



中国航天

中国空间技术研究院

硕士学位论文文献综述

论文题目：基于特征点检测的图像配准研究与应用

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

导 师：周 詮

研 究 生：张晨光

完成时间：2017 年 10 月 25 日

目录

第一章 绪论	1
1.1 课题研究背景及意义	1
1.1.1 研究背景	1
1.1.2 研究意义	1
1.2 国内外研究现状	3
第二章 图像配准基础知识	12
2.1 图像配准的分类	12
2.1.1 根据待配准图像之间的关系	12
2.1.2 根据配准图像信息处理方法	13
2.1.3 根据选取的特征信息	13
2.2 图像配准的原理	15
2.2.1 图像配准问题表述	15
2.2.2 图像配准数学模型	16
2.3 图像变换	16
2.3.1 刚体变换 (Rigid Transform)	17
2.3.2 相似变换 (Similarity Transformation)	17
2.3.3 仿射变换 (Affine Transformation)	18
2.3.3 投影变换 (Projective Transformation)	19
2.3.4 非线性变换	20
2.4 图像配准流程	21
2.5 图像配准评价标准	22
2.5.1 准确率	22
2.5.2 配准精度	23
2.5.3 峰值信噪比	23

2.5.4 相似度检测 SSIM	24
2.5.3 配准速度	24
2.5.3 其他评价标准.....	25
第三章 尺度不变特征变换(SIFT)	26
3.1 SIFT 综述	26
3.2 SIFT 特点	26
3.3 SIFT 流程	27
3.3.1 尺度空间的生成.....	28
3.3.2 检测尺度空间极值点.....	30
3.3.3 精确定位极值点.....	31
3.3.4 求取特征点的主方向.....	33
3.3.5 关键点描述子的生成.....	34
3.3.6 根据 SIFT 进行 Match	36
3.4 本章小结.....	37
第四章 快速鲁棒性特征点检测 SURF.....	38
4.1 SURF 综述.....	38
4.2 SURF 特点	38
4.3 SURF 算法原理	38
4.3.1 构建 Hessian 矩阵, 计算特征值 α	38
4.3.2 构造高斯金字塔.....	41
4.3.3 定位特征点.....	错误!未定义书签。
4.3.4 确定特征点主方向.....	45
4.3.5 构造特征描述子.....	错误!未定义书签。
4.4 本章小结.....	47
第五章 非线性 KAZE 特征.....	48
5.1 KAZE 综述.....	48

5.2 KAZE 原理	48
5.2.1 非线性尺度空间构建	48
5.2.2 特征点检测	49
5.2.3 计算特征点主方向	50
5.2.4 建立描述子	50
5.3 本章总结	51
第六章 几种优化算法	52
6.1 KD 树+BBF 算法	52
6.1.1 K 近邻算法	52
6.1.2 K 近邻算法的实现: KD 树	54
6.1.3 BBF 算法	57
6.2 随机抽样一致性	58
6.2.1 RANSAC 步骤	59
6.2.2 RANSAC 优缺点	60
第七章 总结	61
参考文献	63

第一章 绪论

1.1 课题研究背景及意义

1.1.1 研究背景

眼睛是心灵的窗户。我们很早就认识到视觉的重要性，我们看书看报，看电视，看电影，都离不开视觉的贡献。事实上，通过视觉，人和动物感知外界物体的大小、明暗、颜色、动静，获得对机体生存具有重要意义的各种信息，至少有 80% 以上的外界信息经视觉获得，视觉是人和动物最重要的感觉。

关于人眼成像的原理，实际上就是自然界或者人为的图经过晶状体等在视网膜上面所成的像，这就是所谓的图像。图像由光信号变为电信号，刺激中枢神经，大脑做出相应的反应。

随着技术的发展，我们已经可以通过各种传感器成像，比如遥感多光谱成像，毫米波成像，微波成像，同位素成像等，不同的领域利用不同的传感器，运用不同的图像处理技术，获得、突出对应的图像特征，从而实现特定的目标。数字图像处理起源于 20 年代第一幅数字照片的传输[1]。然而，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[2]。

由于成像条件的不同，得到的图像也有所不同，所以我们需要图像配准技术来将图片进行合成。这项技术在遥感，医用，制图，计算机视觉领域被广泛地应用。

1.1.2 研究意义

图像配准是许多应用问题必须的预处理步骤,比如:时序图像的变化检测或多

模图像融合、图像拼接、图像镶嵌、图像检索、目标识别,这些问题遍及军事、遥感、医学、计算机视觉等多个领域。许多领域都需要图像配准,实际应用过程可能会有所不同,但其中关键的因素是类似的。概括地说,图像配准是对取自不同时间、不同传感器或者不同视角的同一场景的两幅或多幅图像进行最佳匹配的过程,包括像素灰度匹配和空间位置对齐。图像配准是图像处理研究领域中的一个典型问题和技术难点,其目的在于比较或融合针对同一对象在不同条件下获取的图像,例如图像会来自不同的采集设备,取自不同的时间,不同的拍摄视角等等,有时也需要用到针对不同对象的图像配准问题。具体地说,对于一组图像数据集中的两幅图像,通过寻找一种空间变换把一幅图像映射到另一幅图像,使得两图中对应于空间同一位置的点一一对应起来,从而达到信息融合的目的。

图像配准技术在计算机视觉、医学图像处理以及材料力学等领域都具有广泛的应用。根据具体应用的不同,有的侧重于通过变换结果融合两幅图像,有的侧重于研究变换本身以获得对象的一些力学属性。20 世纪以来医学成像技术经历了从静态到动态,从形态到功能,从平面到立体的飞速发展。

图像配准的应用领域概括起来主要有以下几个方面:

(1) 医学图像分析: 肿瘤检测、白内障检测、CT、MRI、PET 图像结构信息融合、数字剪影血管造影术(DSA)等。这是因为单一的图像不能较为全面地提供有病变组织或器官的大小、形状、空间关系等详细信息,单纯依靠单幅图像做出判断的风险较大。因此,需要将不同成像原理和成像条件的异源图像(如MRI,CT,PET,SPECT 等)融合,得到比单一图像信息更丰富的融合图像,利用更为全面的信息更好地指导诊断。

(2) 遥感数据分析: 分类、定位和识别多谱段的场景信息、自然资源监控、核生长监控、市区增长监控等;由于搭载成像传感器的卫星与地面目标的相对运动,以及不同时间或自然条件的影响,使得同一目标区域图像在不同遥感影像上呈现相互联系但又不尽相同的特点,亟需配准技术对其进行校正。通过配准,使得多幅图像可以进行拼接、对比等,为地理数据获取、应急资料采集、气象预报等提供相比于单一图像更丰富的信息。

(3) 模式识别: 目标物运动跟踪、序列图像分析、特征识别、签名检测等;信

息处理过程的一个重要形式是生命体对环境及客体的识别[3]，对人类来说，特别重要的是对光学信息（通过视觉器官来获得）和声学信息（通过听觉器官来获得）的识别。就图像来说，需要样本库中的图像与训练图像之间进行配准。

(4) 计算机视觉：目标定位、自动质量检测等。实现图像理解是计算机视觉的终极目标[4]。

每一种配准方法通常都针对某个具体问题而设计的，众多方法中，唯一的共性就是每个配准问题最终都要在变换空间中寻找一种最有的变换，这种变换能够使两幅图像之间在某种意义上达到匹配，但对于不同的应用领域，对图像类型的要求不同，就需要具体问题具体分析。

1.2 国内外研究现状

图像配准技术最早出现在上世纪七十年代，由美国提出，期初运用于飞行器导航、导弹制导等领域，并得到了美国军方的大力支持和资金上的辅助。历经二十余年的研究，最终成功地将该项技术应用于潘兴 II 式导弹和战斧式巡航导弹上，保证其弹着点平均圆误差半径小于十几米，使得导弹的命中率得到极大地提高。八十年代之后，这项技术已经成为国内外广泛关注的焦点，中外学者对图像配准技术展开了深入的研究，并在各个领域取得广泛的应用，如模式识别自动导航、医学诊断、计算机视觉等领域。图像配准技术在各个领域都呈现出各自的特点和适应性，都是根据各自具体的应用背景，并需要结合实际情况而特殊定制，使得图像配准技术同时具有相似性和差异性。

早在 70 年代，EE.Anuta[5]就提出了使用 FFT 进行互相关图像检测计算的图像配准技术，以提高配准的速度性能；DL.Barnea 等[6]提出了使用模板子图像差值相似性测度的图像配准技术，它比使用 FFT 计算互相关相似性测度进行图像检测的方法有更高的性能；WK.Pratt[7]对图像配准的互相关技术进行了全面的研究；M.Svedlow 等[8]对图像配准的相似性测度和预处理方法进行了比较分析；Eric.Rignot 等[9]在较高层次上对多图像的自动配准技术和要求进行比较和分析；Flussr[10]针对变形图像间的匹配又提出一个自适应映射方法，自动地对两幅遥感图像进行分割，使得分割后两幅图像上相应子块间的相似度较大，从这些子块之间

的空间位置关系来对原来的两幅图像进行匹配。

根据 Institute of Scientific Information(ISI)的调查数据表明,最近十年里至少有超过 1000 篇的学术论文在研究图像配准问题。2012 年网上公布的美国申请专利中,图像配准相关的部分就超过 50 项,IBM 和 GE 等大型跨国企业甚至有自己的工作组专门研究图像配准问题,而在最近的国际顶级学术会议上都有关于图像配准的专题讲座,这足以见得图像配准问题的关注程度。

配准算法大体分为三类,分别为基于灰度信息法、基于变换域法,和基于特征法。

(一)基于灰度信息法

基于灰度信息的配准方法是利用图像本身所具有的灰度统计信息作为相似性测度,并根据标准的灰度级确定图像配准的方位与重叠性区域。一般不需要对图像进行图像分割与特征提取等复杂的预处理,优点是可以充分利用图像自身的灰度信息,但是计算量大,对噪声较为敏感,同时难以旋转、图像的仿射变换也无法被处理。其方法大体可分为互相关法、序贯相似度检测匹配法和交互信息法。

Rosenfeld 提出被称为交叉相关的一种匹配度量,它是通过计算模版图像与搜索窗口之间的相关值来确定匹配程度。若二者到达恰好配准的位置,则交叉相关将出现峰值。Barnea 等人于 1972 年提出一种称为序贯相似性检测算法(Sequential Similarity Detection Algorithms, SSDA)。它的改进在于提出了运算更为简单的相似性准则,认为匹配点即为残差增长最快的点,只有在匹配点附近的点才需计算整个模板,从而减少了运算量,加快了算法运算速度。

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[11]和 Collignon[12] 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原理,判断配准位置。优点是对异源图像配准效果较好,无需预处理以及鲁棒性强,缺点是计算量大,要求图像间有较大的重叠区,容易出现局部极值。该测度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。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[13]。最大互信息算法是图像匹配的有力工具，但其仍然存在确定，如插值引起的局部极值，容易导致误配准，空间信息利用不足等缺点，一些学者也提出了相应的改进方法。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。曹溪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[14]。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。粒子群的惯性权重动态变化提高粒子的搜索范围，并利用烟花算法的高斯变异算子增加粒子的多样性，防止粒子群算法陷入局部极值，并提高了配准的精确度

（二）基于变换域法

该方法根据变换域内图像改变方式完成图像的登记。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。

利用傅里叶变换，通过对比平移、旋转、缩放在时域和频域的对对应关系，求得相应的变换参数。其好处在于其变换对于平移、旋转、缩放有天生的对应关系、利于 FFT 的快速实现、对与频域不相关的噪声有较强的鲁棒性。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，配准性能更好。傅立叶变换方法充分利用相位性质对图像进行配准，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在空域和时域是局部化的，在高频处窗口高而窄，可以精确地定位出突变信号的位置；在低频处窗口矮而宽，适应分析缓变信号的需要。这种特性被称为“变焦”，因而小波又被叫做数学显微镜[16]。这样有利于分析非平稳信号（如遥感图像）。在和图像分解和重建的塔形算法启发下，基于多分辨率分析框架，Mallat 建立了离散正交小波的一种快速算法——Mallat 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[17]。将小波变换技术运用到图像配准领域是目前科技工作者研究的热点。利用小波分解可以方便的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[18]。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[19]。

(三) 基于特征法

根据特征进行图像配准的方法是基于像素特点衍生出来的,在该方法中,涵盖图像信息的特征点都能够被加以使用,例如边缘化的特征点、角点。对于每一幅图像,都有其特有的特征。因此,通过提取图像中容易提取的特征,再通过相关矩阵等方法将两幅待配准图像中的特征进行匹配,去除未配对的特征。利用匹配特征坐标联立方程组,解出变换参数,因此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。

算法中常用的特征有直线、边缘、轮廓、重心、不变矩、特征点等。对于直线、边缘、轮廓等的提取,较为经典的算法包括 Hough 特征提取[23]、Canny 算子边缘检测等[24]。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[25],该方法基于互相关和 Hough 变换,实验表明达到亚像素级精度,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。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[27]、基于 Zernike 多项式的 Zernike 矩特征[28]等。相关研究如 Xiaolong Dai 和 Siamak Khorram[29]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、鲁棒配准; Jan Flusser[30]利用图像中的不变矩特征,实现了目标识别和分类; Jan Flusser[31]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。

针对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、在部分遮挡或相同区域较少时算法失效等缺陷,越来越多的学者将研究重点放到了对局部特征的研究上,一系列性能优异的算法相继被提出。局部特征包括局部区域特征和局部点特征。

基于区域的局部特征一般具有两个特点:第一,提取的特征区域要具有某种显著性;第二,提取的特征区域要具有仿射、尺度、旋转等不变性。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 Extrema-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[33] (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[34] (Union-Find) 算法维护连接区域的列表和面积, 因此经过加速可满足实时应用。Tuytelaars T, Van Gool L[35]等人提出了两种局部区域特征提取算法, 即基于边缘区域 (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断裂, 从而造成特征区域的不匹配。为了弥补这方面的不足, 他们又提出了 IBRs 算法。IBRs 算法利用图像中局部密度极值点和以极值点为中心向外辐射的“光线”上的特征点构造仿射不变区域。Kadir 和 Brady 在 2001 [36]年提出了显著性区域 (Salient Region) 检测算法, 后在 2004 年进行了进一步改进[37]。在某一特定尺度下, 不同的局部图像的信息熵有可能相等, 但在不同尺度下信息熵会表现出不同的变化规律。根据这点, 可以进行某些局部仿射不变特征的检测。

Mikolajczyk、Tuytelaars、Schmid 等人在文献[38]中对几种主流的仿射不变区域检测算子进行了性能测试, 发现 MSERs 算法综合性能最优, Hessian-Affine 算法[39][40]次之; Hessian-Affine 算法[39][40][41]和 Harris-Affine 算法可以提取较多的不变区域, 在复杂场景匹配方面具有优势; 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得分虽然较低, 但在物体分类识别方面表现较好。最早的对局部点特征的研究可追溯至 20 世纪 70 年代末。1977 年, Moravec[42]利用灰度自相关函数检测出“兴趣点”, 即角点。其原理十分简单: 平坦区域的像素灰度相差不多, 而在多个方向都有亮度变化的位置则可判定为角点。缺点是对噪声敏感、不具有尺度和旋转不变性等。

1988 年, Harris 提出了 Harris 角点提取算法[43]。其基本思想是观察一个假定的特征点周围小窗口内的方向性强度平均变化。由于灰度变化会出现在多个方向, 因此要在所有可能的方向上进行计算。判定方法为首先获取平均强度变化最大值对应方向, 接着检查位于它垂直方向的变化是否也很强烈, 同时满足条件的便是一个角点。具体在计算中, 用一个协方差矩阵表现所有方向上的强度变化率。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：若这两个值比较小，那么位于平坦区域；若一个较大，另一个较小，则位于边界上；若两个值都较大，则位于角点处。因此，若两个特征值都大于某一阈值，则此点被判定为角点。Harris 角点对旋转和灰度变换具有不变性。

20 世纪 90 年代，Lindeberg[44][45][46][47][48]系统提出了图像的尺度空间理论，从理论上证明了构造尺度空间唯一的线性核是高斯核。通过分析图像中各个局部的尺度，极大地推动了尺度不变特征提取相关算法的发展，如 Mikolajczyk 和 Schmid[46]将 Harris 角点检测与高斯尺度空间相结合，提出了 Harris-Laplacian 算子，增加了 Harris 角点的尺度不变性。

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[50][51][52]，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对特征点的方向进行归一化处理，并形成 128 维特征向量对特征点进行描述。这样的一系列算法，使得 SIFT 特征在图像旋转、尺度变换、一定的仿射变换和视角变换等条件下都具有良好的不变性。由于其优良的性能，直到现在依然在目标识别、图像拼接、目标跟踪等领域扮演着重要角色，甚至已经成为评判一种算法性能的标准。然而，算法仍存在一些问题，如阈值过多且难以确定和存在一定的错误匹配等。

2006 年，Bay[53]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，在特征描述阶段采用一阶 Haar 小波响应等改进，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 SURF 算法外，比较重要的类 SIFT 算法包括 PCA-SIFT[54]、GLOH[55]、ASIFT[56]、ORB(oriented brief)[57]、CSIFT(colored SIFT)[58]等其他改进的 SIFT 算法[59][60]。

PCA-SIFT 利用了主分量分析技术 (Principal Component Analysis, PCA)。在生成描述符时共分三步: 第一步首先生成一个投影矩阵; 第二步检测特征点; 第三步利用投影矩阵对特征描述符降维, 得到较低维度的描述符。因此, 其核心思想为采用 PCA 技术, 通过寻找高维空间中起主要作用的分量, 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符的维度, 提高了特征的匹配速度。但它的缺点是事先需要对大量典型图像进行学习, 对于不同类型的图像, 此算法不适用。GLOH (Gradient Location-Orientation Histogram) 描述子通过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。在生成梯度直方图时, 由于维数较低, 因此不用进行降维处理。其不足与 PCA-SIFT 一样, 也需要事先对大量样本进行学习。

ASIFT (Affine-SIFT), 是 2009 年提出的一种基于 SIFT 的算法, 在较大的仿射变换下依然可以实现同名点的匹配。近来, 在手机等移动平台上已经有人成功实现了该算法, 体现了算法的较好的应用前景。

Rosten 和 Drummond[61]于 2006 年提出了 FAST (Features from Accelerated Segment Test) 角点检测算法, 并在 2008 年进行了改进[62]。这种算法认为如果一个点周围的一定范围的圆形区域内有足够多的点的灰度比中心点大或小, 则此中心点被判定为角点。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。

Taylor、Rosten、Drummond[63][64]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[66]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显著特征点。但高斯分解是牺牲了局部精度为代价的, 容易造成边界模糊和细节丢失。非线性的尺度分解有望解决这种问题, 但传统方法基于正向欧拉法 (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。如果可以在非线性尺度空间构造和特征提取、描述等方面进行优化，从而达到可以实时运算的要求，则 KAZE 算法一定会有更广阔的应用前景。

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[67]。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[68]。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[69]。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[70]。根据配对点周围的点的相似度配对匹配的正确与否。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[71]。

近几年国内学者针对图像配准的研究重心逐步转移到多模图像配准，提出了很多理论与方案。但是这些方法多基于像素级别的，而随着各个领域对图像质量要求的提高，如卫星遥感、医学领域需要更精确的配准，需要将配准精度控制在一个像素以内，即达到亚像素级。众多的学者对亚像素级配准的方法进行了研究与实验，提出了相关插值法、梯度法、小波变换法。相关插值法，这类方法的精度取决于插值算法的质量；梯度方法，其前提条件是图像灰度保持不变，微分算子对光照较为敏感；相位相关法利用 FFT 的方法提高配准的速度；小波变换法，使得求解配准参数的运算量较小，有效地提高效率。

基于特征点的配准技术研究已成为基于特征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在数量方面与图像的像素点相比有很大程度下降，运算量得到了大量消减。

(2) 特征点在匹配过程中相似性度量函数能够很好地反馈坐标变化，因此能够使精度和准确率在很大程度上有所提升。

(3) 特征点可以适应图像的各种变化，即使图像遭受光照、旋转、尺度和噪声等影响，依然可以提取出质量较好的特征点适应性较强。

可以看出，针对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)对于不同传感器图像与数据的配准技术。不同传感器获得的图像、图像灰度和特征信息往往有很大的不同，如 SAR 图像与可见光图像成像机理不同，同一地物或目标在灰度、统计特性等方面都有很大差异，很多研究者对于 SAR 与可见光图像配准也提出了一些方法。但大多利用河流、湖泊、道路等特征显著的图像进行实验，或在对于图像目标有先验知识情况下进行配准，但算法缺乏广泛的适应性。

(2)自动配准实现。自动配准是不需要人工干预、计算机可以根据既定的程序自动完成图像配准，但目前的研究中很多算法还需要一些人工的干预，无法实现全自动配准。

(3)亚像素级的图像配准技术实现。

(4)快速配准算法的实现。

(5)较大几何位置差别条件下的配准技术的实现。在待配准图像和参考图像之间有较大的尺度比例差别、较大的旋转角及较大的平移时，甚至还伴有各种图像的畸变等情况下的配准也存在很大的困难。

第二章 图像配准基础知识

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设计的，众多方法中，唯一的共性就是每个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，但对于不同的应用领域，对图像类型的要求不同，就需要具体问题具体分析。

1992 年，L. G.Brown 在其论文[65]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。

2.1 图像配准的分类

2.1.1 根据待配准图像之间的关系

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[67]。

多源图像配准，是由异源成像传感器获得的同一场景图像之间的配准。典型应用是多传感器图像的融合，比如医学图像领域的 CT、MRI、PET、SPECT 图像融合；遥感图像领域中多波段的图像融合，比如雷达、微波或者多谱段等；场景分类如建筑物、道路车辆，植被等。方法特点是建立传感器之间的变换模型，利用灰度信息的预配准、物体形状和一些基本标志可使问题简化。

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。典型应用是识别和定位模板，如地图、物体和目标等。遥感数据处理方面，定位和识别定义好的或者已知特征的场景，如飞机高速路、车站，停车场等。模式识别、字符识别、标识确认、波形分析等。方法特点是根据模板预先选定特征信息。

多角度图像配准，对从不同观察点获得的具有不同视场角度的图像间进行配

准。典型应用是对图像的深度或形状进行重建。接收机视觉：从视角差异中构建三维深度和形状信息；目标物体的运动跟踪、序列图像分析。方法特点是变换多为透视变换，多用于视觉几何和表面属性等的假设条件，须考虑阻挡问题。

时间序列图像配准，是在不同时间或者不同条件下获得的同一场景图像间的配准。典型应用是监视和检测变化或增长。医学图像处理中数字剪影、血管造影、注射造影剂前后图像的配准、肿瘤检测、白内障检测。遥感数据处理中自然资源监控、核增长监控、市区增长监控。方法特点是需要容忍图像中部分内容的差异和形变对配准的影响，有时需要建立传感噪声和视点变换的模型。

2.1.2 根据配准图像信息处理方法

(1) 基于待配准图像灰度信息的配准方法

以整幅图像的灰度信息为依据，建立待配准图像和基准参考图像之间的相似性度量，利用某种搜索算法，寻找出使得相似度量达到最优值的变换模型参数。也称之为直接配准法。当光照强度或灰度强弱作为显著信息而不是局部的形状或者结构作为特征信息时效果更好，更受研究者的欢迎。

(2) 基于待配准图像变换域内信息的配准方法

通常是利用傅里叶变换为基础，进行频域内的配准。傅里叶变换能够用于具有平移、旋转、缩放的图像配准中是因为傅里叶变换具有平移、旋转、缩放的特点。

(a) 对于两幅具有平移量的图像经过傅里叶变换后，他们的相位关系是不同的，因为在时域内平移量直接决定在傅里叶变换域内的相位关系；

(b) 对于两幅具有旋转量的图像，在傅里叶变换域内的旋转量保持不变；

(c) 对于两幅具有尺度缩放的图像，先进行坐标系转换，变换到对数坐标系当中，图像的缩放可以转换成图像平移进行处理。

(3) 基于待配准图像特征信息的配准方法

2.1.3 根据选取的特征信息

基于特征的图像配准方法能够将对整个图像的各种分析转化为对图像特征的分析，由于特征的数量相对于整幅图像来说只是一小部分，因此能够大大减小图像

处理过程的运算量，而且特征对灰度变化、图像变形以及遮挡等有很好的适应能力，因此应用也更广泛。根据特征点进行图像配准涉及到特征提取、特征匹配以及对运动关系的求解。

(a) 基于特征点的匹配

如果图像中的一个点具有独特性和不变性，则被认为是一个特征点，也可以被称为关键点或者兴趣点。一般所选取的特征点是相对于其领域表现出来具有某种奇异性的像素点。特征点往往容易被提取到，但是特征点所包含的信息相对较少，只能反映出其在图像中的位置坐标信息，所以在两幅图像中寻找匹配的特征点是关键所在。

Deriche [84]等人对之前的特征点检测算法做了较为全面的总结，发现许多算法存在一个共同的问题：针对不同的图像，算法检测出的特征点位置会有一些偏差，这就会给后续的图像处理过程带来很大的影响。因此，可以将不同图像中相对位置一致的特征点的定位偏差大小作为一种评价特征点检测算法优劣的标准。除此之外，Schmid[85] 等人指出以特征点的重复度（Repeatability）作为评价特征点提取算法性能优劣的一种通用性指标。

$$R = \frac{N_r}{\min(N_1, N_2)} \times 100\% \quad (2-1)$$

其中 N_1, N_2 分别是两幅图像中的特征点数量， N_r 是在两幅图像中重复出现的特征点数量，而 R 就是在两幅图像中提取的特征点的重复度。在相同条件下，重复出现的特征点所占的比重越大则该提取算法的性能就越好。

(b) 基于特征区域的匹配

在图像中寻找某些明显的区域信息作为特征区域，然而在实际应用中寻找到特征区域后，采用最多的还是区域的形心点，所以这类算法要求特征区域提取的精度非常高。

(c) 基于特征边缘的匹配

图像中最明显的特征就是边缘，而边缘特征也是最好提取的特征之一。因此给予边缘的匹配方法鲁棒性较强，适用范围很广，但这类方法对特征边缘的提取要求很高，并且要求边缘信息全部用数学语言予以表述也较为困难。

基于特征的配准方法是目前最常用的配准方法之一，该算法只需要提取待配准图像中的点、线、边缘等特征信息，不需要其它辅助信息，在减少计算量、提高效率的同时，能够对图像灰度的变化有一定的鲁棒性。但是，由于该算法只采用了图像一小部分的特征信息，所以这种算法对特征提取和特征匹配的精度及准确性要求非常高，对错误非常敏感。

2.2 图像配准的原理

2.2.1 图像配准问题表述

多幅图像是由于采用了不同的传感器在同一时间，或者是由一个或数个传感器在不同时间对地面同一目标成像获得的。在这种多幅图像的处理应用中，完成多幅图像之间的配准是其中一个重要的环节。在许多图像配准的文献中，图像配准的数学上的定义如下：给定两幅待配准的图像，称其中之一比如为基准图像，另一个为待配准图像，则称图像配准为两幅图像在坐标位置和灰度级上的双重映射变换。

从推广的范畴考虑，图像配准是产生三维空间的图像集合或者匹配某一目标的过程。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对齐。这是图像的一种广义匹配概念。比如说，对于同一公路场景，它的高分辨率图像可能是一条宽带，带有明晰的道路边界，而其低分辨率图像则可能是一根模糊线条，利用公路和它环境的几何特征可以判断两幅图像是否在几何上对齐，即配准或广义匹配。而不是将一幅图像变换得与另一幅相等。由此可以将图像配准和图像匹配的相关概念定义如下。

定义 1：图像配准是给定两幅待配准图像，对其中一幅图像作变换使得变换后的图像与另一个图像的内容在拓扑上相对应并几何上相对齐。图像配准可以称为图像的广义匹配。

定义 2：狭义的图像匹配是指一幅图像在灰度属性上按像元逼近另一幅图像。

因此，能够进行配准的多幅图像具有一个共同的特点，就是相邻的多幅图像分量之间至少有一部分在逻辑上是相同的，即相邻的图像之间有一部分是重叠的，应反映地面上的同一景物的特征，这是实现图像配准的基本条件。如果确定了相邻的

图像之间的空间几何函数关系即坐标关系，就可以实现图像在拓扑关系上的配准。但是确定所有的像素之间的坐标关系是很困难的，而且是不切实际的。所以人们通常的方法是确定足够数量的同名像素（即两幅或多幅图像中代表相同特征的像素）点，根据所采用的图像之间的数学模型，可以表示图像中其它所有像素之间的坐标关系，从而完成图像的配准。由此可见，图像之间的配准主要包括两方面的内容：其一，是确定足够数量的配准控制点（RCP, Registration Control Point）；其二，是根据这些配准控制点确定两幅或多幅图像的像素之间的坐标对应关系。

图像配准的基本问题就是找出一种图像转换方法，用以纠正图像的形变。造成图像形变的原因多种多样，例如对于遥感图像而言，传感器噪声、由传感器视点变化或平台不稳定造成的透视变化、被拍摄物体的移动、变形或生长等变化、闪电和大气变化，以及阴影和云层遮盖都使图像产生不同形式的形变。

2.2.2 图像配准数学模型

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：

$$f_2(x', y') = g[f_1(h(x, y))] \quad (2-2)$$

其中 h 表示二维空间坐标变换， g 表示灰度或辐射变换，描述因传感器类型的不同或辐射变形所引起的变换。配准的目标就是要找出最佳坐标，灰度变换参数。通常意义的配准只关心位置坐标的变换。灰度或辐射变换可以归为图像预处理部分。

2.3 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叠加，实际上其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。

各种配准技术都要建立自己的变换模型，变换空间的选取与图像的变形特性有关，图像的几何变换可分成全局、局部两类。全局变换对整幅图像都有效，通常

涉及矩阵代数，典型的变换运算有平移，旋转，缩放；局部变换有时又称为弹性映射，它允许变换参数存在对空间的依赖性。对于局部变换，由于局部变换随图像像素位置变化而变化，变化规则不完全一致，需要进行分段小区域处理。

变换模型[73][74]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、透视变换和非线性变换等。

经常用到的主要变换有刚体变换、仿射变换、投影变换、非线性变换。

2.3.1 刚体变换（Rigid Transform）

变换前后两点间的距离依旧保持不变，即变换前后图像保持不变，则被称为刚体变换。刚体变换可分解为平移变换、旋转变换和反转（镜像）变换。

刚体变换模型包括三个参数： t_x, t_y, θ 分别对应了平移和旋转变换。二维空间中的点 (x, y) 到 (x', y') 的刚体变换可以表示为：

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (2-3)$$

其中 t_x, t_y 分别为图像在 x 方向和 y 方向的平移参数， θ 是图像的选择角度。

2.3.2 相似变换（Similarity Transformation）

在刚性变换模型中加入一个尺度变换参数就形成了相似变换模型，相似变换也是一种典型的空空间变换类型。经过相似变换后，图像中的直线仍然为直线，物体的形状保持不变；但是由于尺度变换参数的加入，图像中任意两点间的距离会变化，物体大小也会随之发生改变。

二维空间中的点 (x, y) 到 (x', y') 的相似变换可以表示为：

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = S \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (2-4)$$

可见，相似变换有四个参数： S, t_x, t_y, θ 。容易看出，刚性变换是相似变换

的一种特殊形式，刚性变换中 $S=1$ 。

2.3.3 仿射变换 (Affine Transformation)

仿射变换是一种二维坐标到二维坐标之间的线性变换，保持二维图形的“平直性”（即变换后直线还是直线不会打弯，圆弧还是圆弧）和“平行性”（其实是指二维图形间的相对位置关系不变，平行线还是平行线，相交直线的交角不变）。

相比于相似变换，仿射变换中的尺度变换参数在不同方向上不一定是均匀的，

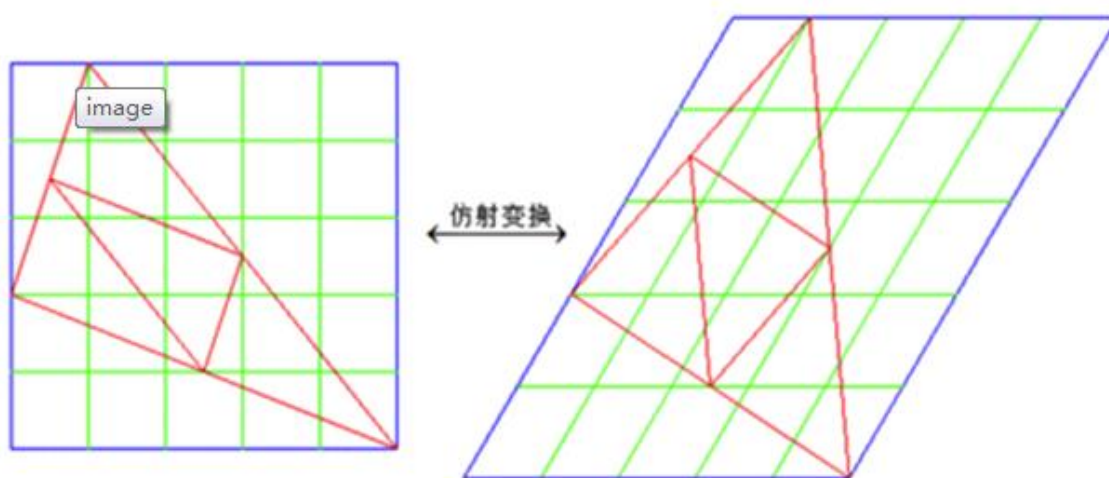


图 2-1：仿射变换，保持了二维图形的“平直性”和“平行性”

这就会形成图像的错切，就是允许图形任意倾斜，而且允许图形在两个方向任意伸缩的变换。仿射变换，可以保持原来的线共点、点共线的关系不变，保持原来相互平行的线仍然平行，保持原来的中点仍然是中点，保持原来在一直线上几段线段之间的比例关系不变。仿射变换可以通过一系列的原子变换的复合来实现，包括：平移 (Translation)、缩放 (Scale)、翻转 (Flip)、旋转 (Rotation) 和剪切 (Shear)。

仿射变换可以用下面公式表示：

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (2-5)$$

可见，相似变换有六个参数： $a_{11}, a_{12}, a_{21}, a_{22}, t_x, t_y$ ，容易看出，仿射变换是相似变换的一般形式，当 $s_x = s_y$ 时，仿射变换退化为相似变换。

2.3.3 投影变换 (Projective Transformation)

如果一幅图像中的直线经过后映射到另一幅图像上仍为直线，但平行关系基本不保持，则这种变换称为投影变换(Projective Transform)。二维平面投影变换是关于齐次三维矢量的线性变换，在齐次坐标系下，二维平面上的投影变换具体可用非奇异 3×3 矩阵形式来描述。

在欧式几何中，如果说在一个平面中任意两条直线必然相交于一点，是不能成立的，因为存在一个例外，即两条直线可以是平行的。这是欧式几何的麻烦之处。而换个说法，可以说平行的两条直线仍然相交于一点，只是它们相交于无限远处。这个在无限远处的交点，叫做理想点(ideal points)。由于增加了“理想点”，我们将传统的欧式空间拓展到了投影空间。在投影空间中，平面上任意两条直线的确是必然相交于一点的。

正侧投影 (Orthographic Projection) 有些地方也翻译成正交投影。

正交是世界中的物体按照平行的光线投射到一张纸上(你的画布)，仿佛被压缩在上面，无论这个物体在世界中多远，投射结果还是原来的大小, 2d 游戏多用此种投影。如图 2-2:

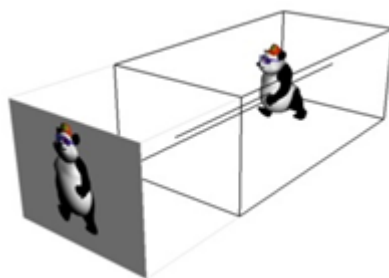


图 2- 2: 正侧投影

透视投影 (Perspective Projection) 透视比较接近我们现实的方式了，你的眼睛发出的光形成一个夹角，离你的眼睛越近，东西越大，范围越小；反之离眼睛越远，东西越小，视野越开阔。因此多用于 3D 中。如图 2-3:



图 2- 3：透视投影

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 3×3 大小的投影变换矩阵的形式。

2.3.4 非线性变换

非线性变换又称为弯曲变换(Curved Transform)，经过非线性变换，一幅图像上的直线映射到另一幅图像上不一定是直线，可能是曲线,在二维空间中，可以用以下公式表示：

$$\left(x',y'\right)=F\left(x,y\right) \tag{2-6}$$

式中， F 表示把一幅图像映射到另一幅图像上的任意一种函数形式。多项式变换是典型的非线性变换，如二次、三次函数及样条函数，有时也使用指数函数。

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说明了图像之间的映射关系将会不同。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。

表 2-1 总结了各种变换模型可以表示的变换，可以看出，非线性变换可以描述更多的变换，但也更加复杂。在对需配准的图像进行空间变换后，要对变换图像进行重新取样，以取得变换后的像素值。

2.4 图像配准流程

目前，很难找到一种普适的方法能够应对所有的配准情况，任何一种配准算法都必须考虑图像的成像原理、几何变形、噪声影响、配准精度等因素。不过，从原理上将，图像配准算法可以大致分为以下四个步骤：

（1）特征提取

采用人工或者自动的方法检测图像中的不变特征，如：闭合区域、边缘、轮廓、角点等。特征提取算法需要满足三个条件

（a）显著性，所提取的特征应该也是比较明显的，分布广泛的、易于提取的特征；

（b）抗噪性，具有较强的噪声抑制能力且对成像条件的变化不敏感；

（c）一致性，能准确地检测出两幅图像的共有特征；

（2）特征匹配

通过特征描述算作及相似性度量来建立所提取的特征之间的对应关系。特征匹配常用到的区域灰度、特征向量空间分布和特征符号描述等信息。某些算法在进行特征匹配的同时也完成了变换模型参数的估计；

（3）变换模型估计

指根据待配准图像与参考图像之间的几何畸变的情况，选择能最佳拟合两幅图像之间变化的几何变换模型，可以分为全局映射模型和局部映射模型。其中，全局映射模型利用所有控制点信息进行全局参数估计；局部映射模型利用图像局部的特征分别进行局部参数估计。常见的变换模型包括仿射变换、透视变换、多项式变换等，其中最常用的是仿射变换和多项式变换。

（4）坐标变换与插值

将输入图像做对应的参数变换，使它与参考图像处于同一个坐标系下。由于图像变换后的坐标点不一定是整数，因此，需要考虑一定的插值处理操作。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；

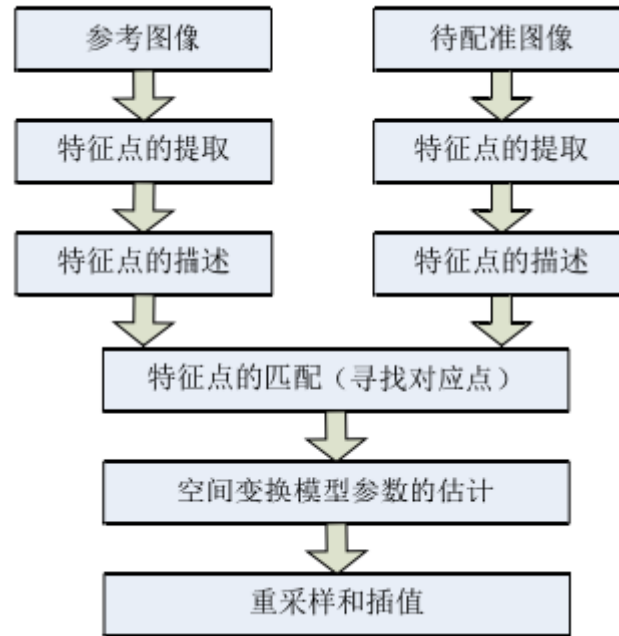


图 2-4：图形配准大致流程

2.5 图像配准评价标准

2.5.1 准确率

准确率 A_c 定义如下：

$$A_c = N_c / (N_c + N_e) \quad (2-7)$$

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。

2.5.2 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。

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

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 数值越小，则表示精度越高，效果越好。

2.5.3 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，它常简单地通过均方差 (MSE) 进行定义。两个 $m \times n$ 单色图像 I 和 K ，如果一个为另外一个的噪声近似，那么它们的均方差定义为：

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

峰值信噪比定义为：

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

其中， MAX_I 是表示图像点颜色的最大数值，如果每个采样点用 8 位表示，那么就是 255。更为通用的表示是，如果每个采样点用 B 位线性脉冲编调制表示，那么 MAX_I 就是 $2^B - 1$ 。

对于每点有 RGB 三个值的彩色图像来说峰值信噪比的定义类似，只是均方差是所有方差之和除以图像尺寸再除以 3。

PSNR 是最普遍和使用最为广泛的一种图像客观评价指标，然而它是基于对应像素点间的误差，即基于误差敏感的图像质量评价。由于并未考虑到人眼的视觉特性（人眼对空间频率较低的对比差异敏感度较高，人眼对亮度对比差异的敏感度较

色度高，人眼对一个区域的感知结果会受到其周围邻近区域的影响等)，因而经常出现评价结果与人的主观感觉不一致的情况。

2.5.4 相似度检测 SSIM

SSIM (Structural Similarity) 结构相似性。 μ_X 、 μ_Y 分别表示图像 X 和 Y 的均值， σ_X 、 σ_Y 分别表示图像 X 和 Y 的方差， σ_{XY} 表示图像 X 和 Y 的协方差。

SSIM 分别从亮度、对比度、结构三方面度量图像相似性。

$$l(X, Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \quad (2-11)$$

$$c(X, Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \quad (2-12)$$

$$s(X, Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3} \quad (2-13)$$

C_1 、 C_2 、 C_3 为常数，为了避免分母为 0 的情况，通常取 $C_1 = (K_1 * L)^2$, $C_2 = (K_2 * L)^2$, $C_3 = \frac{C_2}{2}$ ，一般地取 $K_1 = 0.01$, $K_2 = 0.03$, $L = 255$ ，则

$$SSIM(X, Y) = l(X, Y) * c(X, Y) * s(X, Y) \quad (2-14)$$

SSIM 取值范围[0,1]，值越大，表示图像失真越小。

SSIM 特殊形式如下：

$$SSIM(X, Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \quad (2-15)$$

2.5.3 配准速度

配准速度是指从输入两幅图像到得到图像配准结果的整个过程所花费的时间，时间越长，则配准速度越慢，反之，速度越快。

2.5.3 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[86]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，差值图像越黑表示配准结果越好

第三章 尺度不变特征变换(SIFT)

3.1 SIFT 综述

尺度不变特征转换(Scale-invariant feature transform 或 SIFT)是一种电脑视觉的算法用来侦测与描述影像中的局部性特征,它在空间尺度中寻找极值点,并提取出其位置、尺度、旋转不变量,此算法由 David Lowe 在 1999 年所发表[50],2004 年完善总结[52]。在 Mikolajczyk 对包括 Sift 算子在内的十种局部描述子所做的不变性对比实验中,Sift 及其扩展算法已被证实同类描述子中具有最强的健壮性。

其应用范围包含物体辨识、机器人地图感知与导航、影像缝合、3D 模型建立、手势辨识、影像追踪和动作比对。**SIFT** 特征是基于物体上的一些局部外观的兴趣点而与影像的大小和旋转无关。对于光线、噪声、些微视角改变的容忍度也相当高。基于这些特性,它们是高度显著而且相对容易撷取,在母数庞大的特征数据库中,很容易辨识物体而且鲜有误认。使用 **SIFT** 特征描述对于部分物体遮蔽的侦测率也相当高,甚至只需要 3 个以上的 **SIFT** 物体特征就足以计算出位置与方位。在现今的电脑硬件速度下和小型的特征数据库条件下,辨识速度可接近即时运算。**SIFT** 特征的信息量大,适合在海量数据库中快速准确匹配。

此算法有其专利,专利拥有者为英属哥伦比亚大学。

3.2 SIFT 特点

SIFT 算法的特点有:

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

SIFT 算法可以解决的问题：

目标的自身状态、场景所处的环境和成像器材的成像特性等因素影响图像配准/目标识别跟踪的性能。而 SIFT 算法在一定程度上可解决：

1. 目标的旋转、缩放、平移（RST）
2. 图像仿射/投影变换（视点 viewpoint）
3. 光照影响（illumination）
4. 目标遮挡（occlusion）
5. 杂物场景（clutter）
6. 噪声

SIFT 在图像的不变特征提取方面拥有无与伦比的优势，但并不完美，仍然存在：

1. 实时性不高。2. 有时特征点较少。3. 对边缘光滑的目标无法准确提取特征点等缺点。对模糊的图像和边缘平滑的图像，检测出的特征点过少，对圆更是无能为力。

3.3 SIFT 流程

Sift 特征匹配算法主要包括两个阶段，一个是 Sift 特征的生成，即从多幅图像中提取对尺度缩放、旋转、亮度变化无关的特征向量；第二阶段是 Sift 特征向量的匹配。

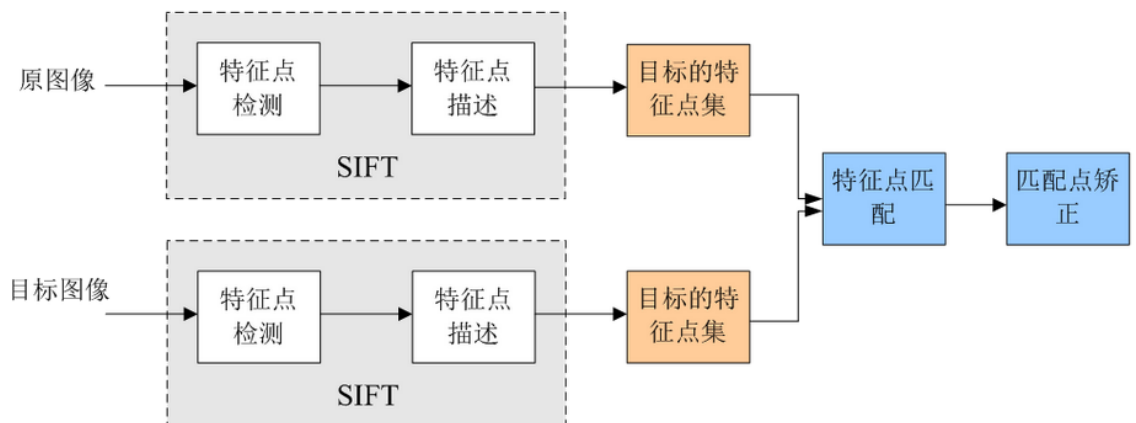


图 3-1：SIFT 图形配准流程

3.3.1 尺度空间的生成

在一定的范围内，无论物体是大还是小，人眼都可以分辨出来。然而计算机要有相同的能力却不是那么的容易，在未知的场景中，计算机视觉并不能提供物体的尺度大小，其中的一种方法是把物体不同尺度下的图像都提供给机器，让机器能够对物体在不同的尺度下有一个统一的认知。在建立统一认知的过程中，要考虑的就是在图像在不同的尺度下都存在的特征点。所以，尺度空间的生成是一个初始化操作，尺度空间理论目的是模拟图像数据的多尺度特征。

高斯卷积核是实现尺度变换的唯一线性核，于是一副二维图像的尺度空间定义为：

$$l(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (3-1)$$

其中 $G(x, y, \sigma)$ 是尺度可变高斯函数， $G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$ ，在二维空间中，这个公式生成的曲面的等高线是从中心开始呈正态分布的同心圆。

(x, y) 是空间坐标， σ 是尺度坐标。 σ 大小决定图像的平滑程度，大尺度对应图像的概貌特征，尺度越大图像越模糊。小尺度对应图像的细节特征。大的 σ 值对应粗糙尺度(低分辨率)，反之，对应精细尺度(高分辨率)。分布不为零的像素组成的卷积矩阵与原始图像做变换。每个像素的值都是周围相邻像素值的加权平均。原始像素的值有最大的高斯分布值，所以有最大的权重，相邻像素随着距离原始像素越来越远，其权重也越来越小。这样进行模糊处理比其它的均衡模糊滤波器更高地保留了边缘效果。理论上讲，图像中每点的分布都不为零，这也就是说每个像素的计算都需要包含整幅图像。在实际应用中，在计算高斯函数的离散近似时，在大概 3σ 距离之外的像素都可以看作不起作用，这些像素的计算也就可以忽略。

构建尺度空间的目的是为了检测出在不同的尺度下都存在的特征点，而检测特征点较好的算子是 $\nabla^2 G$ (高斯拉普拉斯, LoG):

$$\Delta^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \quad (3-2)$$

事实上，2002 年 Mikolajczyk 在详细的实验比较中发现尺度归一化的高斯拉

普拉斯函数 $\sigma^2 \nabla^2 G$ 的极大值和极小值同其它的特征提取函数，例如：梯度，Hessian 或 Harris 角特征比较，能够产生最稳定的图像特征。图像与 LoG 算子进行卷积运算就是求两者之间的相似性，当两者的卷积运算响应达到最大值时，在图像中对应的位置即为特征点的位置。

使用 LoG 虽然能较好的检测到图像中的特征点，但是其运算量过大，通常可使用 DoG（差分高斯，Difference of Gaussina）来近似计算 LoG。这得益于 Lindeberg 早在 1994 年的发现[45]。

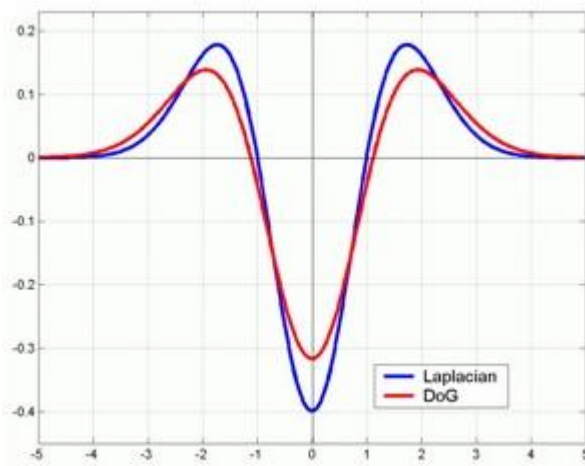


图 3-2：用差分高斯 DoG 近似高斯拉普拉斯 LoG

DoG 曲线表示的是高斯差分算子，而另一条曲线表示的是高斯拉普拉斯算子高斯差分尺度空间（DOG scale-space）：

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

利用不同尺度的高斯差分核与图像卷积生成。

关于尺度空间的理解说明： $2k\sigma$ 中的 2 是必须的，尺度空间是连续的。在 Lowe 的论文中，将第 0 层的初始尺度定为 1.6（最模糊），图片的初始尺度定为 0.5（最清晰）。在检测极值点前对原始图像的高斯平滑以致图像丢失高频信息，所以 Lowe 建议在建立尺度空间前首先对原始图像长宽扩展一倍，以保留原始图像信息，增加特征点数量。

图像金字塔的建立：对于一幅图像 I ，建立其在不同尺度(scale)的图像，也成为子八度（octave），这是为了 scale-invariant，也就是在任何尺度都能够有对应的特征点，第一个子八度的 scale 为原图大小，后面每个 octave 为上一个 octave 降

采样的结果，即原图的 1/4（长宽分别减半），构成下一个子八度（高一层金字塔）。

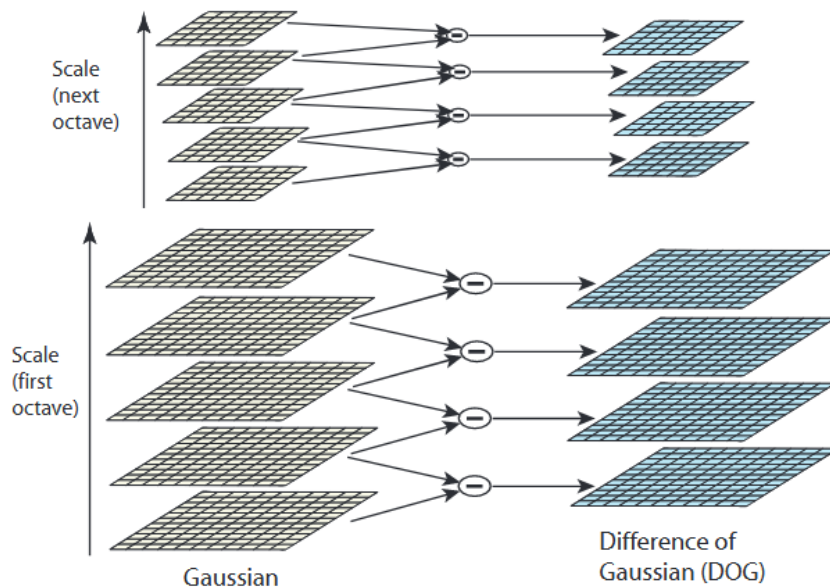


图 3- 3：高斯金字塔的构成

由图片 size 决定建几个塔，每塔几层图像(S 一般为 3-5 层)。0 塔的第 0 层是原始图像(或称 double 后的图像)，往上每一层是对其下一层进行 Laplacian 变换(高斯卷积，其中 σ 值渐大，例如可以是 $\sigma, k*\sigma, k*k*\sigma \dots$)，直观上看来越往上图片越模糊。塔间的图片是降采样关系，例如 1 塔的第 0 层可以由 0 塔的第 3 层 down sample 得到，然后进行与 0 塔类似的高斯卷积操作。

3.3.2 检测尺度空间极值点

为了寻找尺度空间的极值点，每一个采样点要和它所有的相邻点比较，看其是否比它的图像域和尺度域的相邻点大或者小。如图所示，中间的检测点和它同尺度的 8 个相邻点和上下相邻尺度对应的 9×2 个点共 26 个点比较，以确保在尺度空间和二维图像空间都检测到极值点。一个点如果在 DOG 尺度空间本层以及上下两层的 26 个领域中是最大或最小值时，就认为该点是图像在该尺度下的一个特征点，如此可以保证检测到的关键点在尺度空间和二维图像空间上都是局部极值点。如图 3-4 所示。

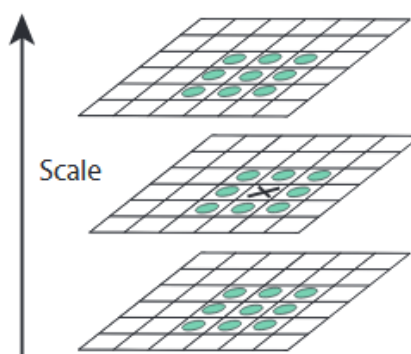


图 3-4: 同一组中相邻尺度（上下层）之间进行寻找

3.3.3 精确定位极值点

通过比较检测得到的 DoG 的局部极值点是在离散的空间搜索得到的，由于离散空间是对连续空间采样得到的结果，因此在离散空间找到的极值点不一定是真正意义上的极值点，因此要设法将不满足条件的点剔除掉。

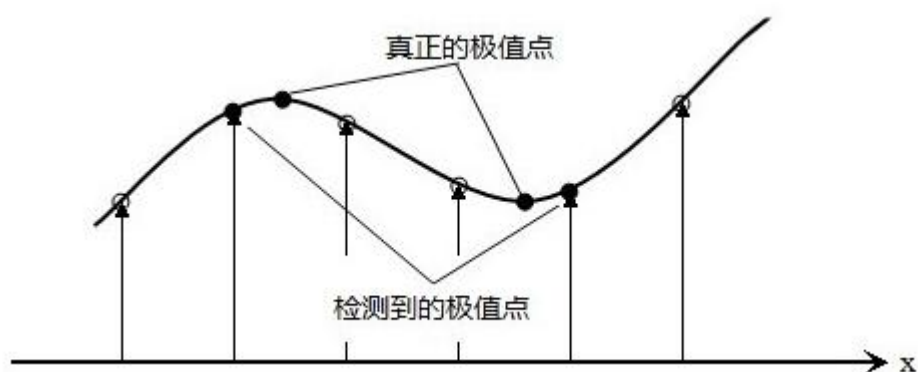


图 3-5: 检测到的极值点不一定是真正的极值点

DOG 值对噪声和边缘比较敏感，所以在第 2 步的尺度空间中检测到的局部极值点还要经过进一步的筛选，去除不稳定和错误检测出的极值点，另一点就是在构建高斯金字塔过程中采用了下采样的图像，在下采样图像中提取的极值点对应在原始图像中的确切位置，也是要在本步骤中解决的问题。

可以通过尺度空间 DoG 函数进行三维二次函数曲线拟合寻找极值点，这一步的本质是去掉 DoG 局部曲率非常不对称的点。

要剔除掉的不符合要求的点主要有两种：

1. 低对比度的特征点；
2. 不稳定的边缘响应点。

对于低对比度的点：通过拟和三维二次函数以精确确定关键点的位置和尺度

(达到亚像素精度),同时去除低对比度的关键点和不稳定的边缘响应点(因为 DoG 算子会产生较强的边缘响应),以增强匹配稳定性、提高抗噪声能力,在这里使用近似 Harris Corner 检测器。

1.候选特征点 x 空间尺度函数泰勒展开式如下:

$$D(x)=D+\frac{\partial D^T}{\partial x}\Delta x+\frac{1}{2}\Delta x^T\frac{\partial^2 D}{\partial x^2}\Delta x \quad (3-4)$$

对比度为 $D(x)$ 绝对值 $|D(x)|$, Δx 是偏移量。因为候选点认为是极值点,

所以对式求导,其为 0,得到精确的位置,即

$$\Delta x = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \quad (3-5)$$

2.把公式(3-5)代入(3-4),即在 DoG Space 的极值点处 $D(x)$ 取值,只取前两

项,得 $D(x)=D+\frac{1}{2}\frac{\partial D^T}{\partial x}\hat{x}$, 设对比度的阈值为 T

若 $|D(\hat{x})|\geq T$, 该特征点就保留下来,否则丢弃。

对于不稳定的边缘响应点:一个定义不好的高斯差分算子的极值在横跨边缘的地方有较大的主曲率,而在垂直边缘的方向有较小的主曲率。候选特征点的 DoG 函数 $D(x)$ 的主曲率与一个 2×2 的 Hessian 矩阵 H 的特征值成正比:

$$H=\begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix} \quad (3-6)$$

令 α 为 H 的较大特征值, β 为较小的特征值,是 x 和 y 方向的梯度,则:

$$\begin{aligned} Tr(H) &= D_{xx} + D_{yy} = \alpha + \beta \\ Det(H) &= D_{xx}D_{yy} - (D_{xy})^2 \end{aligned} \quad (3-7)$$

令 $\alpha=r\beta$, r 是最大特征值与最小特征值的比值。则:

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \quad (3-8)$$

$(r+1)^2/r$ 的值在两个特征值相等的时候最小,随着 r 的增大而增大, r 值越大,说明两个特征值的比值越大,即在某一个方向的梯度值越大,而在另一个方向

的梯度值越小，而边缘恰恰就是这种情况。所以为了剔除边缘响应点，需要让该比值小于一定的阈值，因此，为了检测主曲率是否在某域值 T_r 下，只需检测

$$\frac{Tr(H)^2}{Det(H)} < \frac{(T_r + 1)^2}{T_r} \quad (3-9)$$

If $(\alpha + \beta)^2 / \alpha\beta > (r+1)^2 / r$, throw it out, 在 lowe 的文章中[52], 取 $r=10$.

3.3.4 求取特征点的主方向

上一步中确定了每幅图中的特征点，为每个特征点计算一个方向，依照这个方向做进一步的计算，利用关键点邻域像素的梯度方向分布特性为每个关键点指定方向参数，使算子具备旋转不变性。

稳定的极值点是在不同尺度空间下提取的，这保证了关键点的尺度不变性。为关键点分配方向信息所要解决的问题是使得关键点对图像角度和旋转具有不变性。方向的分配是通过求每个极值点的梯度来实现的。

对于任一关键点，其梯度幅值表述为：

$$m(x, y) = \sqrt{(L(x+1, y) - L(x, y+1))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (3-10)$$

梯度方向为：

$$\theta(x, y) = \alpha \tan 2((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \quad (3-11)$$

其中 L 所用的尺度为每个关键点各自所在的尺度。

分配给关键点的方向并不直接是关键点的梯度方向，而是按照一种梯度方向直方图的方式给出的。要采集其所在高斯金字塔图像 3σ 邻域窗口内像素的梯度和方向分布特征。

具体的方法是：计算以关键点为中心的邻域内所有点的梯度方向梯度直方图的范围是 $0 \sim 360$ 度，其中每 45 度一个柱，总共 8 个柱，或者每 10 度一个柱，总共 36 个柱。将梯度方向直方图中纵坐标最大的项代表的方向分配给当前关键点作为主方向，若在梯度直方图中存在一个相当于主峰值 80% 能量的峰值，则将这个方向认为是关键点的辅方向。辅方向的设计可以增强匹配的鲁棒性，Lowe 指出，大概有 15% 的关键点具有辅方向，而恰恰是这 15% 的关键点对稳定匹配起到关键

作用。

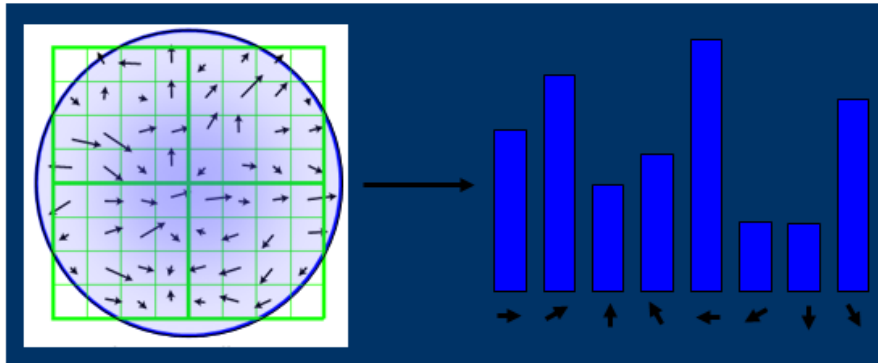


图 3-6：直方图中的峰值作为主方向，其他的达到最大值 80% 的方向可以作为辅助方向

至此，图像的关键点已经检测完毕，每个关键点有三个信息 (x, y, σ, θ) ：位置，所处尺度、方向，由此可以确定一个 SIFT 特征区域。一个 SIFT 特征区域由三个值表示，中心表示特征点位置，半径表示关键点的尺度，箭头表示主方向。

3.3.5 关键点描述子的生成

对关键点的描述是后续实现匹配的关键步骤，描述其实就是一种以数学方式定义关键的过程。描述子不但包含关键点，也包括关键点周围对其有贡献的邻域点。描述的思路是：对关键点周围像素区域分块，计算块内梯度直方图，生成具有独特性的向量，这个向量是该区域图像信息的一种抽象表述。

特征描述符的生成大致有三个步骤：

1. 校正旋转主方向，确保旋转不变性。
2. 生成描述子，最终形成一个 128 维的特征向量
3. 归一化处理，将特征向量长度进行归一化处理，进一步去除光照的影响。

首先将坐标轴旋转为关键点的方向，以确保旋转不变性。以关键点为中心取 8×8 的窗口。图 3-7 中，左边部分的中央为当前关键点的位置，每个小格代表关键点邻域所在尺度空间的一个像素，利用公式求得每个像素的梯度幅值与梯度方向，箭头方向代表该像素的梯度方向，箭头长度代表梯度模值，然后用高斯窗口对其进行加权运算。

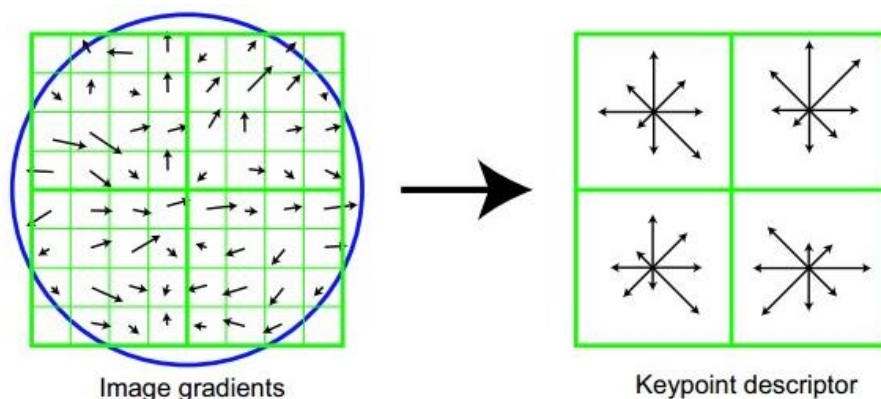


图 3- 7：描述子的生成

图中蓝色的圈代表高斯加权的范围（越靠近关键点的像素梯度方向信息贡献越大）。然后在每 4×4 的小块上计算 8 个方向的梯度方向直方图，绘制每个梯度方向的累加值，即可形成一个种子点，如图 3-7 右部分示。此图中一个关键点由 2×2 共 4 个种子点组成，每个种子点有 8 个方向向量信息。这种邻域方向性信息联合的思想增强了算法抗噪声的能力，同时对于含有定位误差的特征匹配也提供了较好的容错性。

David G.Lowed[51]的实验结果表明：对每个关键点，采用 $4 \times 4 \times 8$ 共 128 维向量的描述子进行关键点表征，综合效果最佳：计算 keypoint 周围的 16×16 的 window 中每一个像素的梯度，而且使用高斯下降函数降低远离中心的权重。

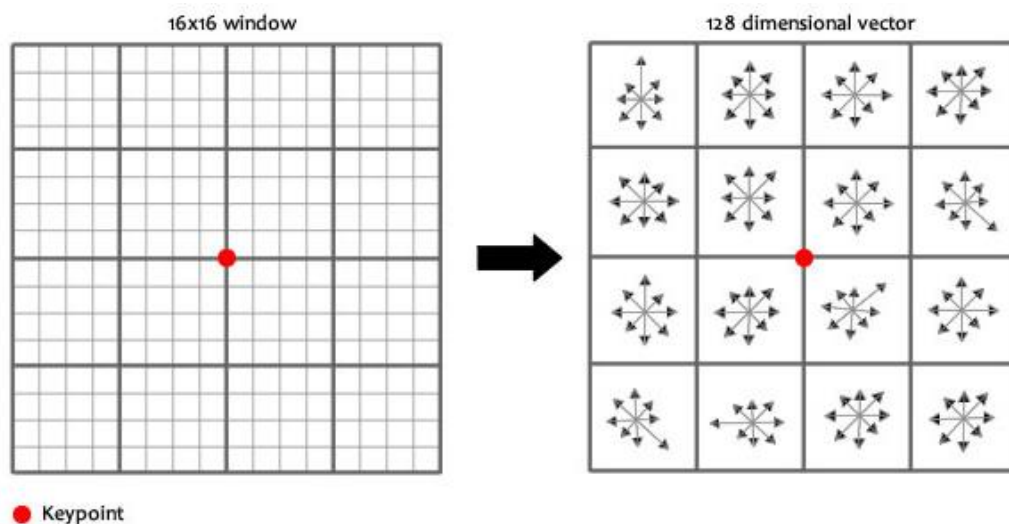


图 3-8: 128 维向量描述子进行关键点表征

在每个 4×4 的 $1/16$ 象限中，通过加权梯度值加到直方图 8 个方向区间中的一个，计算出一个梯度方向直方图。这样就可以对每个 feature 形成一个 $4 \times 4 \times 8 = 128$ 维的描述子，每一维都可以表示 4×4 个格子中一个的 scale/orientation. 将这个向量

归一化之后，就进一步去除了光照的影响。

理想的特征点描述算子也应该具有鲁棒性和独特性的特征。鲁棒性是指描述算子能够在图像几何变形、失真等条件的影响下，仍然具备相对的不变性；而独特性是指不同的特征点之间，描述算子也应该是不同的，这样才能具备不同特征点的可识别性。描述算子的鲁棒性和独特性相互影响而且又具有矛盾性，而且会影响匹配的速度：包含的信息少则鲁棒性更好一些，而且匹配速度快，但相应的独特性就会稍差；而包含信息丰富的话独特性就会更佳，鲁棒性就会相应的减弱，而且给匹配速度也会带来一定的影响。因此，可以将鲁棒性、独特性以及匹配速度作为评价特征描述算子的标准。

3.3.6 根据 SIFT 进行 Match

生成了 A、B 两幅图的描述子，（分别是 $k1*128$ 维和 $k2*128$ 维），就将两图中各个 scale（所有 scale）的描述子进行匹配，匹配上 128 维即可表示两个特征点 match 上了。

实际计算过程中，为了增强匹配的稳健性，Lowe 建议对每个关键点使用 4×4 共 16 个种子点来描述，这样对于一个关键点就可以产生 128 个数据，即最终形成 128 维的 SIFT 特征向量。此时 SIFT 特征向量已经去除了尺度变化、旋转等几何变形因素的影响，再继续将特征向量的长度归一化，则可以进一步去除光照变化的影响。当两幅图像的 SIFT 特征向量生成后，下一步我们采用关键点特征向量的欧式距离来作为两幅图像中关键点的相似性判定度量。取图像 1 中的某个关键点，并找出其与图像 2 中欧式距离最近的前两个关键点，在这两个关键点中，如果最近的距离除以次近的距离少于某个比例阈值，则接受这一对匹配点。降低这个比例阈值，SIFT 匹配点数目会减少，但更加稳定。为了排除因为图像遮挡和背景混乱而产生的无匹配关系的关键点，Lowe 提出了比较最近邻距离与次近邻距离的方法，距离比率 ratio 小于某个阈值的认为是正确匹配。因为对于错误匹配，由于特征空间的高维性，相似的距离可能有大量其他的错误匹配，从而它的 ratio 值比较高。Lowe 推荐 ratio 的阈值为 0.8。但作者对大量任意存在尺度、旋转和亮度变化的两幅图片进行匹配，结果表明 ratio 取值在 0.4~0.6 之间最佳，小于 0.4 的很少有匹配点，大于 0.6 的则存在大量错误匹配点。作者建议 ratio 的取值原则如下：

ratio=0.4 对于准确度要求高的匹配;

ratio=0.6 对于匹配点数目要求比较多的匹配;

ratio=0.5 一般情况下。

也可按如下原则:当最近邻距离 <200 时 ratio=0.6, 反之 ratio=0.4。ratio 的取值策略能排除错误匹配点。

3.4 本章小结

Sift 是 David Lowe 于 1999 年提出的局部特征描述子, 并于 2004 年进行了更深入的发展和完善。Sift 特征匹配算法可以处理两幅图像之间发生平移、旋转、仿射变换情况下的匹配问题, 具有很强的匹配能力。在 Mikolajczyk 对包括 Sift 算子在内的十种局部描述子所做的不变性对比实验中, Sift 及其扩展算法已被证实在同类描述子中具有最强的健壮性。

SIFT 算法的实质是在不同的尺度空间上查找关键点(特征点), 并计算出关键点的方向。Sift 算法就是用不同尺度(标准差)的高斯函数对图像进行平滑, 然后比较平滑后图像的差别, 差别大的像素就是特征明显的点。SIFT 所查找到的关键点是一些十分突出, 不会因光照, 仿射变换和噪音等因素而变化的点, 如角点、边缘点、暗区的亮点及亮区的暗点等。

第四章 快速鲁棒性特征点检测 SURF

4.1 surf 综述

如果说 SIFT 算法中使用 DOG 对 LOG 进行了简化，提高了搜索特征点的速度，那么 SURF 算法则是对 DoH 的简化与近似。虽然 SIFT 算法已经被认为是最有效的，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和专用图像处理器的配合，SIFT 算法以现有的计算机仍然很难达到实时的程度。对于需要实时运算的场合，如基于特征点匹配的实时目标跟踪系统，每秒要处理 8-24 帧的图像，需要在毫秒级内完成特征点的搜索、特征矢量生成、特征矢量匹配、目标锁定等工作，这样 SIFT 算法就很难适应这种需求了。SURF 借鉴了 SIFT 中简化近似的思想，把 DoH 中的高斯二阶微分模板进行了简化，使得模板对图像的滤波只需要进行几个简单的加减法运算，并且，这种运算与滤波器的尺度无关。实验证明，SURF 算法较 SIFT 在运算速度上要快 3 倍左右。

Surf 算法中要用到积分图像的概念。借助积分图像，图像与高斯二阶微分模板的滤波转化为对积分图像的加减运算。积分图像（Integral Image）的概念是由 viola 和 Jones 提出来的，而将类似积分图像用于盒子滤波是由 Simard 等人提出。

4.2 SURF 特点

SURF 特点：

- 1.使用积分图像完成图像卷积（相关）操作；
- 2，使用 Hessian 矩阵检测特征值；
- 3，使用基于分布的描述符（局部信息）。

4.3 Surf 算法原理

4.3.1 构建 Hessian 矩阵，计算特征值 α

surf 构造的金字塔图像与 sift 有很大不同，就是因为这些不同才加快了其检测

的速度。Sift 采用的是 DOG 图像，而 surf 采用的是 Hessian 矩阵行列式近似值图像。我们可以利用 Hessian 矩阵行列式的极大值检测斑点。首先是图像中某个像素点的 Hessian 矩阵：

$$H(f(x, y)) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} \quad (4-1)$$

海森矩阵的混合偏导数是海森矩阵非主对角线上的元素。假如他们是连续的，那么求导顺序没有区别，海森矩阵的行列式，可用于分辨 f 的临界点是属于鞍点还是极值点。

每一个像素点都可以求出一个 Hessian 矩阵。但是由于我们的特征点需要具备尺度无关性，所以在进行 Hessian 矩阵构造前，需要对其进行高斯滤波。这样，经过滤波后在进行 Hessian 的计算，其公式如下：

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

其中， L_{xx} 、 L_{xy} 、 L_{yy} 为高斯滤波后图像 $g(\sigma)$ 在各个方向的二阶导数，也可看成是高斯二阶微分与图像的卷积。 $g(\sigma)$ 是一幅图像在不同解析度下的表示，利用高斯核函数与图像函数在点 (x, y) 的卷积来实现。

二阶导数计算方法为：

$$\begin{aligned} d^2 L(x) / dx^2 &= (L(x+1) - L(x)) - (L(x) - L(x-1)) \\ &= -2 * L(x) + L(x+1) + L(x-1) \end{aligned} \quad (4-3)$$

其中 $L(x) = g(h(x))$ 。 $h(x)$ 为原始图像的灰度值， $L(x)$ 是将 $h(x)$ 高斯滤波处理后的图像。

为了找出图像中的特征点，需要对原图进行变换，在 sift 算法中，是在 DOG 图像中进行在 surf 算法中，该变换图就是原图每个像素的 Hessian 矩阵行列式的近似值构成的。公式为：

$$\det(H_{approx}) = L_{xx}L_{yy} - (0.9L_{xy})^2 \quad (4-4)$$

其中，0.9 为原文作者给出的一个经验值。

由于求 Hessian 时要先高斯平滑，然后求二阶导数，这在离散的像素点是用模板卷积形成的，这 2 种操作合在一起用一个模板代替就可以了。又为了将模板与图像的卷积转换为盒子滤波运算，我们对高斯二阶微分模板进行简化，使得简化后的模板只是由几个矩形区域组成，矩形区域内填充同一值，如下图所示，在简化模板中白色区域的值为正数，黑色区域的值为负数，灰度区域的值为 0。，比如说 y 方向上的模板如下，以 9*9 的模板为例：

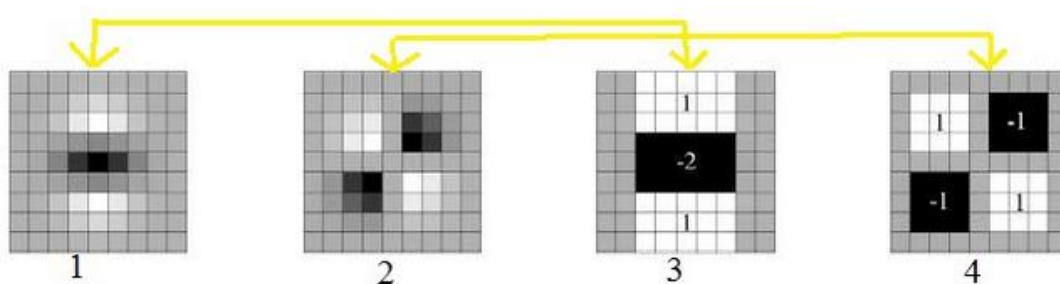


图 4-1：使用盒状滤波器代替高斯核

上图 4-1 中 1 为 y 方向的先高斯滤波然后二阶求导的处理，近似处理为图 3，图 2 为 x 和 y 方向上的先高斯滤波然后二阶混合偏导，近似为图 4。

有了这个近似的模板以后，计算高斯滤波和二阶导数两个步骤就可以一个步骤完成，同时，为了提高计算效率，还引入了积分图像的概念并可以应用于近似后的图像，提高了速度。

积分图像的概念是由 Viola 和 Jones 提出的。就是当前的每个像素的灰度都是它与坐标原点 (0,0) 形成的对角线的矩形内的所有像素的灰度值之和。如图 4-2：

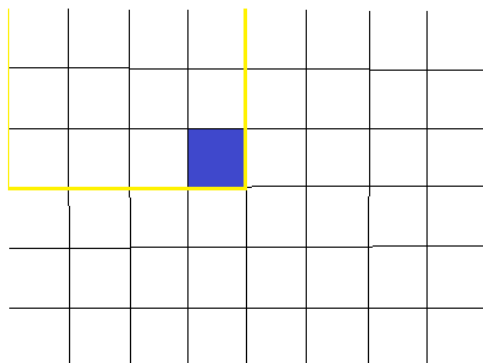


图 4-2：积分图中蓝色像素的积分图像的灰度值就是黄色框内的所有像素的灰度值之和。其他像素同样的计算方法。

有了积分图像的概念，在计算某个矩形框内的像素灰度值之和时，就可以很简

单的得出了。如图 4-3:

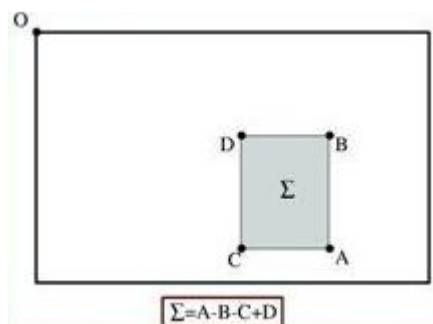


图 4-3: 积分图像, 避免了重复计算。灰色矩形框内的灰度值之和就是 $A+D-C-B$, 就是很简单的一个矩形面积计算公式, A 代表 AO 对角线矩形的面积, 以此类推。

这样, 计算像素的 H 矩阵时, 将会非常的方便, 比如图 4-1 中-2 那一块区域的灰度值, 只需要知道 4 个顶点像素的积分图像的值, 便可以求出那一块的灰度值之和。求积分图像, 只需要对原图像所有像素进行一遍扫描。一旦积分图计算好了, 计算图像内任何矩形区域的像素值的和只需要三个加法。借助积分图像, 图像与高斯二阶微分模板的滤波转化为对积分图像的加减运算。

至此, 我们求出了原图的 Hessian 矩阵, 然后计算行列式的近似值就得出了我们需要寻找特征点的变换图像。

4.3.2 构造高斯金字塔

在计算视觉领域, 尺度空间被象征性的表述为一个图像金字塔, 其中, 输入图像函数反复与高斯函数的核卷积并反复对其进行二次抽样, 这种方法主要用于 Sift 算法的实现, 但每层图像依赖于前一层图像, 并且图像需要重设尺寸, 因此, 这种计算方法运算量较大。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降采样 (1/4 大小); 在每一组里面的几幅图像中, 他们的大小是一样的, 不同的是他们采用的尺度 σ 不同。而且在模糊的过程中, 他们的高斯模板大小总是不变的, 只是尺度 σ 改变。

对于 surf 算法, 由于采用了盒子滤波和积分图像, 所以, 我们并不需要像 SIFT 算法那样去直接建立图像金字塔, 而是采用不断增大盒子滤波模板的尺寸的间接方法。图像的大小总是不变的, 改变的只是高斯模糊模板的尺寸, 当然, 尺度 σ 也是在改变的。

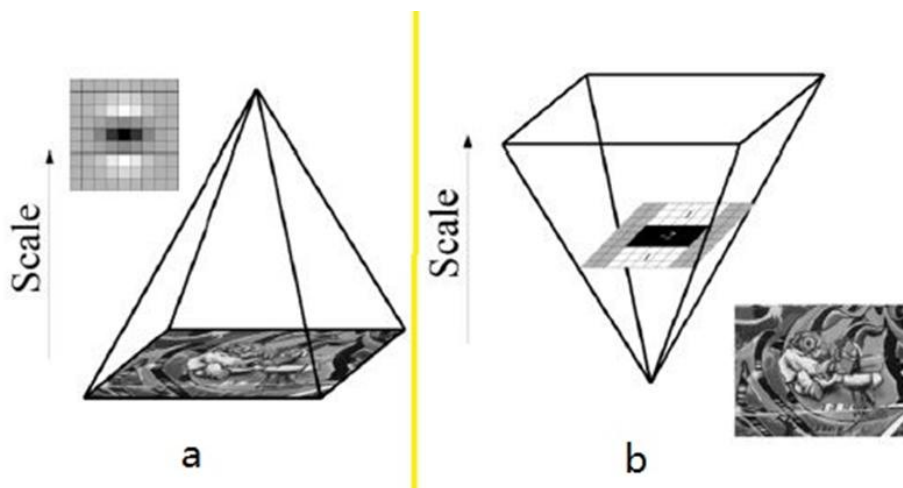


图 4-4: SIFT 中使用的传统图像金字塔与 SURF 采用的图像金字塔对比

SURF 算法申请增加图像核的尺寸，这也是 SIFT 算法与 SURF 算法在使用金字塔原理方面的不同。算法允许尺度空间多层图像同时被处理，不需对图像进行二次抽样，从而提高算法性能。图 4-4(a)是传统方式建立一个如图所示的金字塔结构，图像的尺寸是变化的，并且会反复使用高斯函数对子层进行平滑处理，图 4-4(b)说明 Surf 算法使原始图像保持不变而只改变滤波器大小。因为 surf 算法没有了降采样的过程，因此处理速度得到提高。

与 SIFT 算法类似，我们需要将尺度空间划分为若干组（Octaves）。不同的组代表了逐步放大的滤波模板对同一输入图像进行滤波的一系列响应图。每个组又由若干固定的层组成。由于积分图像离散化的原因，两个层之间的最小尺度变化量是由高斯二阶微分滤波器在微分方向上对正负斑点响应长度 l_0 决定的，它是盒子滤波器模板尺寸的 $1/3$ 。对于 9×9 的模板，它的 $l_0 = 3$ 。下一层的响应长度至少应该在的基础上增加 2 个像素，以保证一边一个像素，即 $l_0 = 5$ 。这样模板的尺寸就为 15×15 。以此类推，我们可以得到一个尺寸增大模板序列，它们的尺寸分别为： $9 \times 9, 15 \times 15, 21 \times 21, 27 \times 27$ ，滤波器尺寸依次增加 6；黑色、白色区域的长度增加偶数个像素，以保证一个中心像素的存在。

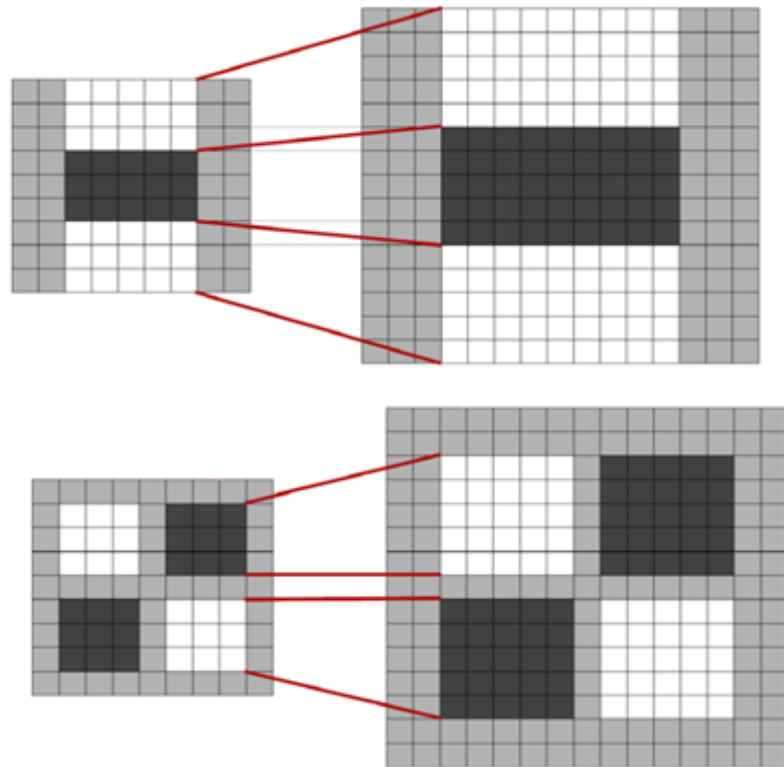


图 4-5: 不同层之间滤波器尺寸依次增加 6 个像素

采用类似的方法来处理其他几组的模板序列。其方法是将滤波器尺寸增加量翻倍（6，12，24，38）。这样，可以得到第二组的滤波器尺寸，它们分别为 15，27，39，51。第三组的滤波器尺寸为 27，51，75，99。如果原始图像的尺寸仍然大于对应的滤波器尺寸，尺度空间的分析还可以进行第四组，其对应的模板尺寸分别为 51，99，147 和 195。下图显示了第一组至第三组的滤波器尺寸变化。

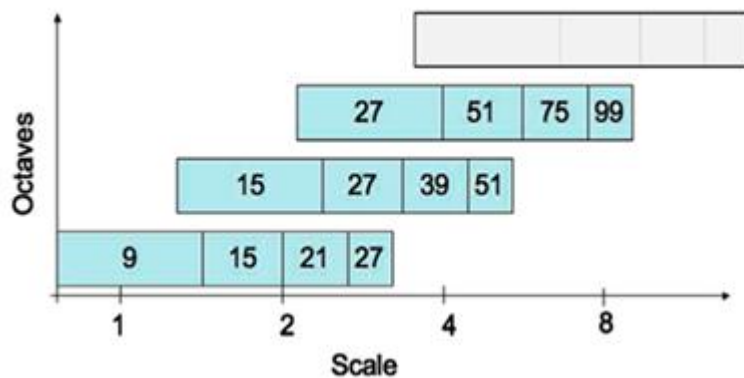


图 4-6: 第一组至第三组的滤波器尺寸变化，第一组增量是 6，第二组增量是 12，第三组增量是 24。

第 n 个 Octave 为第 $n-1$ 个 Octave 模板尺寸序列隔一个抽取得到，重复抽取是为了保证连续性。

滤波器响应长度 l ，滤波器尺寸 L ，组索引 o ，层索引 s ，尺度 σ 之间有严格的数学关系：

$$\begin{aligned} l &= 2^{\sigma+1}(s+1)+1 \\ L &= 3 \times l = 3 \times (2^{\sigma+1}(s+1)+1) \\ \sigma &= 1.2 \times L / 9 \end{aligned} \quad (4-5)$$

在通常尺度分析情况下，随着尺度的增大，被检测到的斑点数量迅速衰减。所以一般进行 3-4 组就可以了，与此同时，为了减少运算量，提高计算的速度，可以考虑在滤波时，将采样间隔设为 2。

如前所述，我们使用 的模板对图像进行滤波，其结果作为最初始的尺度空间层（此时，尺度值为 $s=1.2$ ，近似 的高斯微分），后续(layer)将通过逐步放大滤波模板尺寸，以及放大后的模板不断与图像进行滤波得到。由于采用盒子滤波和积分图像，滤波过程并不随着滤波模板尺寸的增加而使运算工作量增加。

4.3.3 定位特征点

首先初步定为特征点，我们用了 $3 \times 3 \times 3$ 邻域非最大值抑制。如图 4-7，将经过 hessian 矩阵处理过的每个像素点与其 3 维邻域的 26 个点进行大小比较，如果它是这 26 个点中的最大值或者最小值，则保留下来，当做初步的特征点。

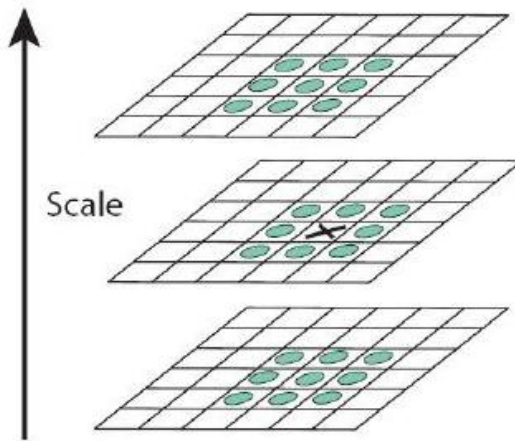


图 4-7：非最大值抑制，在 26 个像素点中找最值

然后，跟 sift 算法类似，采用 3 维线性插值法，Hessian 矩阵行列式的最大值

在尺度和图像空间被插值，得到亚像素级的特征点，同时也去掉那些值小于一定阈值的点，增加极值使检测到的特征点数量减少，最终只有几个特征最强点会被检测出来。

4.3.4 确定特征点主方向

为了保证旋转不变性，需要对每一个特征点分配一个主方向。在 SURF 中，不统计其梯度直方图，而是统计特征点领域内的 **Harr** 小波特征。即以特征点为中心，计算半径为 6σ (σ 为特征点所在的尺度值) 的邻域内，统计 60° 度扇形内所有点在 x (水平)和 y (垂直)方向的 **Haar** 小波响应总和(**Haar** 小波边长取 4σ)，并给这些响应值赋 2σ 的高斯权重系数，使得靠近特征点的响应贡献大，而远离特征点的响应贡献小。然后 60° 度范围内的响应相加以形成新的矢量，**Harr** 特征值反应了图像灰度变化的情况，那么这个主方向就是描述那些灰度变化特别剧烈的区域方向。遍历整个圆形区域，选择最长矢量的方向为该特征点的主方向。这样，通过特征点逐个进行计算，得到每一个特征点的主方向。该过程的示意图如下：

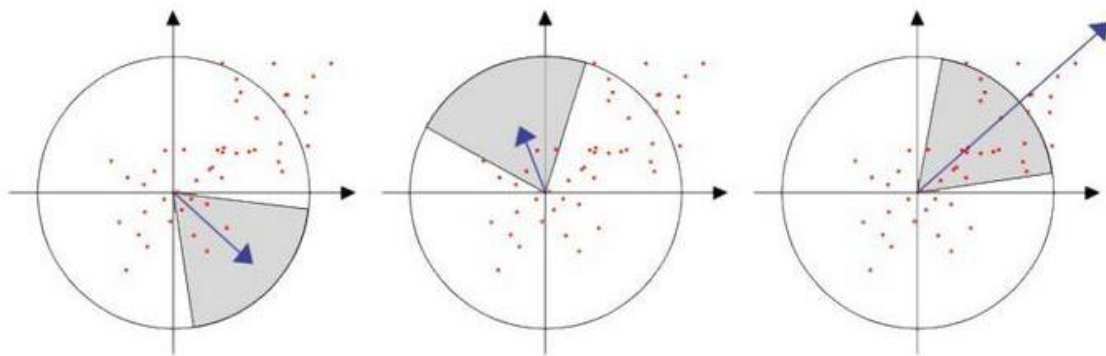


图 4-8: 求取主方向时扇形滑动窗口围绕特征点转动，统计 **Haar** 小波响应值，并计算方向角示意图

4.3.5 构造特征描述子

生成特征点描述子与确定特征点方向有些类似，它需要计算图像的 Haar 小波响应。不过，与主方向的确定不同的是，这次我们不是使用一个圆形区域，而是在一个矩形区域来计算 Haar 小波响应。在特征点周围取一个正方形框，框的边长为 20σ (σ 是所检测到该特征点所在的尺度)。该框带方向，方向当然就是前一步检测出来的主方向了。然后把该框分为 16 个子区域，每个子区域统计 25 个像素的水平方向和垂直方向的 Haar 小波特征，这里的水平和垂直方向都是相对主方向而言的。该 Haar 小波特征为水平方向值之和，水平方向绝对值之和，垂直方向之和，垂直方向绝对值之和。该过程的示意图如下所示：

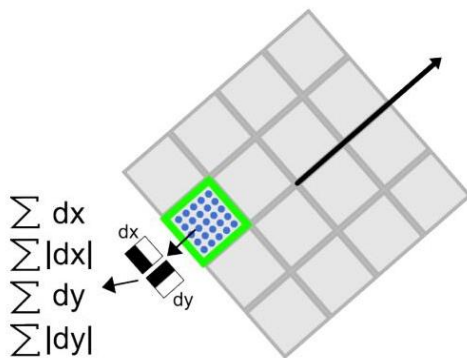


图 4-9：对 Haar 响应值统计形成特征向量

这样每个小区域就有 4 个值，所以每个特征点就是 $16 \times 4 = 64$ 维的向量，相比 sift 而言，少了一半，这在特征匹配过程中会大大加快匹配速度。

Haar 小波响应也可以利用积分图像。为利用积分图像进行 Haar 小波响应计算，并不直接旋转 Haar 小波模板求得其响应值，而是在积分图像上先使用水平和垂直的 Haar 模板求得响应值 dx 和 dy ，然后旋转 dx 和 dy 与主方向保持一致。

为了实现快速匹配，SURF 在特征矢量中增加了一个新的变量，即特征点的拉普拉斯响应正负号（即 Hessian 矩阵迹的正负号）。在特征点检测时，将 Hessian 矩阵的迹的正负号记录下来，作为特征矢量中的一个变量。这样做并不增加运算量，因为特征点检测已经对 Hessian 矩阵的迹进行了计算。在特征匹配时，这个变量可以有效地节省搜索的时间，因为只有两个具有相同正负号的特征点才有可能匹配，对于正负号不同的特征点就不进行相似性计算。

简单地说，我们可以根据特征点的响应值符号，将特征点分成两组，一组是具

有拉普拉斯正响应的特征点，一组是具有拉普拉斯负响应的特征点，匹配时，只有符号相同组中的特征点才能进行相互匹配。显然，这样可以节省特征点匹配的时间。

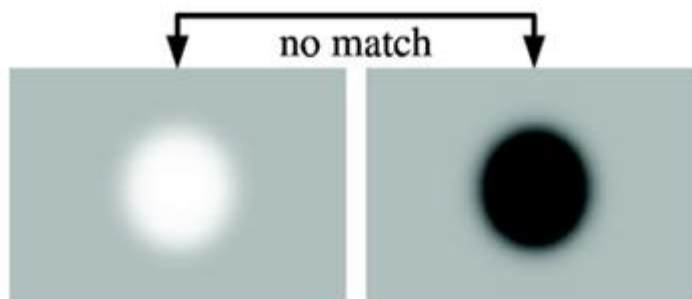


图 4-10 黑背景下的亮斑和白背景下的黑斑，因为它们的拉普拉斯响应正负号不同，不会对它们进行匹配

综上所述，Surf 在速度上比 sift 要快许多，这主要得益于它的积分图技术，已经 Hessian 矩阵的利用减少了降采样过程，另外它得到的特征向量维数也比较少，有利于更快的进行特征点匹配。SURF 采用 Hessian 矩阵获取图像局部最值还是十分稳定的，但是在求主方向阶段太过于依赖局部区域像素的梯度方向，有可能使得找到的主方向不准确，后面的特征向量提取以及匹配都严重依赖于主方向，即使不大偏差角度也可以造成后面特征匹配的放大误差，从而匹配不成功；另外图像金字塔的层取得不足够紧密也会使得尺度有误差，后面的特征向量提取同样依赖相应的尺度，在这个问题上我们只能采用折中解决方法：取适量的层然后进行插值。

4.4 本章小结

SURF (Speeded Up Robust Features)也是一种类似于 SIFT 的兴趣点检测及描述子算法。其通过 Hessian 矩阵的行列式来确定兴趣点位置，再根据兴趣点邻域点的 Haar 小波响应来确定描述子，SURF 的描述子基于 2D 离散小波变换响应 Harr 小波并且有效地利用了积分图。其描述子大小只有 64 维（也可以扩展到 128 维，效果更好），是一种非常优秀的兴趣点检测算法。SURF 算法的概念及步骤均建立在 SIFT 之上，但详细的流程略有不同。

第五章 非线性 KAZE 特征

5.1 KAZE 综述

虽然 sift,surf 等特征在图像特征提取方面已经取得比较好的成绩,但是这类特征都是基于一个通过高斯核进行的线性的尺度空间进行特征检测的,相同尺度下每个点的变换是一样的,由于高斯函数是低通滤波函数,会平滑图像边缘,以至图像损失掉许多细节信息。

针对这一问题, ECCV2012, European Conference on Computer Vision(欧洲计算机视觉国际会议)中出现了一种比 SIFT 更稳定的特征检测算法 KAZE[77]。作者提出了一种基于非线性尺度空间的特征点检测方法,该非线性尺度空间保证了图像边缘在尺度变化中信息损失量非常少,从而极大地保持了图像细节信息。

5.2 KAZE 原理

5.2.1 非线性尺度空间构建

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:

$$\frac{\partial L}{\partial t} = \text{div}(c(x, y, t) \cdot \nabla L) \quad (5-1)$$

其中 div 和 ∇ 分别表示散度和梯度。而函数 $c(x, y, t)$ 表示扩散的传导函数,正是该函数的引入使得扩散能够适应图像局部特征。在这里该传导函数主要由梯度幅值控制,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 年文章 Scale-space and edge detection using annisotropic diffusion[78]提出来的,为了尽量使平滑在区域内而不是区域间,减少在边缘处的扩散,Perona 提出了一个使得传导函数 c 依赖图像梯度幅值的函数:

$$c(x, y, t) = g(|\nabla L_{\sigma}(x, y, t)|) \quad (5-2)$$

其中 L_σ 是原始图像高斯平滑后的梯度。

由于非线性偏微分方程并没有解析解，只能通过一些数学方法近似求解该偏微分方程，本文主要使用 AOS 求解方案[79]。

文章采取的尺度空间的构建与 Sift 尺度空间的构建比较相似，都是通过指数步长的系列组合(O 个组, S 个层)来离散化尺度空间的，但与 sift 不同的是文章中的各组图像都是对原图像的操作而不是对图像进行降采样，而且在同一尺度下各个像素点的变化是非线性的。各个层之间的尺度关系如下：

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

其中 σ_0 是基本尺度， o 表示组序号， 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到时间单元概念，其转换如下：

$$t_i = \frac{1}{2} \sigma_i^2, i = \{0 \dots N\} \quad (5-4)$$

然后根据 AOS 方法进行非线性扩散滤波得到每一层。

5.2.2 特征点检测

特征点的检测与 Sift 特征点检测也非常相似，首先计算各点在本层，上下层 3×3 的立方体邻域内响应值是否为极值，再去掉重复点，再亚像素化精确位置，所不同的是各个像素点处响应值的计算。文章中响应值是采用多尺度下尺度归一化 Hessian 行列式计算的如下：

$$L_{hessian} = \sigma^2 (L_{xx} L_{yy} - L_{xy}^2) \quad (5-5)$$

其中 L_{xx}, L_{yy}, L_{xy} 分别表示 x 方向 2 阶导数， y 方向 2 阶导数， xy 方向混合导

数，其主要过程描述如下：

- (1) 计算各个像素点处响应值，响应值的计算如上
- (2) 查找在 3×3 空间邻域内的极大值
- (3) 删除重复特征点，两点在相邻层并且距离在尺度范围内认为是重复的。
- (4) 亚像素化，精确定位，亚像素化的过程与 sift 中的一样，通过泰勒表达式求导数，倒数为 0 的地方即为极值所在处，该位置即为偏移量：

$$\hat{X} = -\frac{\partial^2 D^{-1}}{\partial X} \frac{\partial D}{\partial X} \quad (5-6)$$

可以通过解方程：

$$\frac{\partial^2 D}{\partial X} \hat{X} = -\frac{\partial D}{\partial X} \quad (5-7)$$

得到 $\hat{X}(x, y, s)$ ，而 $\frac{\partial^2 D}{\partial X}$ 即为对应 $Hessian(x, y, s)$ 矩阵。

5.2.3 计算特征点主方向

文章所用的特征点主方向计算方法与 SURF 算法如出一辙。主要过程如下：

- (1) 以特征点为中心，半径为 $6s$ 区域内的点求 x, y 方向倒数 dx, dy ，然后以标准差为 $2.5s$ 的高斯函数进行加权
- (2) 建立一个 $\frac{\pi}{3}$ 的扇形区域，该扇形以 0.15 弧度为步长进行旋转，然后计算落角度在该扇形区域内点的 x, y 方向导数和
- (3) 获取所有扇形区域中使得 x, y 方向导数和的平方和的最大的扇形，主方向即为该扇形区域 x, y 方向导数和的反正切结果

5.2.4 建立描述子

文章中描述子的建立是采用 M-SURF 描述子，对于一个尺度为 σ_i 的特征，取该特征点邻域 $24\sigma_i \times 24\sigma_i$ 的矩形，再将该矩形分成 4×4 个 $9\sigma_i \times 9\sigma_i$ 的小矩形区域(每个相邻子区域有 $4\sigma_i$ 个像素是重叠的);计算邻域内所有点的 x, y 方向导数，并且以标准差为 $2.5\sigma_i$ 的高斯函数进行加权，然后统计每个子区域内的 x, y 方向导数 sum_dx, sum_dy 以及方向导数绝对值的和 sum_mdx, sum_mdy ，对于每个子区域再通过标准差为 $1.5\sigma_i$ 的高斯函数进行加权，最终每个子区域就能形成一个 4 维

特征, 4×4 个子区域能够形成 64 维特征, 如果想将 64 维特征扩展到 128 维特征, 可以对方向导数的正负值分别统计。

5.3 本章小结

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显著特征点。但高斯分解是牺牲了局部精度为代价的, 容易造成边界模糊和细节丢失。非线性的尺度分解有望解决这种问题, 但传统方法基于正向欧拉法 (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。

第六章 几种优化算法

6.1 KD 树+BBF 算法

无论是 SIFT 还是 SURF，都主要是针对特征点检测的算法，最后都生成一个特征描述子，分别为 128 维和 64 维。假如现在有两副图片都已经提取到特征点，现在要做的就是匹配上相似的特征点。在特征空间中，两个实例点的距离和反应出两个实例点之间的相似性程度。而常见的距离度量有欧氏距离，马氏距离、巴氏距离、曼哈顿距离、汉明距离、夹角余弦、切比雪夫距离、闵可夫斯基距离等。

以欧式距离为度量，又有三种匹配策略：固定阈值、最近邻、最近邻距离比率 (NNDR)。固定阈值：就是设定一个阈值，当距离大于阈值，判为不匹配，否则判为匹配。但是一个问题，是，阈值很难设定。随着移动到特征空间的不同部分时，阈值的有效范围会变化很大，即没有通用的阈值。最近邻：找最近的那个。最近邻距离比率：定义为最近邻距离和次近邻距离的比值。

6.1.1 K 近邻算法

相似性查询有两种基本方式：1.范围查询：即给点查询点和查询阈值，从数据集中找出所有与查询点距离小于阈值的点。2.K 近邻查询：给点查询点及正整数 K，从数据集中找到与查询点最近的 K 个数据，当 K=1 时，就是最近邻查询。

K 最近邻(k-Nearest Neighbor, KNN)分类算法，是一个理论上比较成熟的方法，也是最简单的机器学习算法之一。该方法的思路是：如果一个样本在特征空间中的 k 个最相似(即特征空间中最邻近)的样本中的大多数属于某一个类别，则该样本也属于这个类别。

算法的主要过程为：

- 1、给定一个训练集数据，每个训练集数据都是已经分好类的。
- 2、设定一个初始的测试数据 a，计算 a 到训练集所有数据的欧几里得距离，并排序。
- 3、选出训练集中离 a 距离最近的 K 个训练集数据。

4、比较 k 个训练集数据，选出里面出现最多的分类类型，此分类类型即为最终测试数据 a 的分类。

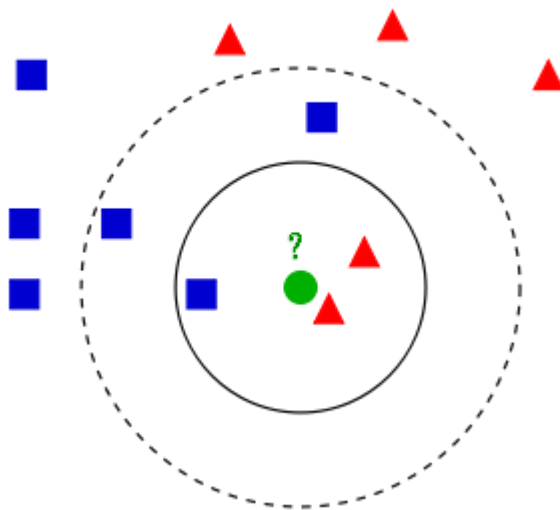


图 6-1: K 近邻法, 根据邻近的 K 个实例判断未知对象的类别

如上图所示, 有两类不同的样本数据, 分别用小正方形和小三角形表示, 而图正中间的那个圆所标示的数据则是待分类的数据。也就是说, 现在, 我们不知道中间那个圆的数据是从属于哪一类 (小正方形 or 小三角形), 下面, 我们就要解决这个问题: 给这个圆分类。

如果 $K=3$, 圆点的最近的 3 个邻居是 2 个小三角形和 1 个小正方形, 少数从属于多数, 基于统计的方法, 判定圆这个待分类点属于三角形一类。

如果 $K=5$, 圆点的最近的 5 个邻居是 2 个三角形和 3 个正方形, 还是少数从属于多数, 基于统计的方法, 判定这个待分类点属于正方形一类。

于此我们看到, 当无法判定当前待分类点是从属于已知分类中的哪一类时, 我们可以依据统计学的理论看它所处的位置特征, 衡量它周围邻居的权重, 而把它归为(或分配)到权重更大的那一类。这就是 K 近邻算法的核心思想。

关于 K 值的选择。统计学习方法中参数选择一般是要在偏差(Bias)与方差(Variance)之间取得一个平衡(Tradeoff)。偏差: 模型输出值与真实值之间的差异。偏差越高, 则数据越容易欠拟合(Underfitting), 未能充分利用数据中的有效信息。方差: 对数据微小改变的敏感程度。假如有一组同一类的样本, 并且这些样本的特征之间只有微小差异, 用训练好的模型进行预测并求得方差。理想情况下, 我们应该得到的方差为 0, 因为我们预料我们的模型能很好处理这些微小的变化; 但现实

中存在很多噪声(即存在不同类别的样本,其特征向量差异很小),即使是特征差异很小的同一类样本也可能达到不同类别的结果。而方差实际上就是衡量对噪声的敏感程度。方差越高,越容易过拟合(Overfitting),对噪声越敏感。

如李航博士的一书《统计学习方法》[81]上所说:如果选择较小的 K 值,就相当于用较小的领域中的训练实例进行预测,“学习”近似误差会减小,只有与输入实例较近或相似的训练实例才会对预测结果起作用,与此同时带来的问题是“学习”的估计误差会增大,方差过大,换句话说, K 值的减小就意味着整体模型变得复杂,容易发生过拟合;如果选择较大的 K 值,就相当于用较大领域中的训练实例进行预测,其优点是可以减少学习的估计误差,但缺点是学习的近似误差会增大,偏差过大。这时候,与输入实例较远(不相似的)训练实例也会对预测器作用,使预测发生错误,且 K 值的增大就意味着整体的模型变得简单。 $K=N$,则完全不足取,因为此时无论输入实例是什么,都只是简单的预测它属于在训练实例中最多的类,模型过于简单,忽略了训练实例中大量有用信息。

在实际应用中, K 值一般取一个比较小的数值,例如采用交叉验证法(简单来说,就是一部分样本做训练集,一部分做测试集)来选择最优的 K 值。

6.1.2 K 近邻算法的实现: KD 树

特征匹配算子可以分为两类:

1.穷举法:即将数据集中的点与查询点逐一计算距离,如果图 1 提取到 N_1 个特征点,图 2 中提取到 N_2 个特征点,用穷举法匹配,要做 $N_1 \times N_2$ 运算,这种方法显然效率低下。

2.建立数据索引:对数据进行分析,对搜索数据空间进行划分,根据划分的空间是否有混叠可以分为 Clipping 和 Overlapping 两种。前者划分空间没有重叠,其代表就是 k-d 树;后者划分空间相互有交叠,其代表为 R 树。

1975 年,来自斯坦福大学的 Jon Louis Bentley 在 ACM 杂志上发表的一篇文章[82]正式提出和阐述的如下图形式的把空间划分为多个部分的 K-d 树。

Kd-树是 K-dimension tree 的缩写,是对数据点在 k 维空间中划分的一种数据结构,主要应用于多维空间关键数据的搜索(如:范围搜索和最近邻搜索)。本质说,

Kd 树是一种平衡二叉树。

Kd 树的构建是一个递归过程，我们对左子空间和右子空间内的数据重复根节点的过程就可以得到一级子节点，同时将空间和数据集进一步细分，如此往复直到空间中只包含一个数据点。而节点的生成有两步，分别是确定 split 域和确定 Node-data 域。

确定 split 域：是对所有描述子数据（特征矢量），统计它们在每个维上的数据方差，以 SURF 为例，描述子为 64 维，可计算出 64 个方差，挑选出方差中最大值，对应的维就是 split 域的值。数据方差最大表明沿该坐标轴方向上数据点分散得比较开，这个方向进行数据分割可以获得最好的分辨率。

确定 Node-data 域，数据点集 Data-set 按其第 split 维的值排序，位于正中间的那个数据点被选为 Node-data。

经典的构造 K-d tree 的规则如下：

随着树的深度增加，循环的选取坐标轴，作为分割超平面的法向量。对于 3-d tree 来说，根节点选取 x 轴，根节点的孩子选取 y 轴，根节点的孙子选取 z 轴，根节点的曾孙子选取 x 轴，这样循环下去。

每次均为所有对应实例的中位数的实例作为切分点，切分点作为父节点，左右两侧为划分的作为左右两子树。

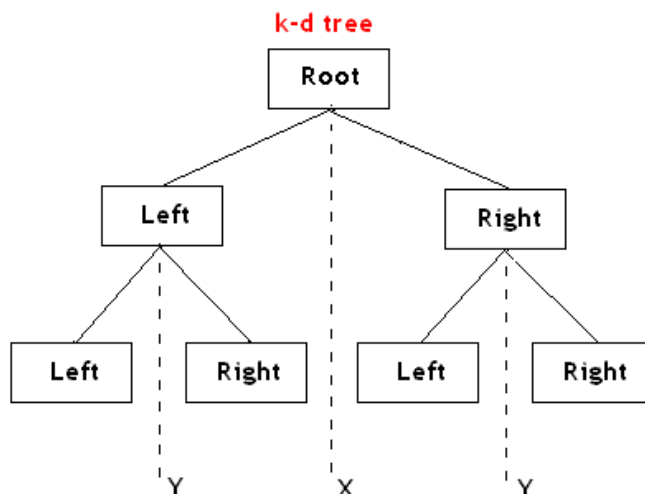


图 6-2

k-d 树算法可以分为两大部分，除了上部分有关 k-d 树本身这种数据结构建立

的算法，另一部分是在建立的 k-d 树上各种诸如插入，删除，查找(最邻近查找)等操作涉及的算法。

构建好 k-d 树之后就在其中进行数据查询。在 k-d 树中进行数据的查找也是特征匹配的重要环节，其目的是检索在 k-d 树中与查询点距离最近的数据点。

k-d 树近邻搜索算法如下：

1. 在 kd 树种找出包含目标点 x 的叶结点：从根结点出发，递归地向下搜索 kd 树。若目标点 x 当前维的坐标小于切分点的坐标，则移动到左子结点，否则移动到右子结点，直到子结点为叶结点为止。
2. 以此叶结点为“当前最近点”。
3. 递归的向上回溯，在每个结点进行以下操作：

(a) 如果该结点保存的实例点比当前最近点距离目标点更近，则更新“当前最近点”，也就是说以该实例点为“当前最近点”。

(b) 当前最近点一定存在于该结点一个子结点对应的区域，检查子结点的父结点的另一子结点对应的区域是否有更近的点。具体做法是，检查另一子结点对应的区域是否以目标点为球心，以目标点与“当前最近点”间的距离为半径的圆或超球体相交：如果相交，可能在另一个子结点对应的区域内存在距目标点更近的点，移动到另一个子结点，接着，继续递归地进行最近邻搜索；如果不相交，向上回溯。

4. 当回退到根结点时，搜索结束，最后的“当前最近点”即为 x 的最近邻点。

如果实例点是随机分布的，那么 kd 树搜索的平均计算复杂度是 $O(N \log N)$ ，这里的 N 是训练实例数。所以说，kd 树更适用于训练实例数远大于空间维数时的 k 近邻搜索，当空间维数接近训练实例数时，它的效率会迅速下降，下降到线性扫描的速度。

因为上述 k 最近邻搜索算法的第 4 个步骤中的所述：“回退到根结点时，搜索结束”，每个最近邻点的查询比较完成过程最终都要回退到根结点而结束，而导致了許多不必要回溯访问和比较到的结点，这些多余的损耗在高维度数据查找的时候，搜索效率将变得相当之低下。

从上述标准的 kd 树查询过程可以看出其搜索过程中的“回溯”是由“查询路

径”决定的，并没有考虑查询路径上一些数据点本身的一些性质。一个简单的改进思路就是将“查询路径”上的结点进行排序，如按各自分割超平面（也称 bin）与查询点的距离排序，也就是说，回溯检查总是从优先级最高（Best Bin）的树结点开始。

6.1.3 BBF 算法

根据前面 k-d 树的搜索过程我们可以知道，在搜索时首先沿着 kd 树找到叶子节点，然后依次回溯，而回溯的路程就是前面我们查找叶子节点时逆序，因此进行回溯时并没有利用这些点的信息。我们接下来介绍的算法就是利用这些信息，回溯时给各个需要回溯的结点以优先级，这样找到最近邻会更快。接下来详细介绍 bbf 算法的流程。

BBF（Best-Bin-First）查询算法，它是由发明 sift 算法的 David Lowe 在 1997 的一篇文章[83]中针对高维数据提出的一种近似算法，此算法能确保优先检索包含最近邻点可能性较高的空间，此外，BBF 机制还设置了一个运行超时限定。采用了 BBF 查询机制后，kd 树便可以有效的扩展到高维数据集上。

其实 BBF 算法的思想比较简单，通过对回溯可能需要的路过的结点加入队列，并按照查找点到该结点确定的超平面的距离进行排序，然后每次首先遍历的是优先级最高（即距离最短的结点），直到队列为空算法结束。同时 BBF 算法也设立了一个时间限制，如果算法运行时间超过该限制，不管是不是为空，一律停止运行，返回当前的最近邻点作为结果。

BBF 的算法流程如下：

输入：kd 树，查找点 x

输出：kd 树种距离查找点最近的点以及最近的距离

流程：（1）若 kd 树为空，则设定两者距离为无穷大，返回；如果 kd 树非空，则将 kd 树的根节点加入到优先级队列中；

（2）从优先级队列中出队当前优先级最大的结点，计算当前的该点到查找点的距离是否比最近邻距离小，如果是则更新最近邻点和最近邻距离。如果查找点在切分维坐标小于当前点的切分维坐标，则把他的右孩子加入到队列中，同时检索它的左孩子，否则就把他的左孩子加入到队列中，同时检索它的右孩子。这样一直

重复检索，并加入队列，直到检索到叶子节点。然后在从优先级队列中出队优先级最大的结点；

(3) 重复(1)和(2)中的操作，直到优先级队列为空，或者超出规定的时间，返回当前的最近邻结点和距离。

6.2 随机抽样一致性

对角点进行初始匹配后，所选定的角点并不能保证全部是正确的点，也可能有误差点，因此，还需要进一步对所选定的角点进行精确匹配。RANSAC(RANdom Sample And Consensus) 随机抽样一致性算法是由 Fischler 和 Bolles 提出的一种鲁棒性的参数估计方法[80]。它的基本思想是在进行参数估计时，不是不加区分地对待所有可用的输入数据，而是首先针对具体问题设计出一个目标函数，然后迭代地估计该函数的参数值，利用这些初始参数值把所有的数据分为所谓的“内点”(Inliers, 即满足估计参数的点)和“外点”(Outliers, 即不满足估计参数的点)，最后反过来用所有的“内点”重新计算和估计函数的参数。

RANSAC 的基本假设是：

- (1) 数据由“局内点”组成，例如：数据的分布可以用一些模型参数来解释；
- (2) “局外点”是不能适应该模型的数据；
- (3) 除此之外的数据属于噪声。

局外点产生的原因有：噪声的极值；错误的测量方法；对数据的错误假设。使用 RANSAC 估计方法，可以最大限度地减少噪声及外点的影响。简单的最小二乘法不能找到适应于局内点的直线，原因是最小二乘法尽量去适应包括局外点在内的所有点。相反，RANSAC 能得出一个仅仅用局内点计算出模型，并且概率还足够高。但是，RANSAC 并不能保证结果一定正确，为了保证算法有足够高的合理概率，我们必须小心的选择算法的参数。

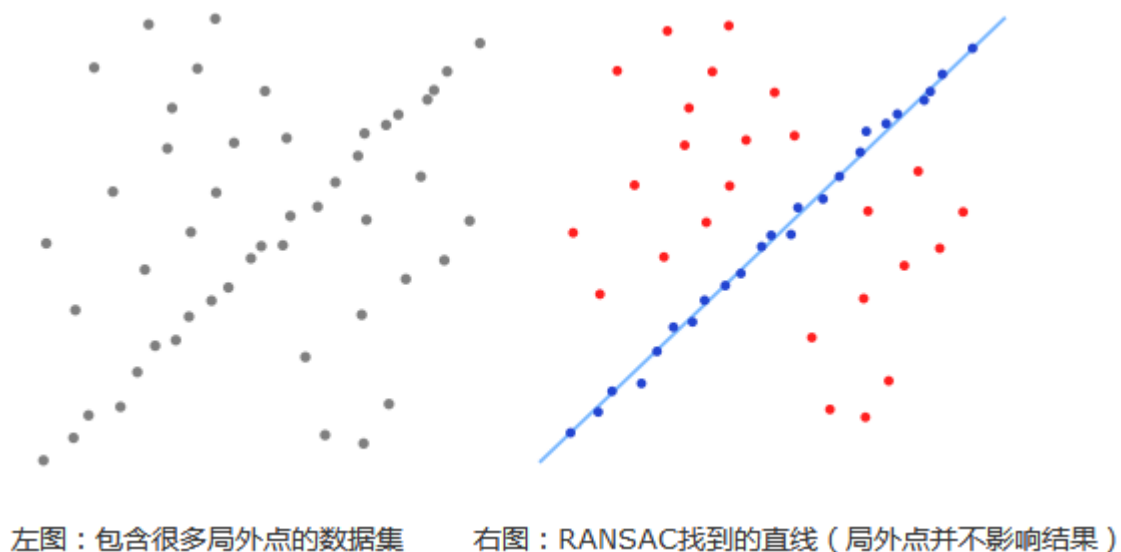


图 6-3:利用 RANSAC 寻找最优参数

6.2.1 RANSAC 步骤

RANSAC 用于选定最佳角点的主要步骤为：

(1)、将初始提取的 2 行 N 列的 N 个角点变成 3 行 N 列的 N 个角点，第三行为全 1；(2)、对初始提取的角点进行归一化；(3)、设置迭代最大次数；(4)、设置每次随机选取时最少的角点个数；(5)、每次随机选取由(4)指定个数的角点；(6)、判断由(5)中选取的角点，是否有部分共线，并设定最大循环次数；(7)、对选定的角点计算 2-D 单应性矩阵；(8)、选定一次内点并记录；(9)、对选定的内点进行判断；(10)、如满足条件(9)则作为选出的最好一组内点，否则，依次循环(5)~(9)；

RANSAC 算法由于在初始时是随机选取角点，因此存在不确定性，即使相同操作对两对同样的特征点进行 RANSAC 算法角点提取，每一次得出的结果也不一定就是相同的，但是最终的效果一般都是比较理想的。

SURF 算法与 ransac 算法相结合可以剔除出错误的匹配对。RANSAC 算法在 SURF 特征筛选中的主要流程是：

- (1) 从样本集中随机抽选一个 RANSAC 样本，即 4 个匹配点对。
- (2) 根据这 4 个匹配点对计算变换矩阵 M 。
- (3) 根据样本集，变换矩阵 M ，和误差度量函数计算满足当前变换矩阵的一致集 consensus，并返回一致集中元素个数。

(4) 根据当前一致集中元素个数判断是否最优(最大)一致集,若是则更新当前最优一致集。

(5) 更新当前错误概率 p , 若 p 大于允许的最小错误概率则重复(1)至(4)继续迭代,直到当前错误概率 p 小于最小错误概率。

6.2.2 RANSAC 优缺点

RANSAC 的优点是它能鲁棒的估计模型参数。例如,它能从包含大量局外点的数据集中估计出高精度的参数。RANSAC 的缺点是它计算参数的迭代次数没有上限;如果设置迭代次数的上限,得到的结果可能不是最优的结果,甚至可能得到错误的结果。RANSAC 只有一定的概率得到可信的模型,概率与迭代次数成正比。RANSAC 的另一个缺点是它要求设置跟问题相关的阈值。

RANSAC 只能从特定的数据集中估计出一个模型,如果存在两个(或多个)模型,RANSAC 不能找到别的模型。

经典 RANSAC 有以下三个主要的局限性:

(1) 效率:经典方法效率与子集大小、类内点比例以及数据集大小有关,因此在某些场景下效率较低。

(2) 精度:经典方法计算参数时选取最小子集是从效率的角度考虑,往往得到的是非最佳参数,在应用产参数 之前还需要再经过细化处理。

(3) 退化:经典方法的目标函数求取最大化的过程基于一个假设:“选取的最小子集中如果有类外点,那么在这种情况下估计的参数获得的目标函数(数据集中点的个数)往往较少”但这种情况在退化发生时有可能是不对的。

第七章 总结

图像配准在图像处理,尤其是海量处理中是非常重要的一个步骤,它把来源、质量、大小、角度参差不齐的多幅图像进行变换,从而与一幅事先选定的标准图像进行配准,进而利于后续的各种图像处理如图像融合、图像拼接、图像镶嵌、目标识别。事实上,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、航空侦察等领域都有广泛的应用。图像配准的质量好坏直接关系到最终的图像处理效果。这使得图像配准成为极具应用价值的技术。

而在图像配准中,图像特征点的检测又是关键和难点。图像的特征提取起到了降维分析的目的。因此一个算法是否成功往往由它使用和定义的特征决定。因此特征提取最重要的一个特性是“可重复性”:同一场景的不同图像所提取的特征应该是相同的。对于尺度不变特征,尺度不变特征变换(Scale Invariant Feature Transform, SIFT)是一种图像特征提取与描述算法。SIFT 算法由 David.G.Lowe 于 1999 年提出并在 2004 年进行了完善总结。SIFT 算法可以处理两幅图像之间发生平移、旋转、尺度变化、光照变化情况下的特征匹配问题,并能在一定程度上对视角变化、仿射变化也具备较为稳定的特征匹配能力。针对 SIFT 不能实时运算的问题, Bay 在 2006 年发表在 ECCV 大会上的 SURF 算法,全称是 Speeded-Up Robust Features。该算子在保持 SIFT 算子优良性能特点的基础上,同时解决了 SIFT 计算复杂度高、耗时长的缺点,对兴趣点提取及其特征向量描述方面进行了改进,且计算速度得到提高。而利用 RANSAC 算法,可以进一步剔除一些错误的匹配对,加快处理速度,改进处理效果。ECCV2012 中出现了一种比 SIFT 更稳定的特征检测算法 KAZE,一种基于非线性尺度空间的特征点检测方法,该非线性尺度空间保证了图像边缘在尺度变化中信息损失量非常少,从而极大地保持了图像细节信息。

国内在图像配准术领域的研究与国际上基本同步,国内的许多科研院所,如北京邮电大学、清华大学、中科院、西安电子科技大学、南京理工大学、天津大学、哈尔滨工业大学、湖南大学等,都在开展这方面的研究工作。国家的 863 计划、

973 计划和国家自然科学基金都也都对这一方向的研究给予了资助。

西安分院目前也有相关课题在进行图像配准方面的工作，周詮研究员于 2005 年开始着手关于图像配准方面的研究，目前获得已授权的与图像配准技术相关的专利申请有 30 余项。在周老师的指导下，其研究生冯新岗，李晓博，魏佳圆，朱厉洪，方海，刘娟妮、张茗茗等一直在丰富图像配准这方面的研究工作。

每种算法都有各自的优缺点，当处理样本不一样时，算法表现也不一样，并且特征点的提取其实不是越多越好，看到的好多论文只是片面地列举能表现自己改进的算法优点的数据，选择有利于自己的图片库。这是存在问题的。论文：A comparison of SIFT, PCA-SIFT and SURF 对三种方法给出了性能上的比较，源图片来源于 Graffiti dataset，对原图像进行尺度、旋转、模糊、亮度变化、仿射变换等变化后，再与原图像进行匹配，统计匹配的效果。效果以可重复出现性为评价指标。由此可见，SIFT 在尺度和旋转变换的情况下效果最好，SURF 在亮度变化下匹配效果最好，在模糊方面优于 SIFT，而尺度和旋转的变化不及 SIFT，旋转不变上比 SIFT 差很多。速度上看，SURF 是 SIFT 速度的 3 倍。但是该论文的评价标准没有量化，只是简单的用 common、good 评价。

所以，本课题的思路一方面是结合最新的研究成果，对已有算法进行优化和改进，如算法主动识别出完全不相关的图像，做出拒判，以期达到更加稳定，更加快速的效果；一方面是用多源、多样的图片库来验证算法，找到算法的不足，进一步调整，希望算法更加有普适性。而这些目标，不仅要求对计算机视觉有更深入的认识，深刻理解最基本的比如 SIFT 算法，因为后期众多算法是 SIFT 的改进；此外还要对人类视觉有更深入的理解，因为图像处理是利用计算机进行处理，但归根结底是服务于人眼的，这就依赖生物学的一些研究。此外，计算机视觉的研究表明，不同视角下物体往往可以通过一个透视矩（ 3×3 或 2×2 ）阵的变换而得到，但这种模型下的图像配准，以观察者的角度说，只配准了观察角度，以图像为观察对象的话，只对它进行了仿射变换，即只进行了旋转，缩放等变换，图像的颜色、分辨率、饱和度等依然没有和参考图像配准。还有，就目前看到的论文，更多的是针对两幅图像之间的配准，没有考虑到对多幅图像进行批量配准。批量配准不应该只是传统意义上的多次配准，应该引入机器学习的概念，在配准过程中不断迭代优化，找到更加快速、稳定的算法。

参考文献

- [1] 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.
- [2] Zitova B, Flusser J. Image registration methods :a survey[J].Image and Vision Computing. 2003,21(11),pp:977-1000.
- [3] (希)西奥多里德斯 等著, 李晶皎 等译.模式识别. 电子工业出版社. 2006-12
- [4] T. F. Cootes. C. J. Taylor. Statistical models of appearance for computer vision: World Wide Web Publication, 2004
- [5] Anuta P E. Spatial registration of multispectral and multitemporal digital imagery using fast Fourier transformation techniques. IEEE Trans. on Geosci. Electron. 1970.8(4),pp:355-368.
- [6] Bamea D I. Silverman H F. A class of algorithms for fast digital image registration. IEEE Trans.on Computers.1972,21(2),pp:179-186.
- [7] Pratt W K(高荣坤等译).数字图像处理学.北京:科学出版社.1984.9.
- [8] Svedlow M, Mac Gillen C D, Anuta P E. Image registration:similarly measure and preprocessing method comparisons. IEEE Trans. on Aerospace and Electronic Systems.1978,AES-14(1),pp:141-150.
- [9] Rignot E, Kwok R, Curlander Jetal. Automate multisensor registration: requirments and techniques. Photogrammetric Eng. Remote Sensing.1991,57(8),pp:1029-1038.
- [10]Flussr Jan. An adaptive method for image registration. Pattern Recognition. 1992. 25(1),pp:45-54.
- [11]P.A. Viola, Alignment by maximization of mutual information, Ph.D. thesis, Massachusetts Institute of Technology,Boston,MA,USA,1995
- [12]A.Collignon,F.Maes,D.Delaere,D.Vandermeulen,P.Suetens,andG.Marchal,Automed 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
- [13]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

- [14]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
- [15]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
- [16]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
- [17]S.Mallat. A Theory for Multiresolution Signal Decomposition: the Wavelet Representation.IEEE Transactions on Pattern Analysis and Machine Intelligence,1989,11(7):674-693
- [18]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
- [19]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
- [20]Viola P.; Wells WM.. Alignment by Maximization of Mutual Information[J]. International Journal of Computer Vision , 1997 , 24 (2):137-154 .
- [21]Reddy B S, Chatterji B N. An FFT-based technique for translation, rotation, and scale-invariant image registration[J]. Image Processing, IEEE Transactions on, 1996, 5(8): 1266-1271.
- [22]Chen Q, Defrise M, Deconinck F. Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1994, 16(12): 1156-1168.
- [23]Hough P V C. Method and means for recognizing complex patterns: U.S. Patent 3,069,654[P]. 1962-12-18.
- [24]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.
- [25]H. Gonalves; J. A. Gonalves; 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 .
- [26]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 .

- [27]Hu M K. Visual pattern recognition by moment invariants[J]. Information Theory, IRE Transactions on, 1962, 8(2): 179-187.
- [28]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.
- [29]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.
- [30]Flusser J. Moment invariants in image analysis[C]//proceedings of world academy of science, engineering and technology. 2006, 11(2): 196-201.
- [31]Flusser J, Suk T. Rotation moment invariants for recognition of symmetric objects[J]. Image Processing, IEEE Transactions on, 2006, 15(12): 3784-3790.
- [32]Matas J, Chum O, Urban M, et al. Robust wide-baseline stereo from maximally stable extremal regions[J]. Image and vision computing, 2004, 22(10): 761-767.
- [33]Murphy-Chutorian E, Trivedi M M. N-tree Disjoint-Set Forests for Maximally Stable Extremal Regions[C]//BMVC. 2006: 739-748.
- [34]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.
- [35]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.
- [36]Kadir T, Brady M. Saliency, scale and image description[J]. International Journal of Computer Vision, 2001, 45(2): 83-105.
- [37]Kadir T, Zisserman A, Brady M. An affine invariant salient region detector[M]//Computer Vision-ECCV 2004. Springer Berlin Heidelberg, 2004: 228-241.
- [38]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.
- [39]Mikolajczyk K, Schmid C. An affine invariant interest point detector[M]//Computer

- Vision—ECCV 2002. Springer Berlin Heidelberg, 2002: 128-142.
- [40]Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors[J]. International journal of computer vision, 2004, 60(1): 63-86.
- [41]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.
- [42]HansP M. Towards automatic visual obstacle avoidance[C]//Proceedings of the 5th international joint conference on Artificial intelligence. 1977: 584-584.
- [43]Harris C, Stephens M. A combined corner and edge detector[C]//Alvey vision conference. 1988, 15: 50.
- [44]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.
- [45]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.
- [46]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.
- [47]Lindeberg T. Edge detection and ridge detection with automatic scale selection[J]. International Journal of Computer Vision, 1998, 30(2): 117-156.
- [48]Lindeberg T. Feature detection with automatic scale selection[J]. International journal of computer vision, 1998, 30(2): 79-116.
- [49]Mikolajczyk K, Schmid C. An affine invariant interest point detector[M]//Computer Vision—ECCV 2002. Springer Berlin Heidelberg, 2002: 128-142.
- [50]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.
- [51]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.
- [52]Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110.
- [53]Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features[M]//Computer Vision–ECCV 2006. Springer Berlin Heidelberg, 2006: 404-417.
- [54]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.
- [55]Mikolajczyk K, Schmid C. A performance evaluation of local descriptors[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2005, 27(10): 1615-1630.
- [56]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.
- [57]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
- [58]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
- [59]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
- [60]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
- [61]RostenE,Drummond T.Machine learning for high-speed cornerd etection [M] // Computer Vision–ECCV 2006. Springer Berlin Heidelberg, 2006: 430-443.
- [62]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.
- [63] 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.
- [64] Taylor S, Drummond T. Binary histogrammed intensity patches for efficient and robust matching[J]. International journal of computer vision, 2011, 94(2): 241-265.
- [65] Lisa Gottesfeld Brown . A Survey of Image Registration Techniques . ACM Computing Surveys[J] , 1992 , 24(4):325~376 .
- [66] Alcantarilla P F, Bartoli A, Davison A J. KAZE features[M]//Computer Vision—ECCV 2012. Springer Berlin Heidelberg, 2012: 214-227.
- [67] 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
- [68] 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.
- [69] 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
- [70] 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
- [71] 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
- [72] Brown L. G.. A survey of image registration techniques. ACM Computer, Surveys, Dec. 1992, Vol. 24, No. 4, p325-376
- [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]opencv2.3 改用的 Rob Hess 的源码 <http://robwhess.github.io/opensift/>, 牛津大学开发的 VLFeat open source library <http://www.vlfeat.org/>
- [76]H.Bay, T. Tuytelaars, L. V. Gool, SURF:Speed Up Robust Features[J],ECCV,2006
- [77]Pablo F.Alcantarilla, Adrien Bartoli, Andrew J.Davision .KAZE Features .[J] ECCV.2012
- [78]P Perona , J Malik. Scale-Space and Edge Detection Using Anisotropic Diffusion. IEEE Computer Society , 1990 , 12 (7) :629-639
- [79]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
- [80]M.A. Fischler and R.C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,” Comm. ACM, vol. 24, no. 6, pp. 381- 395,1981.
- [81]李航.统计学习方法.清华大学出版社.2012.3.1
- [82]Jon Louis Bentley. Multidimensional binary search trees used for associative searching. Communications of the ACM. 1975.:509-517
- [83]Jeffrey S. Beis, David G. Lowe. Shape Indexing Using Approximate Nearest-Neighbour Search in High-Dimensional Spaces. IEEE Computer Society Conference on Computer Vision , 1997 :1000-1006
- [84]Deriche R, Giraudon G. A Computational Approach for Corner and Vertex Detection. International Journal of Computer Vision, 1993, 10(2): 101-124.
- [85]Schmid C, Mohr R and Baucklage C. Comparing and Evaluating Interest Points[J]. International Journal of Computer Vision, 2000, 37(2): 151-172.
- [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.