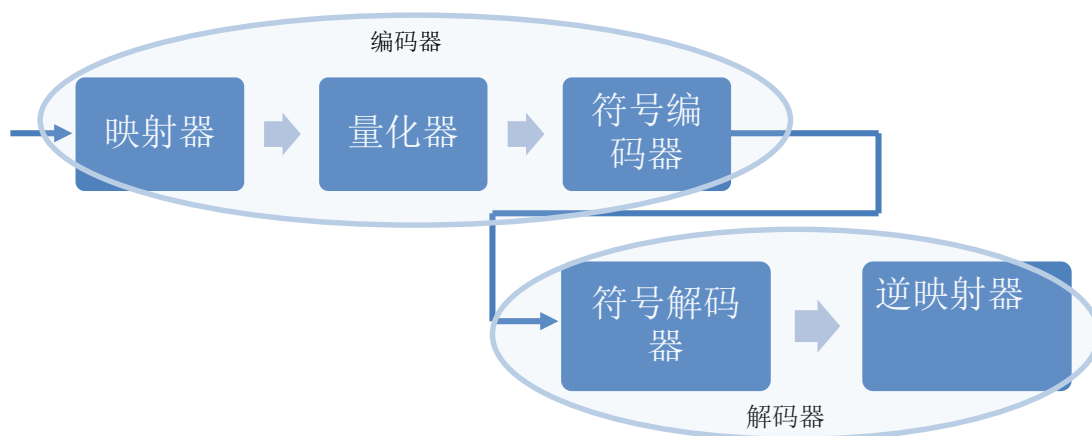


图 5-4 图像压缩流程

常用的图像压缩算法有霍夫曼编码，LZW 编码、行程编码、块变换编码、预测编码、小波编码等。霍夫曼编码已成为 JPEG 等压缩标准中常用的编码算法。霍夫曼编码是一种变成编码，针对信源符号出现的概率不同，即灰度直方图不均匀，根据信源出现的概率大小，使用不同长度的码字进行编码。对信源符号按照出现概率进行排序，对概率较小的信源使用更长的码字表示。同样属于变成编码的还有 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。

5.2. 实验设计

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。这里我们采用相关图像格式和编程函数中默认的压缩算法来对图像或者字符串进行压缩。

5.3. 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 AKAZE 算法,具体实现使用了 OpenCV 工具库。在实现 AKAZE 配准算法的特征点匹配过程中,使用了最近邻比率法,认为最近邻特征点之间的距离小于次近邻的 0.75 时认为是良好的匹配对,在估计最佳变换矩阵时使用了随机抽样一致性算法(Random Sample Consensus, RANSAC),通过迭代的方式剔除了一些不符合估计出的模型的外点。变换模型选取的是透视投影,包含了仿射变换。具体的矩阵形式是

单应矩阵。

先来看图像配准算法和图像减法在 **bike** 图像下的表现,两幅图像与前文中一样。



图 5-5 AKAZE 算法中特征点的匹配

如图所示,小圆圈表示检测得到的特征点,不同的匹配对随机使用不同颜色的线条表示。在 **bike1** 中找到了 1498 个 AKAZE 特征点,在 **bike3** 中找到了 1161 个特征点,可以看到,两幅图像的左上角之间还是存在着一些误匹配点对,这将影响之后的变换矩阵估计,这也是采用 RANSAC 算法的原因。下图是配准后的图像(c),及其与参考图像的差值图像(d)。

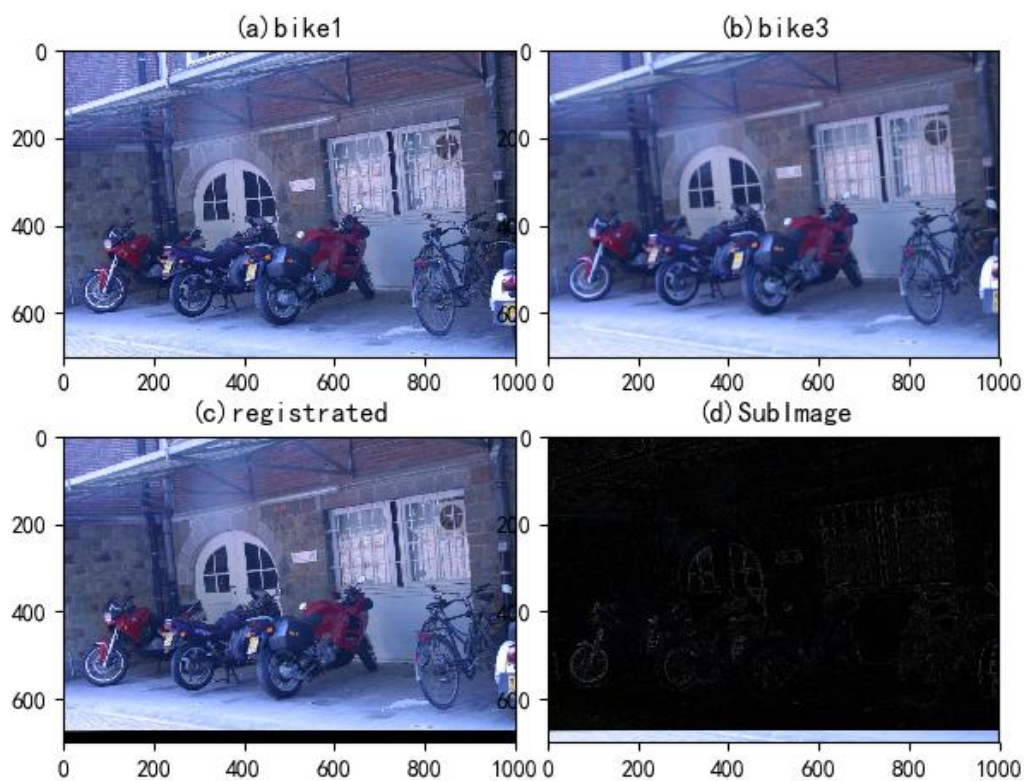


图 5-6 配准后图像与参考图像的差值和复原结果

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，第二组为鲁汶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。

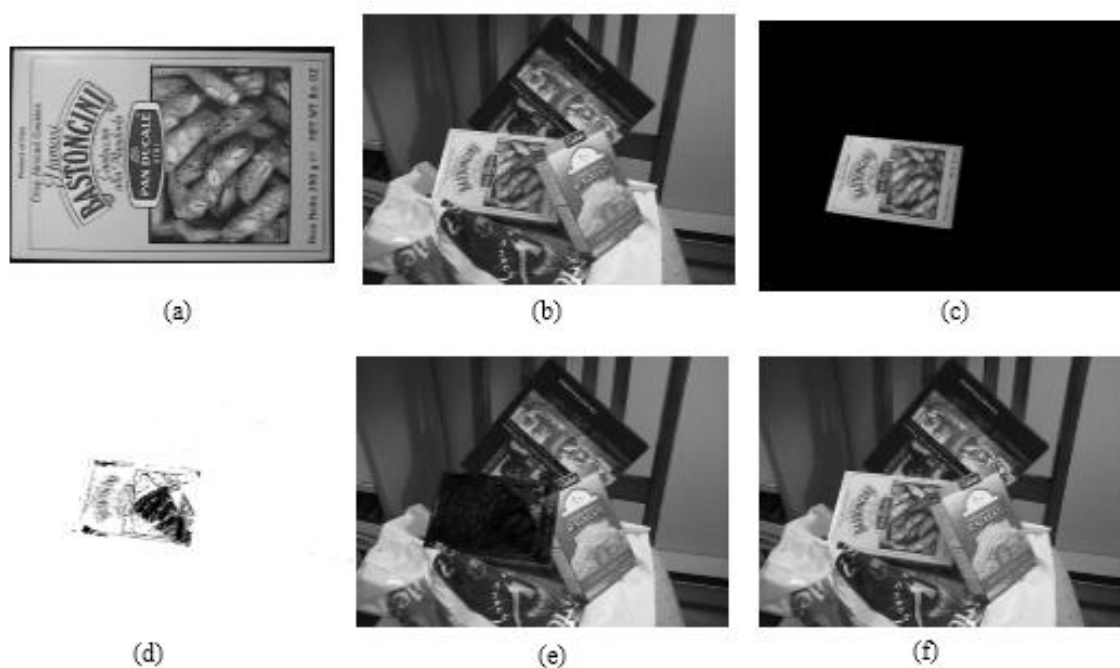


图 5-7 配准后图像与参考图像的差值和复原结果（box 图）

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 0 值和接近于 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。

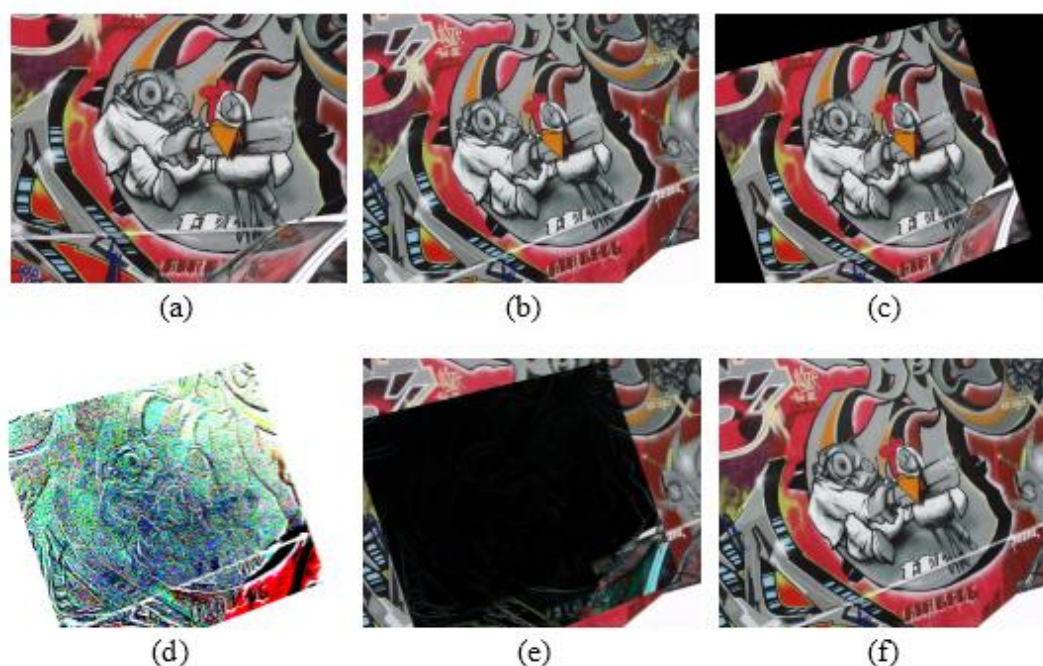


图 5-8 配准后图像与参考图像的差值和复原结果（graffiti 图）

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	汉明 距离	原始图像大 小/Byte	差值图像大 小/Byte	原图压缩 /Byte	差值压缩 /Byte
box.jpg	16	216756	589842	----	214548
box_in_scene.jpg		589842		215228	
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. 本章小结

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。事实上，我们不必要把差值图像保存下来，在本章的实验中只是为了形象化地展示矩阵的稀疏性，实际使用过程中只需保存减法之后得到的字符串。

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,可以检测处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,在进行实验时所选取的图像代表性不够强,算法没有经过海量的真实环境下所成图像的检验。各种优化工作做得不够

完善，距离真正的应用还有一段距离。

（2）对图像配准算法本身的改进有限，对优化算法的理解还不够深入。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] 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.

- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] Wang Y, Wu D. An improved adaptive SSDA based on wavelet-pyramid[C]// American Institute of Physics Conference Series. AIP Publishing LLC, 2017.
- [25] P.A. Viola, Alignment by maximization of mutual information, Ph.D. thesis, Massachusetts Institute of Technology, Boston, MA, USA, 1995
- [26] 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
- [27] 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
- [28] 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
- [29] 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
- [30] 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
- [31] 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
- [32] Hu M K. Visual pattern recognition by moment invariants[J]. Information Theory, IRE TRANSACTIONS on, 1962, 8(2): 179-187.
- [33] 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.
- [34] 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.
- [35] Flusser J. Moment invariants in image analysis[C]//proceedings of world academy of science, engineering and technology. 2006, 11(2): 196-201.
- [36] Flusser J, Suk T. Rotation moment invariants for recognition of symmetric objects[J]. Image Processing, IEEE TRANSACTIONS on, 2006, 15(12): 3784-3790.
- [37] Hough P V C. Method and means for recognizing complex patterns: U.S. Patent 3,069,654[P]. 1962-12-18.
- [38] 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.
- [39] 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 .
- [40] 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 .

- [41] Hans P M. Towards automatic visual obstacle avoidance[C]//Proceedings of the 5th international joint conference on Artificial intelligence. 1977: 584-584.
- [42] Harris C, Stephens M. A combined corner and edge detector[C]//Alvey vision conference. 1988, 15: 50.
- [43] 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.
- [44] 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.
- [45] 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.
- [46] Lindeberg T. Edge detection and ridge detection with automatic scale selection[J]. International Journal of Computer Vision, 1998, 30(2): 117-156.
- [47] Lindeberg T. Feature detection with automatic scale selection[J]. International journal of computer vision, 1998, 30(2): 79-116.
- [48] Mikolajczyk K, Schmid C. An affine invariant interest point detector[M]//Computer Vision—ECCV 2002. Springer Berlin Heidelberg, 2002: 128-142.
- [49] 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.
- [50] 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.
- [51] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International journal of computer vision, 2004, 60(2): 91-110.
- [52] Bay H, Tuytelaars T, Van Gool L. SURF: Speeded up robust features[M]//Computer Vision—ECCV 2006. Springer Berlin Heidelberg, 2006: 404-417.
- [53] 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.
- [54] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors[J]. Pattern Analysis and Machine Intelligence, IEEE TRANSACTIONS on, 2005, 27(10): 1615-1630.
- [55] 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.
- [56] 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
- [57] 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
- [58] 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

- [59] 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
- [60] Rosten E, Drummond T. Machine learning for high-speed corner detection [M] // Computer Vision—ECCV 2006. Springer Berlin Heidelberg, 2006: 430-443.
- [61] 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.
- [62] Taylor S, Rosten E, Drummond T. Robust feature matching in 2.3 μ s [C] // Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on. IEEE, 2009: 15-22.
- [63] Taylor S, Drummond T. Binary histogrammed intensity patches for efficient and robust matching [J]. International journal of computer vision, 2011, 94(2): 241-265.
- [64] Alcantarilla P F, Bartoli A, Davison A J. KAZE features [M] // Computer Vision—ECCV 2012. Springer Berlin Heidelberg, 2012: 214-227.
- [65] 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. 43(2): 281-285
- [66] 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.
- [67] 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
- [68] 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
- [69] 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
- [70] Detone D, Malisiewicz T, Rabinovich A. Deep Image Homography Estimation [J]. 2016.
- [71] 佚名. Three.js 开发指南 [M]. 2015.
- [72] Hill F, Kelley S. 计算机图形学: OpenGL 版 [M]. 清华大学出版社, 2009.
- [73] 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.
- [74] 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.
- [75] P Perona, J Malik. Scale-Space and Edge Detection Using Anisotropic Diffusion. IEEE Computer Society, 1990, 12(7): 629-639
- [76] J. Weickert; B.M. T. Hermen; M.A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. IEEE TRANSACTIONS on Image Processing. 1998
- [77] Wang X, Yu K, Wu S, et al. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks [J]. 2018.
- [78] 张晨光, 周诤, 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.

- [79] Itti L , Koch C , Niebur E . A Model of Saliency-Based Visual Attention for Rapid Scene Analysis[J]. 1998.
- [80] 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.
- [81] 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，即将启程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] 《A novel image registration algorithm based on salient region and SIFT》, The 3rd International Society for Artificial Intelligence and Robotics (ISAIR 2018)
- [3] 《基于 SIFT 特征点检测的一种改进的低复杂度图像配准算法》,第十九届全国图象图形学学术会议 (NCIG2018)