

分类号 TN911  
U D C           

密 级             
编 号 10486

# 武汉大学

## 硕士 学位 论 文

### 图像特征匹配的鲁棒算法研究 及其应用

研 究 生 姓 名: 范翱翔

学 号: 2018202120098

指导教师姓名、职称: 马佳义 教授

专 业 名 称: 信息与通信工程

研 究 方 向: 计算机视觉与模式识别

二〇二一年五月

# A Study of Robust Algorithms in Image Matching and Its Applications

Candidate: FAN AOXIANG

Student Number: 2018202120098

Supervisor: PROF. MA JIAYI

Major: Information and Communication Engineering

Speciality: Computer Vision and Pattern Recognition



Electronic Information School

WUHAN UNIVERSITY

May, 2021

## 论 文 原 创 性 声 明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的研究成果。除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

学位论文作者 (签名)：

年       月       日

## 摘要

图像特征匹配是计算机视觉领域中一个重要且基本的问题，是许多视觉任务，例如图像拼接、目标追踪、视觉定位以及三维重建等的先决条件。从概念上来说，该问题可以定义为在两幅图像中识别出表征相同物理意义的点并建立精确的匹配对应关系。为了解决这一问题，一般而言，在图像特征匹配领域中所使用的标准做法是从图像中提取显著且稳定的特征点，生成对应的描述子来表征局部的场景信息，最后根据描述子之间的欧式距离来建立最终的匹配关系。然而，在实际的图像数据中，由于场景的多样性、复杂性以及观测噪声的影响，由这种方式所建立的匹配将不可避免的受到错误匹配（即离群点）的干扰。因此，针对这一现实问题，研究对于离群点鲁棒且高效的特征匹配算法对于实现智能视觉具有着极为重要的理论研究意义与实际应用价值。

在数据同时受到噪声以及离群点影响的情况下，寻找到潜在的内点在数学上是一个非常复杂的问题，在某些特定的条件下可以证明具有 NP-hard 的特性。本文以图像特征匹配为应用背景，从不同角度出发对该问题进行分析与建模，提出了一系列具有不同特性的高效率算法。在包含大量多样性场景的真实世界数据集中所进行的定性与定量实验，进一步证明了本文所提出算法相较于其他处于领先水平的算法在鲁棒性以及低时间消耗上的性能优势。从更为广义上的层面来说，本文所提出的算法不仅仅局限于图像特征匹配这个单一任务中，也可以为其他领域中对离群点鲁棒的算法设计提供参考价值。

首先，从统计学习的角度来说，含有离群点的数据可以利用一种特殊的概率分布，即高斯分布与均匀分布的混合模型来描述。与此同时，为区分离群点，一种常用的思路是对于内点做出平滑性的假设，将其理解为从一种特殊的平滑函数中采样的结果。此时利用经典的正则化理论，可在再生核希尔伯特空间（RKHS）中寻找复杂度最低的潜在函数来为平滑函数建模。然而，正则化理论的一个严重缺陷是其较高的计算复杂度，假设输入  $N$  个数据点，那么正则化理论的计算复杂度将是  $O(N^3)$ 。基于这一点，本文提出了基于傅里叶基底的表示方法来对假设中的平滑函数进行建模。这种建模方式的优势在于，其对于平滑性函数的表示是紧凑的，即在大多数情况下，在求解时仅

仅需要使用很少的一部分参数量便可以实现对平滑函数高效的近似。本文将这一理论作为正则化理论的替代，并将其嵌入于能够处理离群点的概率模型框架之中，提出了一种全新的算法，其理论上的计算复杂度降低为近似的  $O(N)$ 。本文在实验中验证了所提出算法的有效性，在不损失精度的情况下计算效率得到了大幅度提高，同时相较于其他处于领先地位的算法具有优势。

其次，本文研究了机器学习中经典的基于图模型的理论在图像特征匹配中的应用。图模型是一种应用非常广泛的学习算法，其对于问题的核心假设建立在数据点之间存在的相关性之上，具有很高的灵活性。相较于广义的图模型，本文所研究的概率图模型提供了一种更为严谨的数学描述问题的方式。为了尽可能的实现计算效率的提升，本文使用最初等的二阶图模型来作为问题建模的基础，即只考虑数据中两两之间的相关性。针对图像特征匹配的任务，本文提出了一种广泛成立的基于运动一致性的先验来建模这种相关性。同时，在数学上，本文将概率图模型的求解抽象为一个整数二次规划 (Integer Quadratic Programming, IQP) 问题，并提出了一种基于 Frank-Wolfe 方法的快速近似求解的算法。在实验中表明，本文所提出的基于图模型的算法在匹配数据中存在多种运动模式的场景下，仍然能发挥稳定的作用来去除离群点，同时计算效率相较于其他算法也处于领先地位。

最后，本文结合图像特征匹配中几何模型估计的需求提出了一种鲁棒的模型拟合的算法。具体而言，图像特征匹配一个最为重要的应用方向是三维视觉，例如三维重建或者同步定位与建图。在三维视觉中，根据两视图几何的经典理论，图像特征匹配的结果可以认为编码着重要的几何模型或者几何信息，例如基础矩阵或者单应矩阵，其对于后续还原三维几何结构的任务具有着至关重要的影响。如何在存在离群点的情况下对这些几何模型进行鲁棒的估计是三维视觉中一个极其重要的课题。本文从几何模型估计的数学理论出发，创新性地将的  $l_1$  损失引入进来，使得目标函数具有天然的鲁棒性，因此大幅度提升对于离群点的抵抗能力。同时，在求解方面，本文将该问题与鲁棒子空间估计的理论联系起来，提出了多种基于次梯度下降的快速求解算法。在包含多种具有代表性和挑战性的现实图像数据集中，大量的定性以及定量的实验结果证明了所提出算法在鲁棒模型估计任务中的有效性。

**关键词:** 图像特征匹配，离群点，鲁棒算法，正则化理论，概率图模型，几何模型估计，三维视觉

## ABSTRACT

Image matching is an important and fundamental problem in computer vision, which is also the prerequisite of many visual tasks such as image stitching, target tracking, visual localization and 3D scene reconstruction. Conceptually, the problem can be defined as the task to identify the points that represent the same physical entity in two images and establish accurate correspondences. To resolve the problem, in general, the common practice in the field image matching is to firstly detect salient and stable feature points, secondly generate descriptors accordingly to represent the local scene information, and finally utilize the Euclidean distances between descriptors to determine the correspondences. However, in real-world image data, the established correspondences will inevitably be affected by false matches, i.e. outliers, due to the complexity and diversity of the imaging scenes and measurement noises. Hence, focusing on the problem, the investigation of outlier-robust and efficient methods for image matching is of both theoretical significance and great practical values for intelligent vision.

Given that the data is contaminated by both noises and outliers, it is mathematically a very complicated problem to identify the inliers, and can be proved to be NP-hard under certain assumptions. In this thesis, we base our application on the problem of image matching, analyze and model the problem from several distinct perspectives, and propose a series of efficient algorithms with different properties. Qualitative and quantitative experiments are conducted extensively on real-world image data of great diversities, which demonstrate that the proposed algorithms outperform the state-of-the-art methods in terms of robustness and efficiency. In a more general view, our proposed methods are not designed exclusively for the task of image matching, it can also provide insights for outlier-robust algorithms in many other fields.

First, in the view of statistical learning, the distribution of data contaminated by outliers can be described by a mixture model of a Gaussian distribution and a uniform

distribution. In this probabilistic model, approximate but efficient solutions can be obtained in a classic framework, if we can make reasonable assumptions on the inliers. Considering the prior condition that the inliers should comply with, the most common and intuitive choice is a smoothness assumption, i.e. the inliers are regarded as samples from a smooth function. Based on this assumption, the regularization theory is ubiquitously used in the literature of image matching, i.e. searching for the potential function with the lowest complexity in a Reproducing Kernel Hilbert Space (RKHS). However, a major drawback of regularization theory is its high computational complexity. Assume that we have  $N$  data points as input, the computational complexity is  $O(N^3)$ . To address this issue, we propose to use Fourier bases representation to model the smooth function. The advantage of this modeling is that the representation is compact for smooth functions. As a result, in most cases we only require a small fixed number of parameters to approximate a smooth function. In this thesis we propose a novel method with the compact Fourier representation to play the role of regularization theory, and incorporate it into a probabilistic framework to handle outliers. The new algorithm has a computational complexity reduced to  $O(N)$ . We validate the effectiveness of the proposed method through extensive experiments, and demonstrate that the efficiency has been greatly improved with negligible loss in accuracy.

Second, we investigate the use of graphical model in machine learning theory in the image matching field. Graphical model is a learning algorithm with very broad applications, its core assumption is based on the internal correlations between data points, thus it is very flexible in practice to deal with different forms of data. Compared to general graphical model, probabilistic graphical model (PGM) provides a more rigorous mathematical interpretation, which is the framework we adopt in this thesis. For efficiency concerns, we only consider second-order graph, i.e. we only consider pairwise correlations in the proposed method. Theoretically, the modeling of correlations naturally has distinct forms for different problems. In image matching, we develop a general local motion consistency prior to model the correlations. Moreover, the derivation of the solution of the PGM is converted into an integer quadratic programming (IQP)

problem, an efficient algorithm based on Frank-Wolfe method is proposed for computation. In experiments, we demonstrate that the proposed method based on PGM has the advantage of preserving inliers of different motion patterns while rejecting the outliers, and the time cost is also preferred.

Third, considering the need of geometric model fitting in image matching, we propose a robust geometric model estimation method. Specifically, the most important direction of application for image matching is 3D computer vision, such as Structure-from-Motion (SfM) and Simultaneous Localization and Mapping (SLAM). In 3D vision, based on the classic theory of two-view geometry, the image matching results encode important geometric model or geometric information, such as fundamental matrix and homography, which plays a vital role in subsequent tasks to recover the 3D structure of the imaging scene. The problem of robust geometric model estimation in the presence of gross outliers is a very important topic in 3D vision. Based on the mathematical theory of geometric model estimation, we introduce  $l_1$  loss in the formulation to make the objective inherently robust to outliers. Moreover, we relate the solution of such a problem to the theory of robust subspace recovery, and propose several efficient algorithms based on sub-gradient descent. In real-world datasets containing large numbers of representative and challenging images, extensive qualitative and quantitative experiments demonstrate the effectiveness of the proposed method in the task of robust geometric model estimation.

**Key words:** Image matching, outliers, robust algorithms, regularization theory, probabilistic graphical model, geometric model estimation, 3D vision

# 目 录

摘要	I
<b>ABSTRACT</b>	<b>III</b>
<b>1 绪论</b>	<b>1</b>
1.1 问题定义与分类 . . . . .	1
1.2 研究背景及意义 . . . . .	2
1.3 图像特征匹配研究现状 . . . . .	4
1.3.1 图像特征点的检测 . . . . .	5
1.3.2 直接匹配策略 . . . . .	6
1.3.3 简化匹配策略 . . . . .	8
1.3.4 深度学习方法 . . . . .	11
1.3.5 国内相关研究 . . . . .	12
1.4 图像特征匹配发展趋势 . . . . .	13
1.4.1 图像特征匹配与应用的结合 . . . . .	13
1.4.2 深度学习技术的引入 . . . . .	14
1.4.3 多图像协同匹配 . . . . .	14
1.5 论文研究内容与组织结构 . . . . .	15
<b>2 基于紧凑表示的鲁棒平滑函数恢复算法研究</b>	<b>17</b>
2.1 引言 . . . . .	17
2.2 基于稀疏数据的多元平滑函数回归方法 . . . . .	18
2.2.1 正则化理论 . . . . .	18
2.2.2 基于紧凑表示的平滑函数学习 . . . . .	19
2.3 图像特征匹配中的平滑函数恢复 . . . . .	22
2.3.1 贝叶斯建模 . . . . .	22
2.3.2 基于期望最大化算法的求解 . . . . .	23

2.3.3	复杂度分析以及实现细节	25
2.4	实验结果	26
2.4.1	合成数据中的多元回归	27
2.4.2	鲁棒图像匹配	29
2.4.3	非刚性点集配准	31
2.5	本章小结	33
<b>3</b>	<b>基于概率图模型的误匹配去除算法研究</b>	<b>34</b>
3.1	引言	34
3.2	基于概率图模型的误匹配去除算法	35
3.2.1	问题建模	35
3.2.2	运动一致性度量	36
3.2.2.1	向量场表示	36
3.2.2.2	基于差商的度量	36
3.2.3	局部运动一致性先验	38
3.2.3.1	局部交互结构	38
3.2.3.2	成对交互模型	39
3.2.4	求解算法	40
3.2.5	计算复杂度分析	43
3.2.6	实现细节	44
3.3	实验结果	44
3.3.1	图像特征匹配实验结果	45
3.3.1.1	数据集以及度量指标	45
3.3.1.2	定性研究	46
3.3.1.3	数据集定量研究	46
3.3.1.4	参数分析	48
3.3.1.5	算法分析	48
3.3.2	基础矩阵估计实验结果	49
3.3.3	闭环检测实验结果	52
3.4	本章小结	53

<b>4 基于子空间恢复的鲁棒几何模型估计方法研究</b>	<b>55</b>
4.1 引言 . . . . .	55
4.2 基于子空间恢复的鲁棒几何模型估计方法 . . . . .	56
4.2.1 经典的 DLT 算法 . . . . .	56
4.2.2 鲁棒推广 . . . . .	57
4.2.3 对于线性结构的进一步探索 . . . . .	58
4.2.3.1 仿射相机中的线性结构 . . . . .	58
4.2.3.2 线性结构的推广 . . . . .	61
4.2.4 实现细节 . . . . .	63
4.3 实验结果 . . . . .	63
4.3.1 线性嵌入形式的定性分析 . . . . .	64
4.3.2 基础矩阵与单应变换的估计 . . . . .	65
4.3.3 离群点比率的敏感度 . . . . .	67
4.4 本章小结 . . . . .	68
<b>5 总结与展望</b>	<b>69</b>
5.1 全文总结 . . . . .	69
5.2 未来展望 . . . . .	71
<b>参考文献</b>	<b>73</b>
<b>攻读硕士学位期间发表的学术论文</b>	<b>88</b>
<b>致谢</b>	<b>89</b>

# 1 绪论

图像特征匹配是计算机视觉领域一个经典并且尤为关键的任务，而其中的离群点问题是制约该任务的性能表现的关键因素之一。本文旨在对于图像特征匹配以及其中的离群点问题进行深入的研究，并提出相应的高效鲁棒算法来满足不同应用场景的需求。首先本文将从图像特征匹配的定义出发，对其进行简单的分类，并结合离群点的定义解释这种常见的现象所带来的复杂问题。随后本文将分别从图像特征匹配与离群点的角度，结合其跨领域的应用需求及研究难点，来揭示本文所研究内容在计算机视觉中的重要地位与所提出技术的实际应用价值和理论研究意义。同时，本文也将对图像特征匹配领域的研究现状及发展趋势做出总结，并引出本文的主要的研究动机以及主要内容的划分。

## 1.1 问题定义与分类

图像特征匹配 (Image Matching) <sup>[1,2]</sup> 作为计算机视觉领域一个经久不衰的研究课题，其一般的定义为寻找并提取两幅图像中具有相同物理意义的特征，并建立对应关系的过程。从所研究的图像数据类型的层面来说，具有相同物理意义的目标具有多种不同的表现形式，最常见的一种形式把两幅图像理解为真实世界的三维场景在两个不同角度所形成的二维可见光投影 <sup>[3]</sup>。除此以外，由于特殊的成像原理所形成的多模态图像，乃至不局限于特定目标而仅仅具有相似语义信息的图像，都可以作为图像特征匹配任务的目标 <sup>[2]</sup>。

从特征匹配的角度来说，可以将图像特征匹配分为稠密匹配与稀疏匹配两种形式。稠密的图像特征匹配通常不包含兴趣点的选取，而着眼于得出图像中所有像素之间的匹配对应关系。早期的图像配准算法可以理解为稠密图像匹配的一种形式，该种方法的总体思路一般是通过最小化图像整体的一种分布差异的度量，例如相关系数或者互信息等 <sup>[4-7]</sup>，来搜寻逐像素之间的对应关系。在医学图像配准 <sup>[8]</sup> 中，由于图像之间的变换具有高度非刚性的特性，因此也常使用图像中所有的像素来定义总体的目标函数，

最终通过复杂的变换模型来实现像素与像素之间的匹配对齐。在三维视觉的研究中，也存在着一些图像稠密特征匹配的分支，但常常只能应用于比较受限的场景。例如，光流法<sup>[9]</sup> 所代表的一类方法仅适用于时间连续性较强的两帧图像，而在立体匹配<sup>[10]</sup> 中，所研究的对象一般为极窄基线的两幅图像。

由于稠密匹配常常意味着巨大的计算消耗，或者只能应用于受限的场景之中，稀疏特征匹配<sup>[11]</sup> 成为了图像特征匹配的热点。稀疏特征匹配通常包含数个步骤，即图像特征点的检测与描述子的提取，以及基于描述子的匹配关系的建立<sup>[12, 13]</sup>。在稀疏特征匹配中，得益于显著且稳定的特征的检测与提取，匹配的任务转化为离散的有限个坐标点，避免了针对所有像素的大规模计算。并且，由于特征的局部特性，使得这一路线对于场景变化或者噪声影响具有更强的耐受能力。因此，稀疏特征匹配成为实际应用中一大类视觉任务的主流做法，如图像拼接<sup>[14]</sup>、视觉定位<sup>[15]</sup>、三维重建<sup>[16]</sup> 以及同步定位与建图<sup>[17]</sup> 等等。本文的研究便立足于图像的稀疏特征匹配，若不另加说明，在后文中所使用的图像特征匹配即指代稀疏匹配。

然而，在实际的图像特征匹配的应用中，由于所提取特征的局部特性、特征提取算法的固有的缺陷、以及各种噪声的影响，所建立的匹配关系常常存在不同程度的错误匹配（离群点）。离群点的存在对于后续的任务具有极其严重的负面影响，例如在三维重建的任务中，离群点的存在将使得估计出的相机姿态失去准确性，导致整个任务的失败。这是由于离群点带来的是纯粹的错误信息，将直接误导后续的任务操作。因此，针对离群点这一问题，设计鲁棒的算法具有重大的理论和实际应用价值。更进一步来说，本文所研究的离群点问题虽然主要以图像特征匹配作为主要的应用方向，但其中解决问题的思路及建模方式是一般性的，因此对广义上的离群点问题也有一定的借鉴意义。例如，在三维视觉领域中，不同于 2D-2D 的图像特征匹配，基于特征的点集配准任务<sup>[18]</sup> 以及经典的 Triangulation 问题<sup>[19]</sup> 常常也面临着离群点问题的困扰，而其面临的匹配数据类型转化为 3D-3D 或者 3D-2D。

## 1.2 研究背景及意义

近年来，随着计算机技术的快速发展，人类社会正快速向信息化以及智能化的时代迈进，同时也为视觉领域迎来了新一波的创新浪潮。对于视觉领域的研究来说，其最终的目的是赋予机器以人类的视觉感知能力，以此来完善各种智能化技术，自动化

理论在现实中的应用。而在这一浪潮之中，图像数据作为获取最简单，信息最丰富的一种载体，吸引了全世界研究者最为广泛的关注。而在图像数据相关的多种任务中，图像特征匹配作为实现多种视觉任务的基础，可以在不同图像之间建立联系，是低层次视觉连接高层次视觉的重要桥梁，也是实现信息的识别与整合<sup>[20-22]</sup>，以及从二维图像恢复三维结构的关键途径<sup>[23, 24]</sup>。

具体来说，在图像之间建立匹配关系是诸多高层次视觉任务的先决条件。根据数据获取的条件以及成像条件的差异，图像的特征匹配可以分为不同时间、不同视角以及不同传感器这几种情况，分别对应着不同的应用场景。1) 不同视角的图像匹配：匹配的目标图像是同一场景或目标物体在不同视角所形成的图像，这种类型是图像特征匹配应用最广泛的方向，包括图像拼接，视觉定位以及三维重建<sup>[14-16]</sup> 等等，是三维视觉的核心问题之一。2) 不同成像时间的特征匹配：匹配的目标图像是同一场景或目标物体在不同时间所形成的图像，可以应用于变化检测<sup>[25]</sup>、视频去抖动<sup>[26]</sup> 以及目标跟踪<sup>[27]</sup> 等。3) 不同传感器的特征匹配：由于不同传感器所获得的图像常常表现出不同的特性，捕获了现实场景不同层面的信息，因此这种类型的图像特征匹配常常用于配准任务中，将不同图像进行逐像素的对齐以便于后续的信息融合。代表性的应用领域包括医学影像配准，红外与可见光配准以及遥感图像配准等等<sup>[28]</sup>。针对更加通用，具有多种不可预测的成像条件的情况，图像特征匹配是图像检索应用<sup>[29]</sup> 的一个重要手段。

由此可以看出，图像特征匹配作为一项基础且重要的技术，在诸多的视觉任务中都起着关键的作用，因此对其的研究具有极大的现实意义与应用价值。然而，由于现实数据的复杂性以及各种不确定性的影响，这一技术的应用也面临着多方面的挑战，其中包括广为人知的离群点问题。

由原理上来说，在理想的情况下，从图像中提取的特征可以足够完备且足够精确，便能够从源头处完全消除离群点的存在。然而在实际中这几乎是不可能完成的任务，从一方面来说，由于所提取的特征具有局部的特性，因此便不可避免的具有局限性，在重复或者相近纹理的情况下不能保证精确匹配。从其他方面来说，由于现实场景的高度复杂的特性以及无处不在的噪声的影响，构造普适的特征提取与描述的方法仍旧是计算机视觉领域最具挑战性的课题之一。而离群点，即便只占所有匹配中的一小部分，对于后续的各种视觉任务的影响也却也是灾难性的。举例来说，在三维重建任务中，一

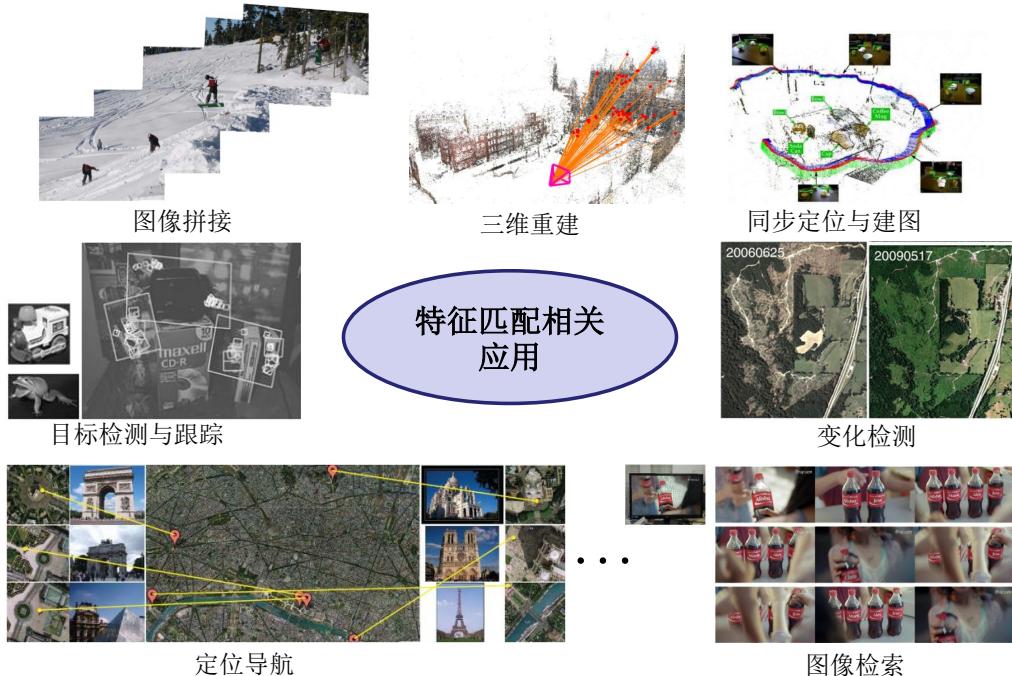


图 1.1 基于图像特征匹配的视觉应用。

个核心的问题是需要利用图像之间的匹配关系来估计相机自身的姿态，而这一任务对于精度的要求是极其严格的，假如输入之中存在离群点，将导致完全错误的姿态估计结果，使得重建任务失败。对于图像特征匹配中的离群点问题而言，还存在着另外的要求。许多以图像特征匹配为基础的视觉任务需要实时性作为依托，例如在视觉同步定位与建图中，必须在毫秒级别的时间内得出结果。因此，对于计算效率的要求也是解决离群点问题的一个重大挑战。

进一步来说，离群点的问题本身在数学上也具有极高的复杂性。在几何模型拟合的任务中，在具有离群点的情况下找到最优解被证明是 NP-hard 的问题<sup>[30]</sup>，因此对于这一问题的研究本身具有一定的理论价值。另一方面，离群点问题的存在在视觉领域中具有广泛的特性，在图像特征匹配之外，点云配准以及 Triangulation 任务中也具有这一挑战<sup>[18, 19]</sup>。综上所述，对于图像特征匹配问题中的离群点问题进行研究具有重要的理论和实际价值。

### 1.3 图像特征匹配研究现状

在本节中，我们将按照在图像特征匹配中处理图像的流程来顺序介绍该领域的研究现状。值得注意的是，由于近年来基于深度学习的文章研究的思路与传统的方法存

在着较大的差异，同时本文所研究的方法主要基于传统的研究范式，本节将基于深度学习的方法总结为单独一节以保证内容的完整性。在最后，我们将国内的相关研究也做出了总结。

### 1.3.1 图像特征点的检测

正如前文所说，图像特征匹配的第一步是检测图像的特征点或者说兴趣点，常见的兴趣点包括图像中常见的角点、交叉点以及闭合区域的中心点等物理信息比较明确的点。提取兴趣点的一般思想可以分为两类，即设计一种特殊的对于某些结构敏感的响应函数<sup>[31,32]</sup>，或者从特征线以及轮廓中进行稀疏采样<sup>[33]</sup>。作为特征点检测最早期的工作，Morevec 于 1977 年首次提出了“兴趣点”的概念并提出了一种基于局部像素灰度差异的方法<sup>[34]</sup>。随后，针对这种方法所具有的缺陷，例如对于方向、尺度、仿射变换以及噪声的敏感度和高时间消耗，一系列的改进算法被提出。经典的 Harris<sup>[35]</sup> 以及 Hessian<sup>[31]</sup> 检测算法利用图像域中的一阶和二阶导数信息来检测图像中潜在的角点，同时保证了方向的不变性。为了避免计算消耗较大的导数计算以满足实时性任务的需要，基于局部区域像素灰度比较的快速方法也得到了充分的研究。典型的例子包括 SUSAN 算子<sup>[32]</sup>，以及采取了不同的像素的比较方法和比较范围的 FAST<sup>[36]</sup> 算法，和在多种实时性视觉任务中得到了广泛应用的 ORB 算法<sup>[17,37]</sup>。总的来说，基于像素的直接比较的特征点检测算法具有效率高、速度快的优点，并且具有一定的方向不变性，然而其显著的缺点是对于尺度以及仿射变换较为敏感，相对而言不够鲁棒。

在特征点检测领域中，带有尺度信息的斑点特征形成了另一类重要的方法。这类方法最早可以追溯到 Lindeberg 等人所提出的高斯拉普拉斯函数 (Laplace of Gaussian, LoG) 函数的响应来实现<sup>[38]</sup>。成为了尺度空间理论的基础。该理论利用不同的标准差所定义的高斯响应函数来构建尺度空间，并通过尺度空间中的极值搜索来提取对尺度、方向以及噪声均鲁棒的尺度信息。为了避免大量的计算，Lowe 等人引入了一种高斯差分 (Difference-of-Gaussian, DoG) 的方法来近似 LoG 的计算以检测具有尺度信息的特征点<sup>[12]</sup>，并且由 Bay 等人利用积分图像以及近似的 Hessian 矩阵进一步加速了这一过程<sup>[39]</sup>。在 Alcantarilla 等人所提出的 KAZE<sup>[40]</sup> 以及 A-KAZE<sup>[41]</sup> 算法中，一种基于非线性扩散的尺度空间被提出来替代原有的 DoG 方法。为了进一步提升匹配的精准度，特征点的检测通常需要精确的位置信息并且保证检测得出的点集具有较高的可重

复性。因此，大多数的特征检测的方法会采用非极大值抑制的方法来提升所检测得到的特征点的显著性与稳定性。同时，在像素空间的插值理论也被用来估计特征点在亚像素空间的精确位置，以减少定位误差。更加具体的关于特征点提取的相关的综述请参考<sup>[7, 42–46]</sup>。

在两幅图像中检测到特征点之后，图像特征匹配的任务便转化为以两个特征点的集合作为输入，以精准的匹配关系作为输出的过程。这一问题虽然已经是图像特征匹配的一个抽象，数据输入的复杂度也大大降低，但由于其组合优化的本质特性，仍旧成为了学术界研究的热点课题之一，存在着大量相关的研究与诸多的分支。广义上而言，此时建立匹配的方式可以粗略分为两类，即直接匹配策略与简化匹配策略。这两种策略存在的差异是本质性的，因此二者形成了各自独立的研究路线，均得到了充分的研究。具体来说，直接匹配策略倾向于保留问题的原始定义，即直接考虑解决两个特征点集之间的匹配关系。这一问题的复杂度是惊人的，即便考虑简单的情形，对于两个可以完全一一对应匹配的具有  $N$  个数据点的点集而言，其搜索空间也存在  $O(N!)$  种不同的情况。与之相对，简化匹配策略考虑了问题的一种退化形式，即首先利用特征点的局部图像信息作为辅助（描述子），仅考虑描述子足够相似的一部分匹配。这种策略大大降低了匹配问题的复杂度，然而离群点问题仍旧是一个难以跨越的障碍。本文的主要研究内容建立在简化策略之上。

### 1.3.2 直接匹配策略

在直接匹配策略中，图像特征匹配被抽象为两个离散的点集的匹配问题，也被称作点集匹配问题<sup>[1]</sup>。假设给定的两个点集分别为  $\mathcal{X} = \{\mathbf{x}_i : i \in \mathbb{N}_N\}$  和  $\mathcal{Y} = \{\mathbf{y}_j : j \in \mathbb{N}_M\}$ ，其中  $\mathbb{N}_N$  表示不超过  $M$  的自然数， $\mathbb{N}_N$  所代表的意义同理。对于匹配问题而言，一般性的目标可以认为是求解一个指派矩阵  $\mathbf{P} = \{p_{ij}\}_{i=1,j=1}^{N,M}$ ，其中的任一元素  $p_{ij}$  都代表着  $\mathbf{x}_i$  和  $\mathbf{y}_j$  的一种对应关系，即当  $p_{ij} = 1$  时代表  $\mathbf{x}_i$  与  $\mathbf{y}_j$  是匹配的，当  $p_{ij} = 0$  时则反之。在最理想的情况下，两个点集之间满足一一对应的关系时，这一类矩阵  $\mathbf{P}$  被称作置换矩阵（Permutation Matrix）。由于置换矩阵定义在离散域，一般将其松弛到连续域来简化问题。同时，在匹配问题的求解中，也存在着一些实际应用中所产生的与置换矩阵含义接近的概念，例如加入了虚拟点（dummy point）来应对离群点的指派矩阵，或者仅对于匹配进行单向的一一对应关系约束（而非置换矩阵所代表的双向

一一对应约束) 的指派矩阵。同时也存在着一类方法并不依赖于指派矩阵的定义, 而将两个点集之间存在的变换作为求解的最终目标。此时, 最终的指派矩阵可以通过将变换之后得到的两个点集进行最近邻匹配来得到指派矩阵。基于直接匹配策略的方法大致可以分为两类, 即基于变换估计的点集匹配与基于图模型的点集匹配两种。

i) 基于变换估计的特征匹配算法。在这一类方法中, 其核心的假设在于两个输入的点集中其中一个点集可以从另一个点集经由某种预定义的变换来得到, 因此还原出这种变换之后就可以自然地建立两个点集之间的匹配关系。这一类方法的建模方式通常以预定义的某种变换模型作为参数, 以两个点集之间重合程度的某种度量作为目标函数, 从而在优化的框架下对问题进行求解。针对图像特征匹配的问题而言, 变换模型的定义也有着多种选择。最为基本的变换包括平移、尺度、旋转、仿射等, 可以使用简单的  $3 \times 3$  的矩阵来表示。考虑到图像中存在的更广泛的场景, 非刚性变换的使用更加灵活, 也更加普适。早期的基于变换估计的点集匹配的算法以经典的 Chui 等人所提出的 TPS-RPM<sup>[47]</sup> 为代表, 其使用的非刚性变换模型是广为人知的薄板样条函数 (Thin-Plate Splines, TPS)。TPS-RPM 同时定义了指派矩阵与变换模型, 并通过变换模型与指派矩阵的交替迭代估计, 附加一种模拟退火的策略, 来得出点集之间的匹配对应关系。继 TPS-RPM 之后, 后续的点集匹配算法通常以概率模型作为求解的框架, 这一类算法同样可以分为两类, 第一类方法以 Myronenko 所提出的一致性点漂移算法 (Coherent Point Drift, CPD)<sup>[48]</sup> 为代表, 将其中一个点集看做从某一个高斯混合模型的概率分布中采样的结果, 而另一个点集视作高斯混合模型的均值, 最终通过最大似然估计 (Maximum Likelihood Estimation, MLE) 的目标函数对问题进行抽象并求解。CPD 算法同时还引入了高斯径向基函数作为变换模型, 并在变分法的框架下证明了这一做法的最优性质, 同时使用 Tikhonov 正则化<sup>[49]</sup> 来约束平滑的变换模型。这一类方法在某种意义上可以视作 TPS-RPM 的扩展, 同样在指派矩阵与变换模型参数估计之间迭代进行求解。第二类以 Jian 与 Vemuri 所提出的 GMM 算法<sup>[50]</sup> 为代表, 其将两个点集均转化为以点集内元素作为均值的高斯混合模型分布, 并通过最小化两个分布之间的  $l_2$  距离来求出最优的变换模型。这一类方法的特点是完全不依赖于指派矩阵, 而只关心如何求出最终的变换模型。值得一提的是, 除非刚性的变换模型之外, 这些方法也都支持刚性变换模型的求解。

ii) 基于图模型的特征匹配算法。从机器学习的角度来说, 基于图模型的匹配代表

着一类适用范围更广的算法。这一类方法并不假设一种变换模型的存在，而通过将两个点集分别构建为具有顶点与边的图结构来对问题进行分析求解，因此也叫作图匹配 (Graph Matching)。在早期的图匹配的研究中，通常在概念上将问题进行高度的抽象，对于边的关系仅考虑是否连接两种情况，从而定义为一种图同构的问题<sup>[51, 52]</sup>。然而这种建模方式过于严格要求图的相似性，因此求解难度极高并且不能满足实际视觉问题的需求。现在的图匹配对于图的定义具有更大的包容性，对于各种边的定义也依托于问题先验。在计算机视觉的研究中，图匹配一般被建模为一种二次规划问题 (Quadratic Programming) 的特殊形式，二次分配问题 (Quadratic Assignment Problem, QAP)。如前文所说，这个问题的求解空间正是所谓的置换矩阵，定义在离散域且具有多种线性约束。由于该问题是一种典型的组合优化问题，且被证明具有 NP-complete 的特性，主流的研究图匹配的算法一般将问题进行简化，以期在合理的时间消耗中获得高质量的解。这一类方法的求解一般集中于对问题的约束进行不同的松弛，转化为更加容易求解的形式的思路。具有代表性的图匹配方法包括基于谱松弛<sup>[53, 54]</sup>，凸松弛<sup>[55, 56]</sup>，凸至凹松弛<sup>[57, 58]</sup>以及连续性松弛<sup>[59, 60]</sup>的各种方法。总而言之，基于图模型的匹配方法不同于基于变换模型的匹配方法，其对于数据的假设更为宽松，仅仅通过概念上的边的定义来挖掘相关性进行最终的求解，因此具有较强的理论研究意义。

总的来说，直接匹配的策略，不论是基于变换模型还是基于图模型的匹配算法，其计算复杂度由于问题本身的组合优化特性，均达到非常高的地步，同时也较容易受到噪声以及离群点的影响，因此基于简化的策略由于其高效率的特点在实际的视觉任务中得到了更多的应用。

### 1.3.3 简化匹配策略

在简化匹配的策略中，一般首先针对两个特征点集的每一个特征点进行特征描述，而根据描述子的欧氏距离，将建立起一个粗略的匹配集合。即便这种做法可能会损失一部分的正确匹配，但无疑大大简化了图像特征匹配这一问题，使得后续高效准确的求解成为可能。

描述子在简化匹配的策略中具有至关重要的意义，其是特征点周围的场景信息的有效表征。常用的传统的特征描述子可以粗略的分为居于局部图像梯度统计的浮点型描述子和基于像素灰度比较的二值型描述子。最为代表性的描述子可能是 Lowe 所提

出的 SIFT 描述子<sup>[12]</sup>, 其通过对于局部区域进行网格的划分并利用梯度主方向的信息构造一个向量作为描述子。SIFT 描述子在深度学习提出前是最受认可的图像匹配算法, 对于光照、旋转、尺度以及图像的噪声均有较强的鲁棒性。SIFT 的提出激发了一系列的改进工作, 在 Bay 等人所提出的 SURF<sup>[39]</sup> 中, 引入了 Haar 小波相应的策略, 并使用高斯函数进行局部的加权, 最终统计扇形区域的导数方向生成描述子。其他较为流行的 SIFT 改进包括基于特征降维的 PCA-SIFT<sup>[61]</sup>, 以及 A-SIFT<sup>[62]</sup>、DSP-SIFT<sup>[63]</sup> 等。基于像素灰度比较的二值特征的描述方法也存在着一系列的工作。Michael 等人于 2010 年在特征点局部的矩形区域内针对不同的采样形式进行了对比测试, 提出了一种被称作 BRIEF 的二值描述方法<sup>[64]</sup>, Stefan 等人于 2011 年提出了一种基于变尺度同心圆采样形式的特征描述方法 BRISK<sup>[65]</sup>。进一步, Alexandre 等人提出了一种基于视网膜采样的二进制描述子 FREAK<sup>[66]</sup>。在工业界得到广泛应用的 ORB 特征<sup>[37]</sup> 中也使用了二值描述方法, 具体来说, ORB 在 BRIEF 特征描述方法的基础上, 结合一种贪心算法来确定二进制的编码方式。其他较为流行的二进制描述子包括 Yang 等人提出的 LDB<sup>[67]</sup> 以及 Alcantarilla 等人所提出的 A-KAZE<sup>[41]</sup> 等改进形式。总的来说, 二值型描述子生成以及进行相似度度量时相较于浮点型描述子有更快的速度, 但一般不如浮点型描述子准确。针对特征描述的相关文献的综述, 可以参考<sup>[42-46]</sup>。在从真实数据中抽象出来的带有较强结构信息的点集中, 同样可以构建针对离散点的描述子。这一类描述子脱离了图像本身的信息, 而从离散点集合的拓扑结构的角度来辅助匹配算法。这一类算法的代表是 Belongie 等人所提出的形状上下文<sup>[33]</sup>(Shape Context)。

然而, 无论使用何种特征描述的算法, 构建的匹配集合中都将不可避免的含有着或多或少的离群点。这一方面是由于描述子算法本身固有的缺陷, 在复杂的现实世界的图像数据中常常不能有效地表征所编码的图像, 造成信息的损失甚至畸变。另一方面, 局部的图像描述本身是不可靠的, 在处理相同或相似的纹理时也会失效。因此, 设计针对离群点问题的鲁棒算法是有必要的。

在使用特征描述的算法后, 图像特征匹配问题便可以抽象为从一个给定的初始匹配集合中找出正确匹配的问题, 以实现鲁棒性的匹配。这一类方法大致可以分为两个类别, 即基于全局变换的算法以及基于局部一致性的算法。基于全局变换的算法其目的在于由受到离群点污染的数据中还原一种潜在的平滑的变换, 从而识别出正确的匹配。Li 等人所提出的 ICF 方法<sup>[68]</sup> 是这类方法的一种早期代表, 其定义了一种双向的

匹配函数对于平滑变换进行建模，并利用支持向量回归 (Support Vector Regression, SVR) 的鲁棒特性来学习这一函数的参数。Ma 等人提出了一种基于向量场一致性的方法<sup>[69]</sup>(Vector Field Consensus, VFC)，其利用 RKHS 作为函数定义的空间，并使用 Tikhonov 正则化理论来对函数的平滑性进行约束，最终在一种考虑离群点的鲁棒概率模型框架之下还原出一种最优的函数作为平滑的变换。这种 VFC 方法由于其简洁有效的数学理论，成为了许多后续工作的基础<sup>[69-78]</sup>，例如基于局部线性迁移的匹配算法 LLT<sup>[79]</sup> 以及基于流形正则化的算法 MR-RPM<sup>[80]</sup> 等。基于局部一致性的方法由于其高效的特性也得到了充分的研究，这些算法首先包括早期的图匹配方法，由于对于约束的松弛，一些图匹配的方法也可以作为鲁棒算法来处理离群点问题，例如 SM<sup>[53]</sup> 以及 GS<sup>[54]</sup> 等。近年来，这一类方法得到了充分的关注与研究，例如基于局部邻域一致性的保留的 LPM 方法<sup>[81]</sup>，以及基于网格划分的运动特性统计算法 GMS<sup>[82]</sup> 等。这一类方法由于其极具优势的速度得到了广泛的应用。同样地，在文献中也存在着一类综合了全局与局部特征的鲁棒匹配算法。在 CODE 算法<sup>[83]</sup> 中，通过设计一种结合点坐标与其运动的双边变换将匹配问题转换到一个新的域中，并使用非线性回归技术实现正确与错误匹配的自动判别。在 BD 算法<sup>[84]</sup> 中，使用了分段仿射模型对局部变换进行建模，但同时也在全局添加了一种对于变换的剧烈程度的有界限制。

结合 3D 视觉的原理，从图像特征的匹配中估计出具有物理意义的几何模型也是计算机视觉学界研究的重点方向之一。由于离群点的存在所带来的理论上的限制，许多研究者由不同角度出发提出了一系列解决这一问题的思路。从一般的特性上来说，鲁棒的模型估计算法可以分为确定性与非确定性两类算法。非确定性的算法以经典的随机采样一致性算法<sup>[85]</sup>(RANdom SAMple Consensus, RANSAC) 为代表，由于随机采样过程的引入，因此算法每次运行的结果会有不同。近二十年里，由于 RANSAC 算法简单有效的特性，一系列的算法被提出来从 RANSAC 算法的各个阶段来从精准性及计算效率等方面改进这一算法。代表性的算法有使用邻域信息或者匹配可信度的先验信息加速采样过程的 GroupSAC<sup>[86]</sup>, PROSAC<sup>[87]</sup>, NAPSAC<sup>[88]</sup> 等算法，加速模型验证过程的 SPRT<sup>[89]</sup> 以及  $T_{d,d}$  test<sup>[90]</sup> 方法，以及改进模型质量评估过程的 MLESAC<sup>[91]</sup>。同时，LO-RANSAC<sup>[92,93]</sup> 创新性的引入了局部优化 (local optimization) 步骤，有效改善了 RANSAC 的稳定性与精准度。在两视图几何的具体应用中，DEGENSAC<sup>[94]</sup> 考虑到了基础矩阵估计中的退化场景，并提出了在 RANSAC 的基础上提出了改进的

采样策略。综合了基于 RANSAC 算法的一系列有效的改进，USAC<sup>[95]</sup> 被提出以达到最优的鲁棒模型估计效果。近年来，GC-RANSAC<sup>[96]</sup> 在局部优化的过程中引入 graph cut 算法，进一步提高了模型估计的质量。MAGSAC<sup>[97]</sup> 以及 MAGSAC++<sup>[98]</sup> 的提出突破了传统 RANSAC 方法的局限，不再需要事先定义内点与外点的阈值界限，从而增强了算法的实用性。

大体上来说，确定性的鲁棒模型估计的算法可以分为两类，一类着重在理论上为这一问题提供新的见解，并给出具有全局最优保证的解法<sup>[99-101]</sup>，代表性的分支存在着基于 branch-and-bound<sup>[102]</sup>，tree search<sup>[103]</sup> 或者 enumeration<sup>[104]</sup> 的各种算法。由于本身这一类问题的复杂特性，这种具有全局最优保证的解法通常不能完全脱离穷举搜索的局限，时间复杂度极高，因此学界更多的将注意力放在一类仅仅保证局部最优但计算效率较高的算法中，在近年来也得到了长足的发展。理论上，这一类算法通常将鲁棒估计的问题建模为针对一种非凸的损失函数求解其相对于模型参数变化的最优解的过程，即便求解过程仅仅是近似的或者渐进性精确的。代表性的方法包括基于平滑的替代函数的优化方法<sup>[105]</sup>，基于惩罚函数的优化方法<sup>[19]</sup>，基于 ADMM 理论的优化方法<sup>[19]</sup> 以及基于双向凸规划的优化方法<sup>[106]</sup> 等等。值得一提的是，即便这一类方法相较于全局优化的确定性方法效率更高，在大多数情况下基于随机采样的方法通常具有更优的速度，因此在实际中的得到的应用更加广泛。

### 1.3.4 深度学习方法

近年来，深度学习方法由于其遥遥领先的学习与表征能力，在人工智能的各个领域均掀起了一波研究的浪潮。在图像特征匹配的各个方向与各个阶段，基于深度学习的方法都展现出了卓越的效果。

深度学习在图像特征匹配中最为常见，也最受关注的应用是图像特征的检测与描述方面。相对于 SIFT 等人工所设计的图像特征，由大量数据所驱动的深度网络被认为可以完全取代传统的方法，并达到更好的效果。在这一领域，根据深度网络所替代的功能，基于深度学习的方法大致可以分为三类，即特征点检测学习、描述子提取学习以及端到端的特征点检测以及描述子提取联合学习。特征点检测方向的工作包括 TILDE<sup>[107]</sup>，TCDet<sup>[108]</sup>，QuadNet<sup>[109]</sup> 以及 Key.Net<sup>[110]</sup> 等，其目的是使用深度学习网络学习检测图像中重复率以及分布特性较好的点的位置。描述子提取方向的工作

其目的是利用深度网络强大的表征能力来将图像块编码为紧凑表示的低维向量，包括 Deepdesc<sup>[111]</sup>, TFeat<sup>[112]</sup>, L2-net<sup>[113]</sup>, HardNet<sup>[114]</sup>, SOSNet<sup>[115]</sup> 以及 LogPolarNet<sup>[116]</sup> 等。不同于传统的做法，也存在一些方法尝试使用图像信息之外的信息，例如几何场景信息或者全局上下文关系等来得到更好的描述子，这方面的工作包括 GeoDesc<sup>[117]</sup> 以及 ContextDesc<sup>[118]</sup> 等。端到端的深度网络将特征点检测以及描述子的提取集成到同一学习框架中，代表性的工作包括 LIFT<sup>[119]</sup>, DELF<sup>[120]</sup>, SuperPoint<sup>[13]</sup>, LF-Net<sup>[121]</sup>, D2-Net<sup>[122]</sup> 以及 R2D2<sup>[123]</sup> 等。值得一提的是，在描述子匹配的场景中，不同于传统的最近邻匹配策略，SuperGlue<sup>[124]</sup> 运用图神经网络对这一任务进行了改进，取得了卓越的效果。

在图像特征匹配的其他方面，同样存在着使用深度网络达到更好效果的一系列工作。例如在图匹配的方向，许多工作使用深度学习网络将图像信息融合到图匹配的过程，并在一个统一的网络框架中学习<sup>[125-128]</sup>。同样，深度学习的方法在几何模型估计中也发挥了突出的作用。这一类方法通常将深度学习引入来解决离群点的问题，并且将其和三维视觉的应用紧密联系在一起。这一方向的论文首先自然地包含对于 RANSAC 算法的深度网络的改进，例如 DSAC<sup>[129]</sup> 以及 NeuralGuidedRANSAC<sup>[130]</sup> 等，同样包括对于经典的确定性鲁棒估计算法的深度网络改进，例如 DFE<sup>[131]</sup>。同样，受到点云数据处理领域的一些重要学习算法，例如 PointNet<sup>[132]</sup>，的影响，一系列的工作将类似的思路引入匹配数据中以解决离群点的问题。这一类的方法通过网络来提取匹配数据的特征，并在最终得出其正确与否的分类，从而为后续的视觉任务扫除障碍。这些方法包括 Yi 等人所率先提出的 LFGC-Net<sup>[133]</sup>，以及在其基础上的一系列改进方法，包括 OANet<sup>[134]</sup>，ACNe<sup>[135]</sup> 等方法。

### 1.3.5 国内相关研究

国内学者对于图像特征匹配领域也存在着许多的研究工作。国防科技大学赵健<sup>[136]</sup> 以及复旦大学的宋智礼<sup>[137]</sup> 在各自的博士论文中分别对于点模式的匹配问题以及图像配准相关的技术与应用做出了系统的研究工作。华中科技大学马佳义<sup>[1]</sup> 在博士论文中重点研究了基于非参数模型的点匹配算法框架，利用正则化理论提出了一系列的点集匹配以及图像特征匹配的算法。哈尔滨工业大学于伟<sup>[138]</sup> 使用卷积神经网络来解决图像匹配的问题，对深度学习的深层特征表达以及语义匹配做出了较为深入的研究。华

中科技大学柳成荫<sup>[139]</sup>在其博士课题中，针对实际的应用需求，对于多模与多视角的图像配准问题做出了专门的研究。

综上所述，图像特征匹配是一个方向众多，同时问题也较为复杂的领域。由于实际应用中图像场景的复杂性，现有的特征匹配算法通常具有多方面的局限。本文的目的在于，针对最为常见的离群点问题，进行深入而系统的研究，以提出高效率的鲁棒算法，从而突破图像特征匹配问题的现有局限。

## 1.4 图像特征匹配发展趋势

图像特征匹配的问题由来已久，是计算机视觉领域的核心研究内容之一。然而由于图像数据天然具有的复杂性，以及实际应用所面临的各种限制，图像特征匹配仍然是一个充满活力，有许多问题待研究的开放性课题。对于这一问题而言，综合当前的研究现状与相关的应用需求，图像特征匹配技术的发展趋势主要包括以下几个方面。

### 1.4.1 图像特征匹配与应用的结合

图像特征匹配作为一个基础性的研究问题，在诸多视觉任务中均居于关键性的地位。然而，在实际中，图像特征匹配的质量就其本身来说是难以评估的，这为设计更加先进的算法带来了困难。或者说，即便针对图像特征匹配本身的指标得到了提高，也并不意味着其在实际的应用中，存在后续任务的前提下，仍然能够带来性能上的提高。这一点在最近的研究中，藉由三维重建任务得到了充分的体现。三维重建的算法是一个复杂的系统，图像特征匹配是其中一个重要的步骤，但最终的总体性能也将受到系统中其他部分的制约。在尽可能创造了一个公平的环境，在所有步骤中集合了现有的领先算法之后，才能得到对于各种算法的性能评估得到一个准确的结论，而在此之前，许多声称达到了最先进水平的工作的结论是极不精确的。因此，对于图像特征匹配来说，下一个可能的突破方向可能在于将其与一些特定的应用相结合。考虑到特定应用的限制条件与需求，例如图像拼接、图像检索、目标追踪等，图像特征匹配的实际作用有望得到进一步的提高。

### 1.4.2 深度学习技术的引入

深度学习的技术在计算机视觉中许多的领域均实现了突破性的进展，在图像特征匹配中，其引起的革命性浪潮仍在发展之中。深度学习技术由于其强大的特征提取与表达的能力，在图像特征匹配中最为典型的应用是将其用于图像的特征检测与提取中。在这一方向中，由于图像数据的复杂性，传统的人工设计的方法常常具有各种局限性，不能适应多变的图像场景与各种干扰，因此可以期待数据驱动的深度学习技术能够突破传统的技术的局限，将这一任务推向一个新的层次。深度学习的技术可以被用于图像特征匹配的各个阶段，包括特征检测、特征提取以及特征的匹配的过程等等。对于现有的框架而言，深度学习技术有望在许多图像特征匹配阶段性的任务中替代传统的算法设计，另外，尽可能综合更多的步骤进去，以接近端对端的解决方案也是一个值得考虑的方向。同时，在特征被抽象为点的集合之后，数据的特性也将发生巨大的变化，具有离散的点数据所特有的无序性，如何对这一类型的数据进行处理以促进图像特征匹配的性能也是一个研究热点。图像特征匹配在三维视觉中的应用常常设计到内在的接结构的计算，这些内在的几何结构是一种物理世界规则的反映，简单而直接。如何将深度学习与这一类的内在集合结构、物理规则有机的结合起来同样是一个巨大而意义重大的挑战。

### 1.4.3 多图像协同匹配

在图像特征匹配的研究中，大多数的方法考虑的是一对图像之间的匹配任务，而多幅图像的协同考虑的问题由于其复杂度而常常被忽略。从概念上来说，考虑成对图像相较于考虑多幅图像的协同匹配具有天然的劣势，在成对图像中所可能遇到的约束不足以及歧义性问题等可以在使用其它图像所提供的额外的信息和约束的前提下得到解决。同时，协同匹配的概念也自然地满足了对于一个系统来说的一致性的特点，而仅仅考虑成对图像的匹配任务，常常可能会导致综合后整体上的矛盾结果。

对于多图像协同匹配的任务来说，如何对其进行有效的建模并提出高效的解法其本身便是一个极具理论研究价值的问题，是现有算法迎来新的突破的重要契机。多图像协同的理论在检测、分割以及超分等任务中已经得到验证并取得了不错的成效。然而在图像特征匹配的相关研究中，多图像协同的做法尚未得到足够的重视。值得注意的是，多图像协同的理念在许多任务中是自然的，例如三维重建与图像拼接中，具有

重合的场景的图像是现实存在并且容易获得的。并且，在这些任务中，视觉算法的精度常常是需要优先考虑的对象，这也是多图像协同匹配的优势所在。因此多图像协同的匹配也是未来图像特征匹配的重要发展路线之一，具有极其广阔前景。

## 1.5 论文研究内容与组织结构

图像特征匹配是一个研究方向众多的复杂任务，且有着巨大的理论研究意义与实际的应用价值。本文着力于研究其中所涉及到的广泛存在的离群点问题，并结合多种应用的需求提出高效率的鲁棒算法来解决这一挑战。本文的主要研究内容与组织结构的安排如下：

第1章从图像特征匹配的定义以及主要的分类入手，引出本文所重点研究的图像特征匹配的任务，并且结合当前的研究背景与应用需求介绍图像特征匹配所具有的实际应用价值和理论研究意义，同时分析归纳改问题学术研究的现状与今后的发展趋势，最后对于本文的研究内容与组织结构进行总结。

第2章由统计学习的角度切入，主要研究一种从受离群点污染的数据之中恢复内点所代表的潜在平滑函数的方法。传统的学习算法研究中，这一任务通常使用正则化理论来进行建模求解，即将待求的函数定义在再生核希尔伯特空间 (RKHS) 中，并使用 Tikhonov 正则化理论来保证所求得的函数的平滑性。本文研究一种鲁棒的基于傅里叶基的紧凑表示方法来建模这一类平滑函数，从理论上解释，并结合图像特征匹配任务在实验中证明了所提方法的优越性，在不损失精度的前提下，运行效率相较于传统的算法得到了超过一个数量级的巨大提升。

第3章考虑经典的图模型在图像特征匹配问题中的应用。图模型的建模方式具有灵活便捷的特点，本文结合图像特征匹配的应用需求，提出了一种基于概率图模型的鲁棒误匹配去除方法。所提方法将概率图模型与图像特征匹配的现实问题结合起来，将模型转化为一类特殊的整数二次规划问题，并提出了基于 Frank-Wolfe 方法的快速求解算法。实验中证明，所提方法在多种图像匹配的数据中均具有优势，并且可以处理图像中存在非平滑的多模式运动的情况。

第4章结合图像特征匹配在三维视觉中的应用需求，提出了一种鲁棒的模型估计的算法。计算机视觉中经典的多视图几何的理论揭示了图像特征匹配在还原真实世界三维结构中所起到的重要作用，具体来说，图像特征匹配中蕴含着还原三维结构关键

的几何信息，例如基础矩阵以及单应矩阵。本文主要研究在存在大量离群点场景下的模型估计问题。特别地，所提方法考虑了几何模型的数学结构，创新性的引入了  $l_1$  损失来提高求解对于离群点的鲁棒性。在求解方面，所提方法将问题与鲁棒子空间恢复理论联系在一起，并在此基础上提出了多种基于问题特性的求解算法，实现了鲁棒且高效率的几何模型估计。

第 5 章对全文工作内容进行了总结，同时对课题研究成果的创新性进行了说明，并针对本文中理论研究方法与实际应用表现上的不足对下一步的研究工作进行了展望。

## 2 基于紧凑表示的鲁棒平滑函数恢复算法研究

### 2.1 引言

在图像特征匹配，以及一大类的机器学习相关研究问题中，离群点问题的解决首先需要对于内点做出某种广义上的通用的假设。从这一点来说，最自然同时也是应用最广泛的是将内点的数据视作从某种平滑性的函数中采样得来的结果。基于这种思路，在现有的工作中，正则化理论<sup>[140, 141]</sup>得到了最广泛的使用。正则化理论是众多机器学习算法的基石之一，其核心思想是定义某种再生核希尔伯特空间 (Reproducing Kernel Hilbert Space) 作为所研究函数的来源，并通过正则化的方式对于函数的复杂度加以限制，从而保证最终得到的函数是平滑的。然而，正则化理论一个巨大的弱点是在其基础上所研究的算法通常具有非常高的计算复杂度，时间上为  $O(N^3)$ ，这一点大大限制了它的实用价值。

图像匹配问题中，这种基于平滑函数假设的思路代表着一类重要的方法，将特征点之间的映射视作一种特殊的平滑函数对于解决实际问题具有重要的作用。在这种假设之下，对于实际图像场景的高复杂度具有较高的适应性，甚至可以在一定程度上处理非刚性的情况，例如图像产生局部的形变或者场景包含一定程度的不连续运动。图像特征匹配中著名的向量场一致集<sup>[69]</sup>(Vector Field Consensus, VFC) 方法可以视作平滑函数假设与鲁棒学习方法的一个结合，获得了巨大的成功。为解决正则化理论所带来的高复杂度的问题，一些快速近似的算法在 VFC 的基础上被提出，包括 FastVFC<sup>[69]</sup> 以及 SparseVFC<sup>[142]</sup>。在 FastVFC 中，对于核矩阵的一种低秩分解的方法被使用，使其在迭代计算的环节的复杂度降低到了  $O(N)$ ，然而低秩分解的算法本身需要  $O(N^3)$  的复杂度，因此没有完全的解决问题。在 SparseVFC 中提出了一种稀疏随机基选取的方法，可以将复杂度降为  $O(N)$ ，然而这种做法的理论上的论文尚不充分，在实际中常常导致数值的不稳定性导致速度大幅度降低。

在本章中，我们提出了一种新的从受到离群点污染的数据中学习平滑函数的方法。不同于正则化理论，我们研究了一种基于傅里叶基的表示方法来建模平滑函数。这种

表示的好处是，由于其紧凑性，对于平滑函数的表达一般只需要很少的一部分参数，这一点从理论上保证了所提方法高效率的特点。我们将所提的方法用于解决图像特征匹配中的离群点问题，即在图像特征匹配的过程中，我们使用简化匹配的策略。首先在图像中检测特征点并提取描述子，然后根据描述子的欧氏距离建立最近邻匹配的关系，此时我们可以得到一个存在着错误匹配（即离群点）的初始匹配集，此时本章所提的算法可以用于识别出真正的匹配而提出错误的匹配。在实验中，我们证明所提的方法可以达到与基于正则化理论的方法相近的精度，并且计算效率达到超过一个数量级的提升。同时，我们也将所提方法用于点集配准的任务中，证明了所提方法的泛用性。

## 2.2 基于稀疏数据的多元平滑函数回归方法

本节首先以广义的稀疏数据的多元平滑函数回归任务为例，来介绍我们提出的基于紧凑表示的方法。首先我们介绍经典的正则化理论，并由此引出我们的方法的具体介绍。

### 2.2.1 正则化理论

假设我们首先具有一个稀疏的采样集合  $S = \{(\mathbf{x}_n, y_n)\}_{n=1}^N \subset \mathcal{X} \times \mathcal{Y}$ ，其可以视作从某种定义在  $\mathcal{X} \times \mathcal{Y}$  的概率分布  $P$  中采样的结果。特别地，输入空间  $\mathcal{X}$  可以看做  $\mathbb{R}^D$  的一个子集，输出空间可以视作  $\mathbb{R}$  的一个子集。这一任务的目标是学习一个函数  $f$  使得误差的期望  $E[V(y, f(\mathbf{x}))]$  最小，其中该期望是相对于概率分布  $P$  的，而  $V$  是一个预先定义的损失函数，例如平方损失。解决这一问题的经典的方式是通过正则化理论，即将学习的目标函数  $f$  视作最小化如下的正则化风险泛函的一个解：

$$\min_{f \in \mathcal{H}} \sum_{n=1}^N V(y_n, f(\mathbf{x}_n)) + \lambda \|f\|_{\mathcal{H}}^2, \quad (2.1)$$

其中  $\|f\|_{\mathcal{H}}$  是 RKHS  $\mathcal{H}$  中的范数，由正定的核函数  $K$  来确定， $\lambda$  是平衡经验风险项与函数的复杂度（或者说平滑度）的一个参数。正则化理论表明，问题 (2.1) 的最优解必然具备以下的形式：

$$f(\mathbf{x}) = \sum_{n=1}^N c_n K(\mathbf{x}, \mathbf{x}_n), \quad (2.2)$$

其中  $\{c_n\}_{n=1}^N$  代表一个实数参数集合。核函数  $K$  的性质在于, 对于  $\mathbf{x} \in \mathcal{X}$ , 有  $K(\mathbf{x}, \cdot) \in \mathcal{H}$ , 对于  $f \in \mathcal{H}$ , 有  $\langle f, K(\mathbf{x}, \cdot) \rangle_{\mathcal{H}} = f(\mathbf{x})$ 。因此考虑到 (2.2) 有:

$$\|f\|_{\mathcal{H}}^2 = \sum_{m=1}^N \sum_{n=1}^N c_m c_n K(\mathbf{x}_m, \mathbf{x}_n). \quad (2.3)$$

同时考虑 (2.2) 和 (2.3), 则 (2.1) 可以简化为

$$\min_{\mathbf{c}} \sum_{n=1}^N V(\mathbf{y}, \mathbf{K}\mathbf{c}) + \lambda \mathbf{c}^T \mathbf{K} \mathbf{c}, \quad (2.4)$$

其中  $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$ ,  $\mathbf{c} = [c_1, c_2, \dots, c_N]^T$  与  $\mathbf{K}$  是  $N \times N$  的矩阵, 其第  $mn$  个元素为  $K(\mathbf{x}_m, \mathbf{x}_n)$ 。公式 (2.2) 所反映的结果即是所谓的表示定理 (*representer theorem*)。这意味着只要损失函数  $V$  是凸的, 那么问题 (2.4) 的全局最优解可以使用简单的数值算法来求得。考虑最简单的平方损失, 那么其解的形式可以通过求解如下的线性系统来获得:

$$(\mathbf{K} + \lambda \mathbf{I})\mathbf{c} = \mathbf{y}. \quad (2.5)$$

显然, 求解线性系统 (2.5) 的时间复杂度是  $O(N^3)$  的, 并且由于矩阵  $\mathbf{K}$  是  $N \times N$  的。这种计算复杂度极大地限制了其实用性。一个直接的想法是类似于 VFC 中提供的方法, 使用矩阵  $\mathbf{K}$  的低秩近似来加速计算, 然而 SVD 分解本身也具有  $O(N^3)$  的时间复杂度, 因此不能完全解决该问题。下面我们介绍如何使用紧凑表示的傅里叶基来解决这一问题。

### 2.2.2 基于紧凑表示的平滑函数学习

紧凑表示的主要思想在于利用特定的稀疏结构。在正则化理论中, 所寻找的函数定义在一个具有非常强的表征能力的函数空间内, 即 RKHS, 并且使用一个正则化项来控制该函数的复杂度。我们主要的思想在于, 对于平滑函数来说, 可以有更加紧凑的表示方法, 而发现并使用它将是设计高效率算法的关键。接下来我们介绍如何去构造这样一种表示方法。

不失一般性, 我们考虑输入空间为一个  $D$  维空间  $\Omega := [0, 1]^D$ , 这一点在实际中比较容易满足, 仅需要在事先进行一次数据的标准化操作即可。首先, 我们考虑在  $\Omega$  中所定义的拉普拉斯算子  $\Delta$ , 以及其所具有的特征向量  $\{\phi_1, \phi_2, \dots\}$  和特征值  $\{\mu_1, \mu_2, \dots\}$ :

$$-\Delta\phi_k = \mu_k\phi_k. \quad (2.6)$$

为了明确所研究函数空间的性质，我们需要首先确定边界条件。自然地，边界条件一般具有三种选择：

$$\phi_k(\mathbf{x}) = 0 \quad (\mathbf{x} \in \partial\Omega) \quad (\text{Dirichlet}), \quad (2.7a)$$

$$\frac{\partial}{\partial n}\phi_k(\mathbf{x}) = 0 \quad (\mathbf{x} \in \partial\Omega) \quad (\text{Neumann}), \quad (2.7b)$$

$$\frac{\partial}{\partial n}\phi_k(\mathbf{x}) + h\phi_k(\mathbf{x}) = 0 \quad (\mathbf{x} \in \partial\Omega) \quad (\text{Robin}), \quad (2.7c)$$

其中  $\frac{\partial}{\partial n}$  代表朝向输入空间定义区域之外的法向导数，而  $h$  则是一个取正值的常数。对于图像特征匹配或者点集配准任务来说，由于向量场在边界附近也可以具有较大的值，因此 Dirichlet 边界条件并不是一个好的选择。同时，Robin 边界条件较为复杂，并不是必要的，因此在本节提出的方法中我们选取 Neumann 边界条件。此时，这些特征向量具有解析的形式，即是傅里叶基中的 cosine 项<sup>[143]</sup>：

$$\mathcal{B}_\phi = \{\phi : [0, 1]^D \rightarrow \mathbb{R}, \mathbf{x} \rightarrow \prod_{d=1}^D \cos(x_d)\pi j_d | \mathbf{j} \in \mathbb{N}^D\}, \quad (2.8)$$

其中  $x_d$  代表  $\mathbf{x} \in \Omega$  中的第  $d$  位的值， $j_d$  代表  $\mathbf{j}$  中的第  $d$  个值。对于每个特征向量来说，相应的特征值是  $\pi^2 \|\mathbf{j}\|^2$ 。可以证明，这些基底  $\mathcal{B}_\phi$  共同形成了一个定义在  $\Omega$  中且由可测且平方可积的函数构成的函数空间  $L^2(\Omega)$  中的完备基<sup>[144, 145]</sup>。在  $D = 2$  的情况下，我们将一些基函数的可视化结果在 Fig. 2.1 中给出。

我们使用  $\{\phi_1, \phi_2, \dots\}$  来代表这些傅里叶基底  $\mathcal{B}_\phi$ ，其排列顺序由特征值从小到大来决定： $0 \leq \mu_1 \leq \mu_2 \leq \dots \nearrow \infty$ 。不难看出，由于  $\mathcal{B}_\phi$  中的基底在概念上与频率相关，使用其来定义平滑的函数将更加简单直观。同时，这也意味着选取的基底的数目可以视速度或者表征能力的需要来自由设定。具体来说，我们可以将  $T$  个函数截断来作为平滑函数的基底，这即是本章所研究的紧凑表示的基底：

$$f(\mathbf{x}) = \sum_{n=1}^T a_n \phi_n(\mathbf{x}). \quad (2.9)$$

为了进一步对所求函数进行限制，我们将所求参数  $\mathbf{a} = [a_1, a_2, \dots, a_T]^T$  视作服从正态

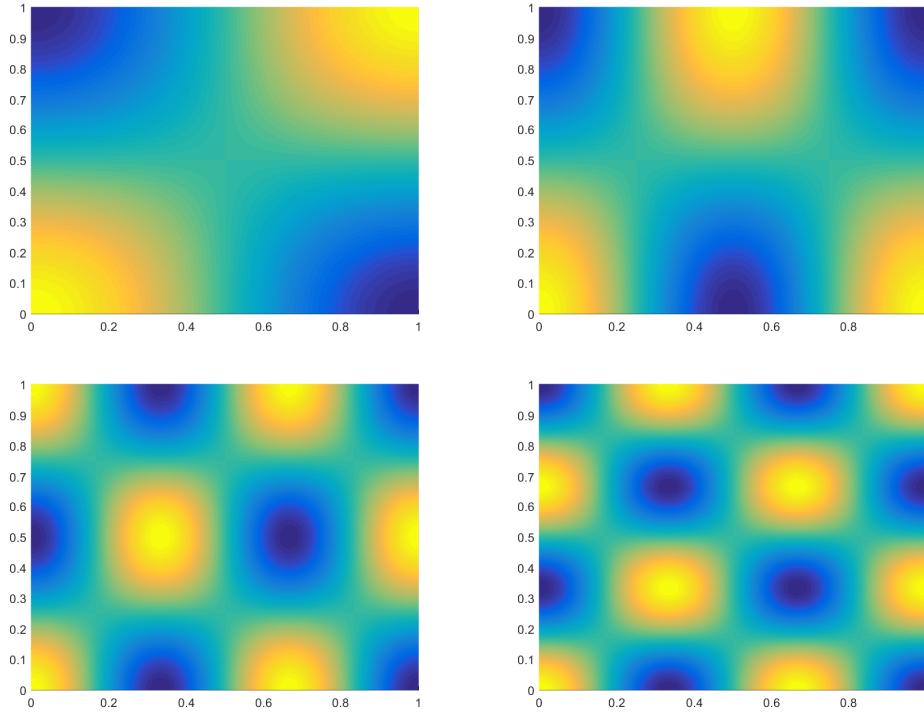


图 2.1 定义在  $[0, 1]^2$  中且边界条件为 *Neumann* 条件时的傅里叶基函数的可视化结果。按照从左到右，从上到下的顺序，频率是增加的，且相应函数中的  $\mathbf{j}$  分别为  $[1, 1]$ ,  $[2, 1]$ ,  $[3, 2]$  与  $[3, 3]$ 。颜色从蓝色到黄色的渐变意味着所代表的值由小增大。

分布  $a \sim \mathcal{N}(0, \mathbf{R})$  的随机变量，其中  $\mathbf{R} := \frac{1}{\lambda} \text{diag}(\omega_1, \omega_2, \dots, \omega_T)$ 。权重  $\omega_k$  由特征值来构造：

$$\omega_k = \mu_k^{-\frac{D}{2}}. \quad (2.10)$$

这种做法的数学背景来自于 Karhunen-Loeve expansion<sup>[146]</sup>，其更进一步地降低高频成分的权重，因此使得函数更加平滑。因此，通过紧凑表示 (2.9)，变分意义上的多元回归问题的最优解应有以下的有限维的形式：

$$\min_{\mathbf{a}} \sum_{n=1}^N V(\mathbf{y}, \mathbf{Fa}) + \mathbf{a}^T \mathbf{R}^{-1} \mathbf{a}, \quad (2.11)$$

其中  $\mathbf{F}$  是一个  $N \times T$  的矩阵，对于每个矩阵元素有  $\mathbf{F}_{mn} = \phi_n(\mathbf{x}_m)$ 。

相似的，在最简单的使用平方损失的情况下，最优解可以通过求解如下的线性系统来得到：

$$(\mathbf{F}^T \mathbf{F} + \mathbf{R}^{-1}) \mathbf{a} = \mathbf{F}^T \mathbf{y}. \quad (2.12)$$

显然，相对于 (2.5), (2.12) 的意味着更高的计算效率，其计算复杂度为  $O(N)$ 。

## 2.3 图像特征匹配中的平滑函数恢复

在本节中，我们主要关注图像特征匹配这一问题，作为我们所提出方法的一个重要应用。图像特征匹配可以视作偶从数据中学习多个多元平滑函数的问题，而对于精度没有太高的要求，因此我们所提出的近似算法较为适用于这一任务。首先，我们介绍一种贝叶斯的理论框架来对离群点进行建模，从而达到对离群点鲁棒的目的。

### 2.3.1 贝叶斯建模

在简化的匹配策略中，假设我们已经获得了一个初始匹配的集合  $S = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^N$ ，其中  $\mathbf{x}_n, \mathbf{y}_n \in \mathbb{R}^D$  表示  $D$  维的列向量，其中通常  $D = 2, 3$ 。为了找到内点集合  $\mathcal{I} \subseteq S$ ，我们的目的是从集合  $S$  中恢复一个潜在的变换，或者说一个向量场  $\mathbf{f} : \mathbb{R}^D \rightarrow \mathbb{R}^D$ ，使得对于  $(\mathbf{x}_n, \mathbf{y}_n) \in \mathcal{I}$  来说，有  $\mathbf{y}_n = \mathbf{f}(\mathbf{x}_n)$ 。这样以来，内点集可以通过检测其于所求的平滑函数的一致性来得出。

就这种方法而言，对于一个匹配  $(\mathbf{x}_n, \mathbf{y}_n)$ ，我们将其考虑为一个在  $\mathbf{x}_n$  位置的采样数据，同时我们假设这些采样的匹配是独立同分布的。对于内点来说，我们假设其对应的每个分量的噪声是服从高斯分布的，均值为 0 而标准差为  $\sigma$ 。对于离群点来说，由于采样值可以认为是在  $\mathbb{R}^D$  中随机分布的，因此我们将其假设为一种均匀分布，其概率密度为  $1/a$ ，其中  $a$  表示输出空间的体积。为了完成内点与离群点的区分任务，我们对于每个匹配都构造一个隐变量  $z_n \in \{0, 1\}$ ，其中  $z_n = 1$  代表匹配  $(\mathbf{x}_n, \mathbf{y}_n)$  是一个内点， $z_n = 0$  则反之。每一个隐变量  $z_n$  都应服从一个离散的概率分布，即  $p(z_n = 1) = \gamma$  且  $p(z_n = 0) = 1 - \gamma$ ，其中  $\gamma \in [0, 1]$ 。令  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T$  作为采样点的数据，并令  $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)^T$  作为采样值。在独立同分布的假设之下，总体的概率似然函数应具有如下的形式：

$$\begin{aligned} p(\mathbf{Y}|\mathbf{X}, \theta) &= \prod_{n=1}^N \sum_{z_n} p(\mathbf{y}_n, z_n | \mathbf{x}_n, \theta) \\ &= \prod_{n=1}^N \left( \frac{\gamma}{(2\pi\sigma^2)^{D/2}} e^{-\frac{\|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2}{2\sigma^2}} + \frac{1-\gamma}{a} \right), \end{aligned} \quad (2.13)$$

其中  $\theta = \{\mathbf{f}, \sigma, \gamma\}$  代表待求的未知参数。在贝叶斯的观点下，我们对于  $\theta$  也具有先验的信息，据此可以通过确立一种先验分布  $p(\theta)$  来进一步地对估计过程进行正则。

使用贝叶斯公式，则最大似然估计 (maximum a posteriori, MAP) 的过程可以表述为如下的形式：

$$\theta^* = \arg \max_{\theta} p(\theta | \mathbf{X}, \mathbf{Y}) = \arg \max_{\theta} p(\mathbf{Y} | \mathbf{X}, \theta) p(\theta). \quad (2.14)$$

这一问题可以等价地表示为寻找最小化如下能量函数的参数：

$$E(\theta) = -\ln p(\theta) - \prod_{n=1}^N \ln \sum_{z_n} p(\mathbf{y}_n, z_n | \mathbf{x}_n, \theta). \quad (2.15)$$

### 2.3.2 基于期望最大化算法的求解

为了解决 (2.15)，我们考虑广为人知的期望最大化算法 (Expectation-Maximization, EM)，其是一种高效率的解决具有隐变量的学习问题的通用方法。基本上来说，这一算法主要在两个步骤中迭代：Expectation step (E-step) 与 Maximization step (M-step)。考虑负对数后验函数 (2.15)，对于完整数据的对数似然期望为：

$$\begin{aligned} \mathcal{L}(\theta, \theta^{old}) = & -\frac{1}{2\sigma^2} \sum_{n=1}^N p_n \|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2 - \frac{D}{2} \ln \sigma^2 \sum_{n=1}^N p_n \\ & + \ln(1 - \gamma) \sum_{n=1}^N (1 - p_n) + \ln \gamma \sum_{n=1}^N p_n - \ln p(\theta), \end{aligned} \quad (2.16)$$

其中  $p_n = P(z_n = 1 | \mathbf{x}_n, \mathbf{y}_n, \theta^{old})$  代表  $z_n$  的后验概率。EM 算法中的 E-step 和 M-step 的详细说明如下：

*E-step*: 在这一步中，将根据现有的参数值来衡量后验的概率。由于独立同分布的假设，这一步可以分解为对于每个单独匹配的操作：

$$p_n = \frac{\gamma e^{-\frac{\|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2}{2\sigma^2}}}{\gamma e^{-\frac{\|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2}{2\sigma^2}} + (1 - \gamma) \frac{(2\pi\sigma^2)^{D/2}}{a}}. \quad (2.17)$$

*M-step*: 在这一步中，将对于现有的参数进行修正  $\theta^{new}$  using (2.16):  $\theta^{new} = \arg \max_{\theta} \mathcal{L}(\theta, \theta^{old})$ 。对于  $\sigma$  与  $\gamma$  来说，修正的规则可以由令 (2.16) 的导数为零的操作来导出。令  $\mathbf{P} = diag(p_1, p_2, \dots, p_N)$  为一个对角矩阵，那么我们有：

$$\sigma^2 = \frac{tr((\mathbf{Y} - \mathbf{T})^T \mathbf{P} (\mathbf{Y} - \mathbf{T}))}{D \cdot tr(\mathbf{P})}, \quad (2.18)$$

$$\gamma = \frac{tr(\mathbf{P})}{N}, \quad (2.19)$$

其中  $\mathbf{T} = (\mathbf{f}(\mathbf{x}_1), \mathbf{f}(\mathbf{x}_2), \dots, \mathbf{f}(\mathbf{x}_n))^T$ ,  $tr(\cdot)$  代表取迹的操作。

在下一步中，我们考虑更新向量场函数  $\mathbf{f}$  的方式。对于  $\sigma$  与  $\gamma$  来说，我们不对其做出更多的假设，或者等价来说假设其服从均匀分布，此时  $p(\theta)$  等价为  $p(\mathbf{f})$ 。将关于  $\mathbf{f}$  的项抽离出来，我们可以得出如下的泛函的形式：

$$\varepsilon(\mathbf{f}) = \frac{1}{2\sigma^2} \sum_{n=1}^N p_n \|\mathbf{y}_n - \mathbf{f}(\mathbf{x}_n)\|^2 - \ln p(\mathbf{f}). \quad (2.20)$$

这是对于向量场学习来说关键的一步，我们使用所提出的紧凑表示的方法来作为这一步的基础。

注意到 (2.20) 实际上要求计算一种以向量为值的函数  $\mathbf{f}$ 。一般来说，这与通常所说的多任务或者多输出学习相关，其目标是用不同人物或不同输出的相关性来提升总体的学习效果。在我们的框架之下，不同的任务对应着由  $\mathbf{x}_n$  对应  $\mathbf{y}_n$  的每个分量的学习。我们将对于第  $i$  个分量的映射记为  $\mathbf{f}_i : \mathbb{R}^D \rightarrow \mathbb{R}$ 。然而，在先前的图像特征匹配的文献中，一般隐式地对  $\mathbf{f}_i$  做出相互独立的假设 [48, 69]。这是由于在图像特征匹配问题中，每个映射函数之间的关联尚不明确，并且做出独立性的假设可以使算法的速度得到提升。因此在本章中，我们也沿用这一假设，并使用多元函数回归的思路来解决这一问题。

我们将每一个映射  $\mathbf{f}_i$  建模为潜在的平滑函数的紧凑表示的表达形式：

$$\mathbf{f}_i(\mathbf{x}) = \sum_{n=1}^T a_n^i \phi_n(\mathbf{x}). \quad (2.21)$$

正如前文所说，我们将  $\mathbf{a}_i = [a_1^i, a_2^i, \dots, a_T^i]^T$  视作随机变量，此时  $-\ln p(\mathbf{f}_i)$  转化为  $\mathbf{a}_i^T \mathbf{R}^{-1} \mathbf{a}_i$ 。最终 (2.20) 将可以被分解为解决如下问题：

$$\begin{aligned} & \min_{\mathbf{f}_i \in \text{span}(\mathcal{B}_T)} \frac{1}{2\sigma^2} \sum_{n=1}^N p_n \|y_n^i - \mathbf{f}_i(\mathbf{x}_n)\|^2 - \ln p(\mathbf{f}_i) \\ &= \frac{1}{2\sigma^2} \|\mathbf{P}^{1/2}(\mathbf{Y}_i - \mathbf{F}\mathbf{a}_i)\|^2 + \mathbf{a}_i^T \mathbf{R}^{-1} \mathbf{a}_i, \end{aligned} \quad (2.22)$$

其中  $y_n^i$  代表  $\mathbf{y}_n$  的第  $i$  个分量， $\mathbf{Y}_i$  表示  $\mathbf{Y}$  的第  $i$  列， $\mathbf{F}$  表示  $N \times T$  的矩阵，其每个

**Algorithm 1** Compact Representation Consensus

---

输入: 初始匹配集合  $\mathcal{S}$ , 基函数集合  $\mathcal{B}_T$ , 常数  $\lambda$  和内点判断门限  $\tau$

输出: 内点集合  $\mathcal{I}$

- 1: 初始化  $\gamma, a, \mathbf{T} = \mathbf{0}_{N \times 2}, \mathbf{P} = \mathbf{I}_{N \times N}$ ;
  - 2: 利用 (2.18) 初始化  $\sigma^2$ ;
  - 3: 利用  $\mathcal{B}_T$  构造矩阵  $\mathbf{F}$ ;
  - 4: **while**  $\mathcal{L}$  未收敛 **do**
  - 5:   **E-step:**
  - 6:     利用 (2.17) 更新  $\mathbf{P}$ ;
  - 7:   **M-step:**
  - 8:     对于  $i = 1, 2, \dots, D$  利用 (2.23) 更新  $\mathbf{a}_i$ ;
  - 9:     利用 (2.21) 更新  $\mathbf{T}$ ;
  - 10:    利用 (2.18) 以及 (2.19) 更新  $\sigma^2$  and  $\gamma$ ;
  - 11: **end while**
  - 12: 利用 (2.24) 决定内点集合。
- 

元素的意义为  $\mathbf{F}_{mn} = \phi_n(\mathbf{x}_m)$ 。

可以看出 (2.22) 是一个凸的问题, 并且最优解可以通过解决如下的线性系统来得出:

$$(\mathbf{F}^T \mathbf{P} \mathbf{F} + \sigma^2 \mathbf{R}^{-1}) \mathbf{a}_i = \mathbf{F}^T \mathbf{P} \mathbf{y}_i. \quad (2.23)$$

在 EM 算法收敛之后, 我们可以通过使用一个预定义的门限  $\tau$  来找到内点:

$$\mathcal{I} = \{(\mathbf{x}_n, \mathbf{y}_n) : p_n > \tau, n \in \mathbb{N}_N\}. \quad (2.24)$$

我们将本章所提出的算法记为 Compact Representation Consensus (CRC), 并且将所提算法的细节总结于算法1中。

### 2.3.3 复杂度分析以及实现细节

在本章所提出的 CRC 算法中, 主要的计算消耗来源于求解线性系统 (2.23), 由于  $\mathbf{F}$  是  $N \times T$  的且  $N \gg T$ , 其复杂度为  $O(N)$ 。因此该算法总体的时间复杂度可写为  $O(cN)$ , 其中  $c$  代表 EM 算法的迭代次数。由于存储  $\mathbf{F}$  的需要, 该算法的空间复杂度是  $O(N)$  的。即便我们所提出的 CRC 算法在理论上与 SparseVFC 具有相同的时间复杂度, 由于后者的数值稳定性较差, 在实际中 CRC 要远远快于 SparseVFC。

在 CRC 中,  $T, \lambda, \gamma$  以及  $\tau$  等参数需要预先设置。其中  $T$  代表着所取的基函数的数目, 我们根据章节 2.4.1 的结果将其设置为 15。参数  $\lambda$  用于控制对于  $\mathbf{a}$  的正则化的程度, 我们经验上将其设置为 1。参数  $\gamma$  反映的是对于内点比率的估计, 我们经验

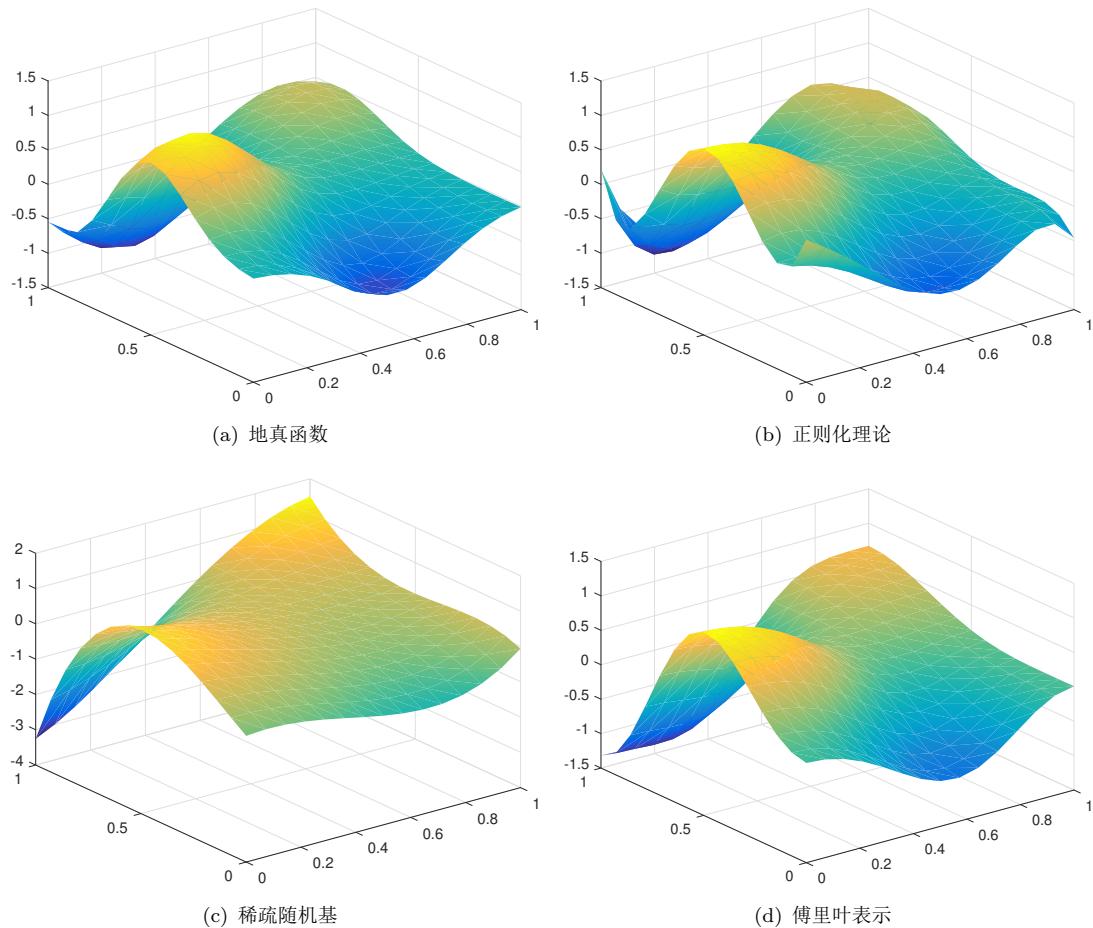


图 2.2 一个由受离群点污染的稀疏采样数据中恢复平滑函数的例子，平滑函数被可视化为三维空间中的一个表面。

上将其设置为 0.95。参数  $\tau$  是用于在最终决定内点的门限，我们将其设置为 0.75。数据在算法处理前均经过归一化使其分布在  $[0, 1]^2$  中。

## 2.4 实验结果

在本节中，我们使用数值实验的结果来说明所提方法的优越性。首先，我们使用人工合成的数据，在存在离群点的多元回归任务中来比较所提方法与经典正则化理论以及其变体。然后，作为所提 CRC 算法的应用，我们考虑图像特征匹配的应用，并且使用真实的图像数据做实验。此外，我们也在点集配准的任务上说明了所提方法的优越性。所有的实验均在配置为 2.3GHz Intel Core CPU，16GB 内存的笔记本电脑中使用 Matlab 代码来实现。

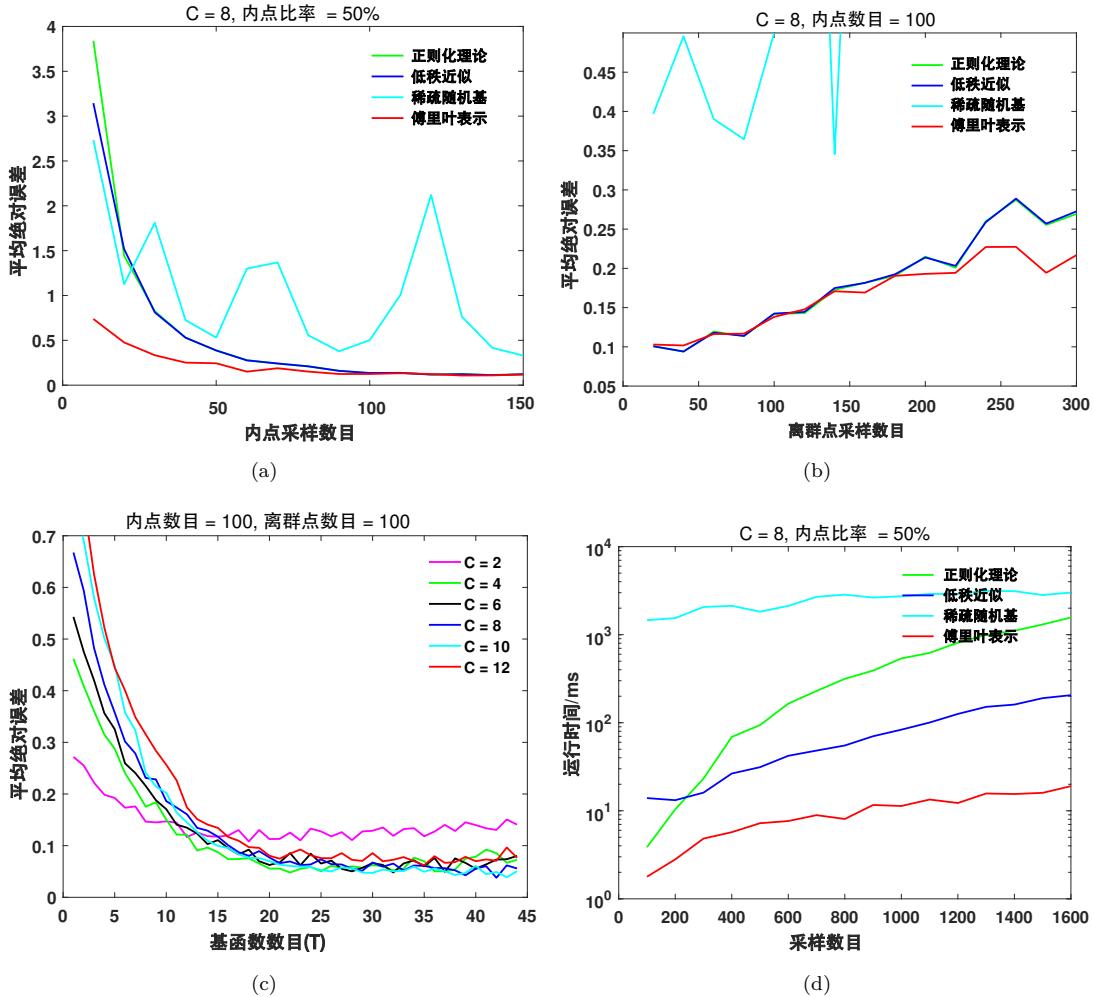


图 2.3 所提方法与对比方法在合成数据中的指标对比，其中每个数据点都代表着在 50 个随机产生的平滑表面上的结果的均值。

#### 2.4.1 合成数据中的多元回归

从一种广义的角度来看，我们所提出的算法可以视作经典的正则化理论的一个替代，来从稀疏的数据中学习平滑的函数。我们使用合成数据来创造一个可以衡量所提方法与正则化理论的理想环境。我们将地真的平滑函数构造为  $C$  个高斯函数的混合，其均具有相同的协方差  $0.04\mathbf{I}$ ，而均值位于在  $[0, 1]^2$  中的随机点，幅值为限制在  $[0, 1]$  中的随机值。利用这种方式，所构造的函数保证是平滑的，且复杂性随着  $C$  的增高而增高。一个  $C = 8$  时的可视化的结果作为 3D 空间的表面展示在图 2.2 中。为了构造合成的数据，我们从这个函数中进行随机的采样并且人工添加一定的离群点。最终的性能使用所恢复得到的函数与地真的函数之间的差异性来衡量，其中差异性定义为在均匀分布在  $[0, 1]^2$  中的采样点所定义的平均绝对误差 (mean absolute error, MAE)。

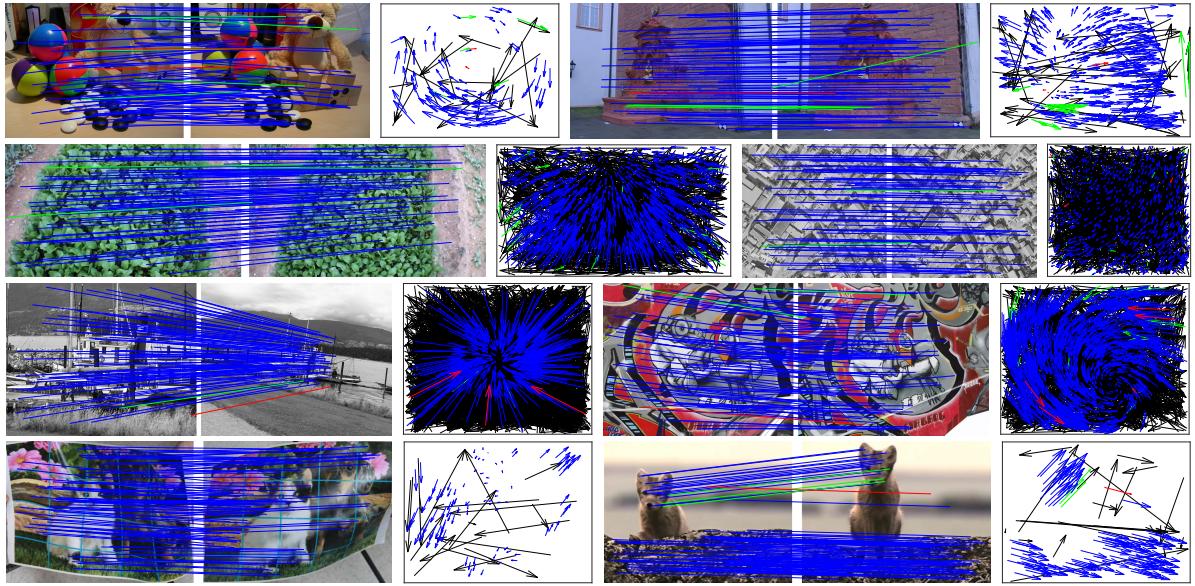


图 2.4 所提方法在 8 对来自于 DAISY, CRS, 以及 VGG+ 数据集中的代表性的图像对中的特征匹配结果。在运动场的可视化结果中，箭头的头部与尾部分别代表特征点在两幅图像中的坐标（蓝色 = 真阳性，黑色 = 真阴性，绿色 = 假阴性，红色 = 假阳性）。考虑到视觉上的效果，在所有的图像对中，我们随机选取了 100 个随机的匹配显示，真阴性的结果则不被显示。

我们选取三种竞争性的方法来做比较，包括经典的正则化理论<sup>[140, 141]</sup> 以及两种近似的方法，即基于低秩的近似<sup>[69]</sup> 与基于稀疏随机基底的近似<sup>[142]</sup>。对于低秩的近似而言，我们使用核矩阵的前 30 个最大的特征值以及其所对应的特征向量。对于稀疏随机基底而言，我们按照原论文<sup>[142]</sup> 的推荐，随机选取 80 个基底。所有的方法都使用相同的贝叶斯框架来处理离群点。图 2.3 中展示了所有的测试结果，其中每个数据点均是 50 个随机产生的合成数据的统计结果。图 2.3(a)与图 2.3(b)展示了所有方法随内点的采样数目以及离群点的采样数目的性能变化。可以看出，在我们的设置中，100 个样本足以恢复平滑功能。基于紧凑傅里叶基表示的方法具有明显的优势，因为它需要较少的内部样本进行回归，并且对离群点更具鲁棒性。正则化理论和低秩近似具有相似的性能，然而，稀疏随机基的策略则无法给出准确的结果。如图 2.2 所示，这一方法产生了过度光滑的表面。在图 2.3(c)中，我们展示了所提的方法相对于基函数的数目  $T$  的性能。可以看到，通常 15 个基函数已经足够用于光滑函数的回归，这清楚地表明了 Fourier 表示的紧凑性。图 2.3(d)显示了运行时间的统计信息。可以观察到，我们的方法具有最低的计算复杂度，并且可以比其他方法快数个数量级。而稀疏随机基方法由于其数值不稳定性而效率最低。

### 2.4.2 鲁棒图像匹配

在本节中，我们将重点讨论所提方法在面对真实图像的特征匹配中的应用。为了解决图像特征匹配中所常见的离群点问题，所提算法的目的是区分从受离群点污染的数据中区分出内点。如上所述，我们的方法与 VFC<sup>[69]</sup> 及其变体 FastVFC<sup>[69]</sup> 和 SparseVFC<sup>[142]</sup> 具有相同的原理，即通过拟合一个平滑函数以剔除异常值。我们的评估采用了图像特征匹配文献中广泛使用的三个数据集：

**DAISY**<sup>[81]</sup>: 该数据集由宽基线图像对与地真深度图组成，包括两个短图像序列和几个单独的图像对。我们总共创建了 52 对图像进行评估。由成像场景决定，这些匹配通过极线约束进行关联。

**CRS**: 该数据集由文章<sup>[147]</sup> 中使用的服从刚性变换以及投影变换的遥感数据集组成，包含 161 对遥感图像，包括无人机拍摄的彩色图像、合成孔径雷达拍摄的图像、全色航空照片和彩色红外航空照片。图像对之间的关系可用单应变换来描述。

**VGG+**: 该数据集中的图像对是 VFC<sup>[69]</sup> 中使用的数据集合，包括单应和非刚性变换相关的图像对和宽基线图像对。我们使用 SIFT 为每一对建立初始匹配的集合，并手动标注内点作为地真值。由于较低的 inlier 比率，这个数据集是相当有挑战性的。

我们使用所采用的数据集中的一些具有代表性的图像特征匹配的典型例子来测试所提出的 CRC，如图 2.4 所示。这些数据涵盖了不同类型的特征匹配场景，包括单应相关的图像对、基本矩阵和非刚性变换。对于每组结果，左侧图像对显示直观的匹配结果，右侧运动场提供初始匹配集合中每个匹配判定的正确性的结果。可以看出，无论图像对之间具有怎样的复杂关系，甚至在存在高离群点比率的情况下，我们的 CRC 算法总能产生令人满意的结果。

我们选取了 9 种最新的鲁棒特征匹配方法进行比较，即 RANSAC<sup>[85]</sup>、ICF<sup>[68]</sup>、SM<sup>[53]</sup>、VFC<sup>[69]</sup>、FastVFC<sup>[69]</sup>、SparseVFC<sup>[142]</sup>、LPM<sup>[81]</sup>、GMS<sup>[82]</sup>，以及最近的深度学习算法 ACNe<sup>[135]</sup>。所有算法都是基于公开的代码实现的，并且参数经过优化调整之后固定。定量的比较结果如图 2.5 所示。在第一行中，我们提供了三个数据集上初始内点比率的累积分布。我们可以看到 VGG+ 数据集是最具挑战性的数据集，具有相当低的内点率。第二、第三、第四和第五行分别总结了三个数据集的精度、召回率、F-分数和运行时间的统计结果。经典的 RANSAC 算法在这三个数据集中有不同的性能，这是由于其随机的性质。这种随机性在存在大量的离群点的情况下是非常不利的。

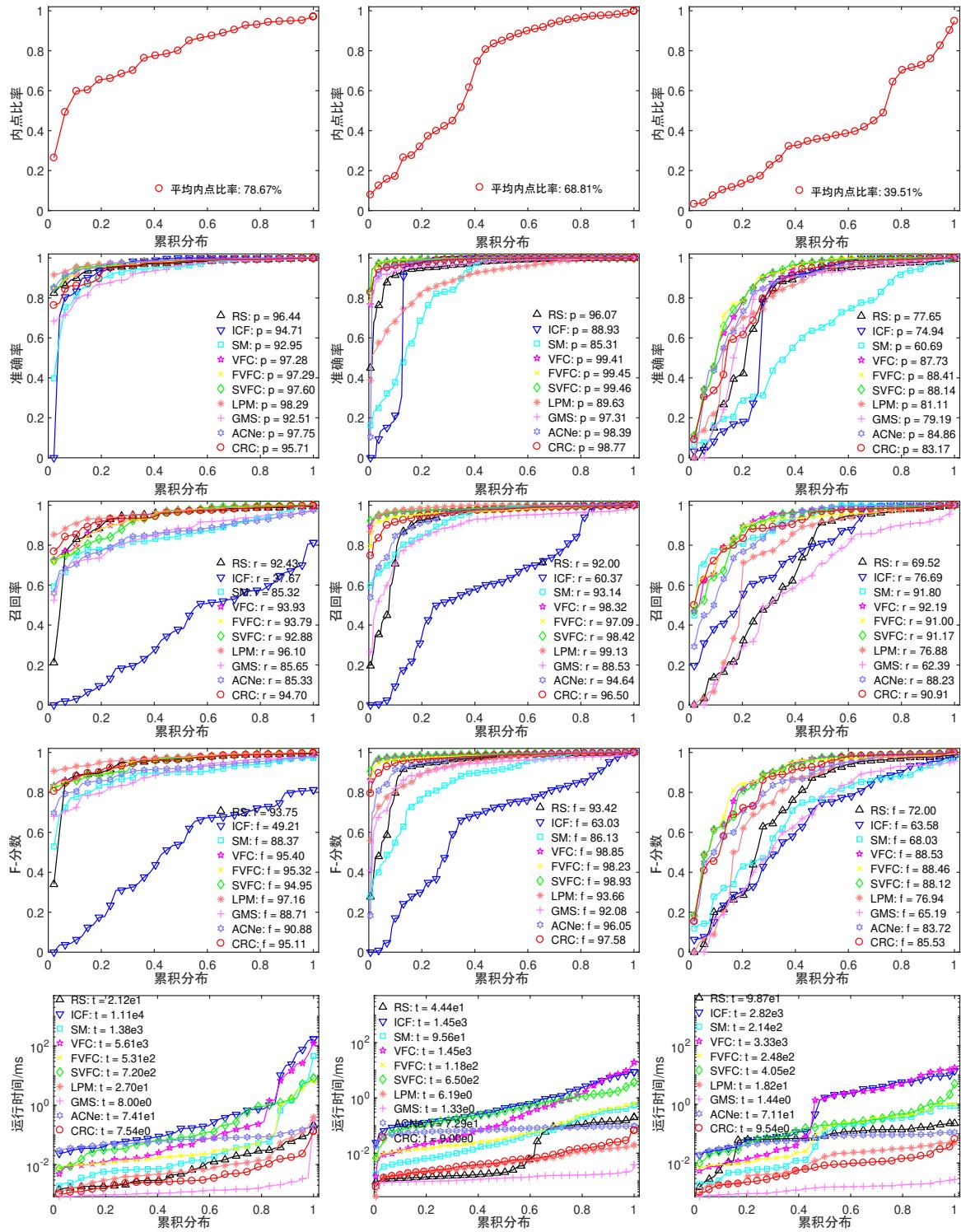


图 2.5 所提出的 CRC 方法和九个竞争算法在三个数据集上的图像特征匹配的定量比较结果, (从左到右) DAISY, CRS, VGG+. 从上到下, 给出了初始内点比率、准确率、召回率、F-分数和运行时间等的累积分布。每个方法的平均统计数据在图例中给出。FastVFC 和 SparseVFC 分别缩写为 FVFC 和 SVFC。

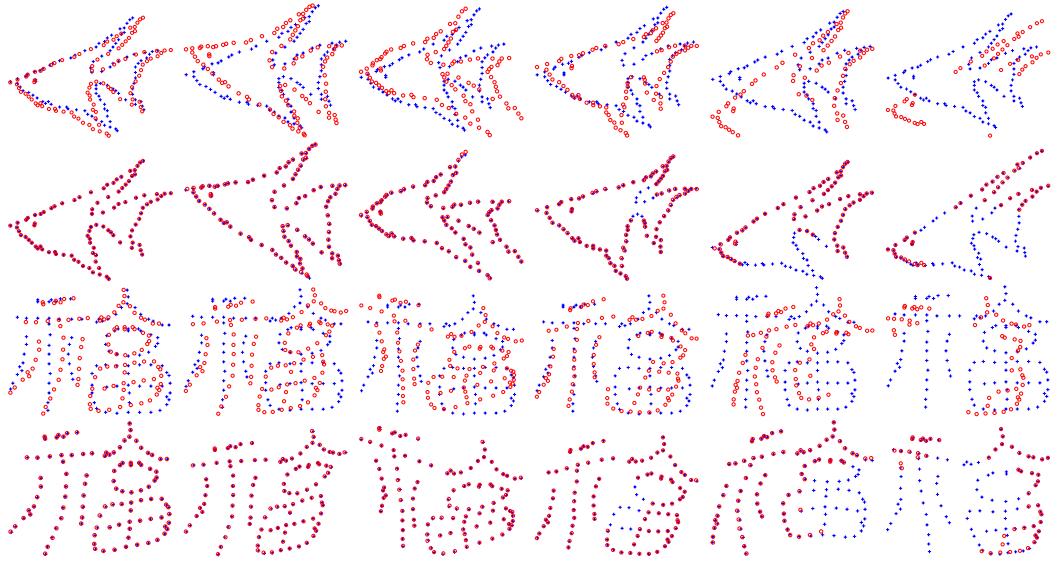


图 2.6 我们所提出的方法在 *fish* 和 *Chinese character* 的数据上的配准结果。配准任务的目标是将模型点集（蓝色加号）与目标点集（红色圆圈）对齐。在前三列中，点集受到不同程度的变形，在后三列中，点集受到不同程度的遮挡。第一行和第三行是初始点集，第二行和第四行表示配准结果。

ICF 算法也试图恢复平滑函数以区分内点与离群点，但它对于参数设置非常敏感，性能因而较差。SM 算法利用匹配之间两两之间的相关性获得可信度高的匹配，具有较好的通用性，但精度较低，因此其性能比较一般。VFC 家族，即 VFC、FastVFC 和 SparseVFC，具有非常接近的性能，并且尽管它们的运行时间很高，但其性能优于所有其他方法。LPM 算法和 GMS 算法效率更高，但它们的性能还不能与 VFC 系列相媲美。相比之下，我们的 CRC 在精确度和计算复杂度上都非常有竞争力。由于算法的原理类似，它的性能与 VFC 系列相差不多，但要快一个数量级以上。与 SparseVFC 不同的是，在我们的实验中，我们的 CRC 算法实现了真正的线性复杂度，运行时间一般小于 10 毫秒，这证明了我们方法的优越性。此外，我们的方法在准确性和效率方面都显示出与当前最先进的深度学习技术相比的明显优势。

#### 2.4.3 非刚性点集配准

点集配准是计算机视觉中的一个经典问题，其目的是通过恢复点集之间的潜在的变换使一个点集与另一个点集对齐。对于非刚性情况，学术界最流行的解决方案通常是基于正则化理论。与图像特征匹配类似，在实际应用中，这种变换通常被简化为两个独立的多元函数的组合。因此，我们可以应用我们的 CRC 来鲁棒地恢复平滑的变换函数并对于点集进行配准。

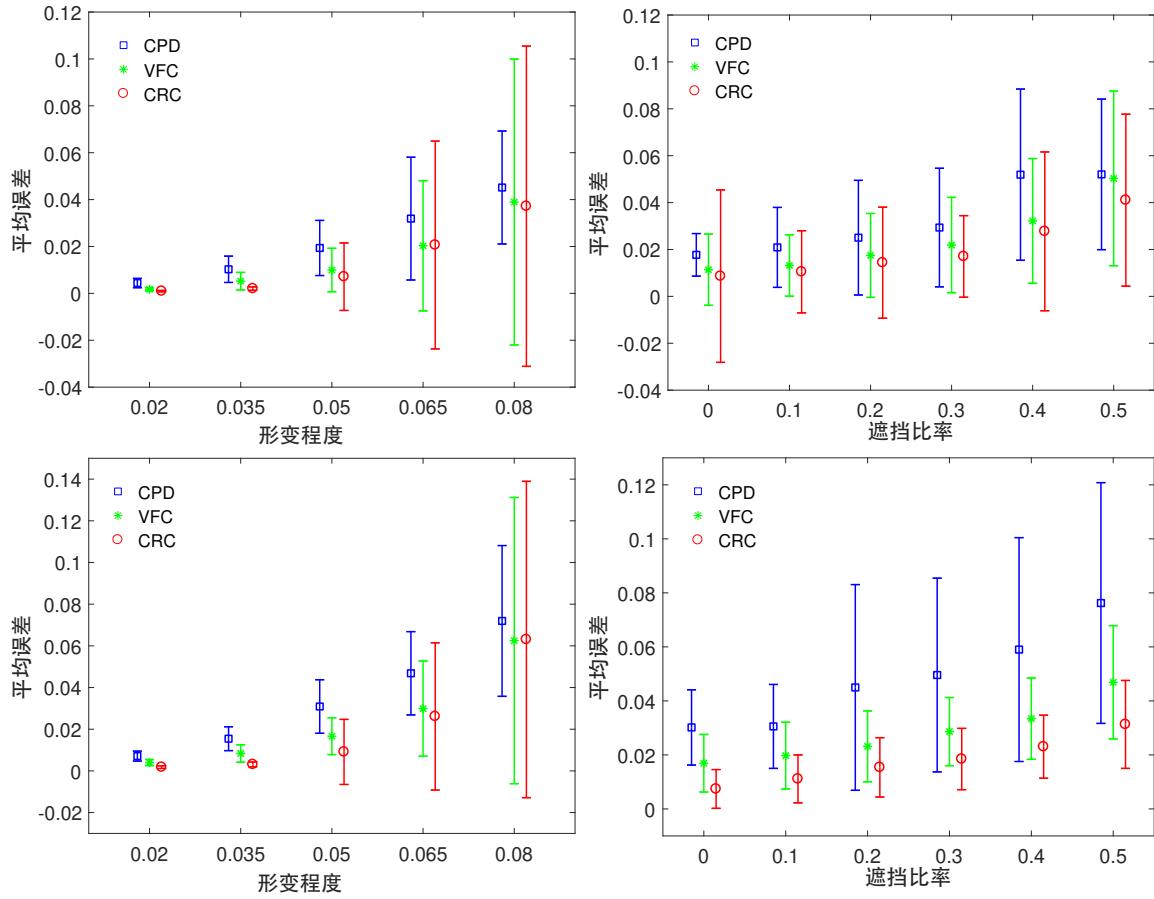


图 2.7 CPD、VFC 和所提出的 CRC 方法在 *fish* (上) 和 *character* (下) 数据上的定量比较。图中的误差线统计为在 100 个实例中的配准误差均值和标准差。

与 VFC 的做法相同，我们的算法遵循类似 ICP 的过程。在每次迭代中，我们使用 ShapeContext 建立点集之间的对应关系，然后使用 CRC 对变换函数进行鲁棒的估计。我们经验性地使用经过 10 次迭代的变换后的点集作为最终的配准结果。实验中使用了 *fish* 和 *Chinese character* 数据，我们的方法对这两种形状数据的定性结果如图 2.6 所示。我们将结果展示成两行：第一行是初始点集，第二行是对应的配准结果。从结果可以看出，即使在存在严重退化，如非刚性变形和遮挡的情况下，我们的方法仍旧能够产生准确的配准结果。

作为比较，我们采用了著名的 CPD 方法<sup>[48]</sup> 和 VFC 方法<sup>[69]</sup>，这两种方法本质上均是基于正则化理论发展起来的。两个点集之间的配准误差由变换后的模型点集与目标点集的地真值匹配的平均欧氏距离来表征。对于每个干扰的级别，我们计算在所有 100 个实例上的配准误差的平均值和标准差，以得出总体的统计数据。定量结果如图 2.7 所示。从结果中，我们可以看到，由于 CPD 在配准过程中没有使用局部特征，

它的效果通常差于 VFC 和我们所提出的 CRC 方法。VFC 和我们的 CRC 的主要区别在于对于变换函数的建模。可以看出，在我们的紧凑表示的方法之下，配准误差得到了一致的降低，特别是在遮挡的情况下。这证明了我们的方法作为以经典正则化理论为基础的方法的一种有效替代的优越性。

## 2.5 本章小结

在本章中，我们提出了一种由受离群点污染的数据中恢复潜在的多元平滑函数的方法，并在实验中论证了所提方法的有效性。相较于传统的正则化理论，本章所提出的方法采用紧凑的傅里叶基表示来建模平滑函数，因而所需的参数更少并且计算效率得到巨大的提升。将这种建模方式与一种对离群点显式建模的贝叶斯框架相结合，我们提出了一种基于 EM 算法的快速求解方法，并赋予了所提方法对离群点的鲁棒性。我们在合成数据中的多元回归任务，鲁棒特征匹配任务以及点集配准任务中进行了大量实验并与处于领先水平的现有算法进行对比，结果显示本章所提出方法的优越性。

### 3 基于概率图模型的误匹配去除算法研究

#### 3.1 引言

在简化匹配的策略中，图像特征匹配的问题转化为在初始匹配集中寻找内点的过程。就这一点而言，在上一章，我们介绍并提出了一种新的基于平滑函数恢复的方法。这一类方法的有着全局性的特点，即在其将内点与离群点区分开的过程中，同时建立了一个全局性的平滑函数。与全局性算法相对，局部性的算法在图像特征匹配中也有着重要的作用。例如在图像中的形变非常剧烈，图像中存在较大的深度差，或者图像中存在着多个相对运动的物体的情况下，局部性的算法通常可以发挥更好的效果，这是由于图像场景本身所具有的性质决定的。与全局性的算法不同，局部性的算法对问题的建模一般更加简单直接。在简化匹配的策略中，局部性的鲁棒算法一般直接将问题建模为从初始匹配中去除错误匹配的过程，不涉及对于运动场进行函数回归的问题。

在广义的机器学习领域中，基于图模型的算法具有很高的灵活性，在多种任务中都发挥着重要的作用。而图模型能发挥灵活性的一个重要的原因正是其局部性的特点。具体来说，最简单的二阶图模型仅仅考虑数据之中两两之间的相关性，而其目标函数是所有这种局部的相关性的累加作用的结果。当然，由这一点来说，局部性的累加也构成了某种意义上的全局性，因此，也可以说图模型达到了全局性与局部性的一种恰到好处的折中。在图像特征匹配领域的现有的文献中，局部性的算法往往仅仅考虑相邻的匹配对于误匹配去除的影响，因此完全失去了全局性，从而造成算法的较差表现。

在本章中，我们采取一种基于概率图模型<sup>[148]</sup>(probabilistic graphical model, PGM)的框架来对于误匹配去除问题进行建模。概率图模型总体来说思想与基本的图模型类似，但不同的是其建立在更加严谨的概率框架之下，因此我们选取概率图模型作为所提出算法的基础。为了尽可能提高计算效率，我们仅仅考虑二阶概率图模型，即仅仅对匹配两两之间的相关性进行建模。为了解决图像特征匹配的问题，我们提出了一种基于运动一致性的相关性建模方式。最终，我们将概率图模型下的求解问题转化为一种特殊的整数二次规划问题，并提出了一种基于 Frank-Wolfe 方法<sup>[149, 150]</sup>的快速求解

算法。实验验证了所提算法的有效性。

## 3.2 基于概率图模型的误匹配去除算法

### 3.2.1 问题建模

假设我们从给定的图像中获得了一组初始的匹配特征对应关系  $S = \{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^N$ , 其中  $\mathbf{u}_i$  和  $\mathbf{v}_i$  是分别表示来自两个不同图像的特征点的空间位置的二维列向量。我们的目标是从整初始匹配集合  $S$  中识别真正的匹配集合  $\mathcal{I}$ 。

令设  $G = (V, E)$  为一个无向图 (也称为 Markov 网络), 节点集合为  $V$ , 边的集合为  $E$ , 其中  $V$  中的每个节点  $v_i \in V$  表示一个  $S$  中的匹配  $(\mathbf{u}_i, \mathbf{v}_i)$ , 每条边  $e_k \in E$  均表示着两个节点之间的交互。节点的势函数和边的势函数是描述定义在  $G$  上的联合分布的必要条件。对于每个节点  $v_i$ , 节点的势函数  $\psi(v_i)$  对其状态的先验分布进行编码。对于每个边  $e_k$ , 边的势函数  $\psi(e_k)$  编码两个相连接的节点之间的交互, 这是 PGM 的核心。我们将在章节 3.2.3 节中讨论势的定义。在这里, 我们假设已经指定了势函数, 然后进行解码的过程来推断节点状态并由此解决误匹配消除的问题。

对于概率图模型来说, 解码的过程是这样的。作为状态指示符, 令  $x_i$  作为节点  $v_i$  所索引的随机变量, 其中  $x_i = 1$  表示内点,  $x_i = 0$  则表示离群点。从技术上讲, 解码是为了找到使所有随机变量的联合概率分布函数最大化的最佳配置, 如下所示:

$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in \{0,1\}^N} p(x_1, x_2, \dots, x_N), \quad (3.1)$$

其中  $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$  是解空间中的一个向量,  $p$  指代联合概率分布。在仅考虑二阶的势函数的基础下,  $p$  可以定义为如下的形式:

$$p(x_1, x_2, \dots, x_N) = \frac{1}{Z} \prod_{v_i \in V} \psi(v_i) \prod_{e_k \in E} \psi(e_k), \quad (3.2)$$

其中  $Z$  是一个归一化常数。显然, 在解码过程完成之后, 内点的集合  $\mathcal{I}$  可以直接经由最优解  $\mathbf{x}^*$  的状态来决定。关于求解的过程我们将在章节 3.2.4 中解释。

为了在图像特征匹配的背景下建立 PGM 模型, 我们首先引入一种新的运动一致性度量来描述两个匹配的相容性, 然后讨论如何建立一个包含局部运动一致性先验知

识的交互模型。最后，我们将解码过程转化为一个 IQP 问题，并给出一个高效率的求解算法。

### 3.2.2 运动一致性度量

在图像特征匹配中，运动场可以理解为特征点的坐标在两张图像上的不同位置所定义的一种偏移，而运动一致性的概念也在诸多的现有文章中得到了研究<sup>[69, 82, 83]</sup>。我们从一个新的角度，在运动/矢量场的背景下研究了它，并导出了一个空间一致性度量来描述两个匹配的相容性。

#### 3.2.2.1 向量场表示

我们使用匹配的向量场表示来作为对于运动建模的方式。向量场可以定义为对任意一个点  $\mathbf{p} \in R^D$ ，其具有一个  $\mathbf{q} \in R^P$  的向量作为映射。因此，对于一个匹配  $(\mathbf{u}_i, \mathbf{v}_i)$  来说，其可以转化为一个向量场的采样形式： $\mathbf{p}_i = \mathbf{u}_i$ ,  $\mathbf{q}_i = \mathbf{v}_i - \mathbf{u}_i$ ，其中  $\mathbf{p}_i$  代表特征点的空间位置，而  $\mathbf{q}_i$  代表相应的运动向量。在这种向量场的表示形式之下，初始匹配的集合  $S = \{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^N$  可以被转化为一个向量场的采样样本集合  $F = \{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^N$ 。实际上， $(\mathbf{p}_i, \mathbf{q}_i)$  可以被看做是节点  $v_i \in V$  的一种属性。

向量场的表示允许我们以显式的方式考虑运动一致性。直观地说，我们可以把  $F$  看作一个受到污染的采样集合，其中存在一个由内点所决定的平滑向量场。因此，我们可以通过找到与平滑向量场一致的样本来识别真实的匹配对应关系<sup>[69]</sup>。值得注意的是，与文章<sup>[69]</sup> 不同，我们的方法还可以处理由独立运动所带来的分段光滑场。

#### 3.2.2.2 基于差商的度量

在图像特征匹配中，特征点的维度为 2，因此我们在向量场表示中有  $P = D = 2$ （也可以推广到更高维度）。这种向量场可以有一个函数  $\mathbf{f}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  来定义：

$$\mathbf{f}(p_1, p_2) = (f_1(p_1, p_2), f_2(p_1, p_2)), \quad (3.3)$$

其中  $f_i : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  是  $\mathbf{f}$  的第  $i$  个分量。向量场  $\mathbf{f}$  的导数具有如下的形式（即 Jacobian 矩阵）：

$$D_{\mathbf{p}} \mathbf{f} = \begin{bmatrix} \partial f_1 / \partial p_1 & \partial f_1 / \partial p_2 \\ \partial f_2 / \partial p_1 & \partial f_2 / \partial p_2 \end{bmatrix}. \quad (3.4)$$

由于特征点是离散的并且我们只有潜在的向量场  $\mathbf{f}$  的一些采样，我们转而考虑向量场的差商。考虑任意两个  $F$  中的采样数据  $(\mathbf{p}_i, \mathbf{q}_i)$  与  $(\mathbf{p}_j, \mathbf{q}_j)$ ，其中  $\mathbf{p}_i = (p_{1i}, p_{2i})$ ,  $\mathbf{q}_i = (q_{1i}, q_{2i})$ ,  $\mathbf{p}_j = (p_{1j}, p_{2j})$  且  $\mathbf{q}_j = (q_{1j}, q_{2j})$ 。那么作为向量场的导数的一阶近似，差商可以定义为：

$$DQ_{ij} = \begin{bmatrix} \frac{q_{1j} - q_{1i}}{p_{1j} - p_{1i}} & \frac{q_{1j} - q_{1i}}{p_{2j} - p_{2i}} \\ \frac{q_{2j} - q_{2i}}{p_{1j} - p_{1i}} & \frac{q_{2j} - q_{2i}}{p_{2j} - p_{2i}} \end{bmatrix}. \quad (3.5)$$

差商的概念可以解释为向量场由  $\mathbf{p}_i$  到  $\mathbf{p}_j$  的变化率。我们使用  $\|DQ_{ij}\|_F$  来衡量  $\mathbf{p}_i$  与  $\mathbf{p}_j$  至今的空间连续性，其中  $\|\cdot\|_F$  表示 Frobenius 范数。显然，假如两个采样点均与一个潜在的平滑向量场一致，那么  $\|DQ_{ij}\|_F$  将具有较小的值，反之  $\|DQ_{ij}\|_F$  则将具有较大的值。具体来说， $\|DQ_{ij}\|_F^2$  的形式如下：

$$\begin{aligned} \|DQ_{ij}\|_F^2 &= [(q_{1j} - q_{1i})^2 + (q_{2j} - q_{2i})^2] \\ &\times \left[ \frac{1}{(p_{1j} - p_{1i})^2} + \frac{1}{(p_{2j} - p_{2i})^2} \right]. \end{aligned} \quad (3.6)$$

在预先定义的图  $G$  中，考虑连接  $v_i$  与  $v_j$  的一条边。我们可以看到  $\|DQ_{ij}\|_F$  是一个自然的  $v_i$  与  $v_j$  之间相容性的度量方式。为使这种度量更加稳定，我们定义引入一个常数  $s$  并定义  $c_{ij}$  作为  $v_i$  与  $v_j$  之间运动一致性的最终度量，其形式为：

$$\begin{aligned} c_{ij}^2 &= [(q_{1j} - q_{1i})^2 + (q_{2j} - q_{2i})^2] \\ &\times \left[ \frac{1}{(p_{1j} - p_{1i})^2 + s} + \frac{1}{(p_{2j} - p_{2i})^2 + s} \right]. \end{aligned} \quad (3.7)$$

为了对于运动一致性给出一个直观的解释，我们在图. 3.1中提供了一个概念性的解释。

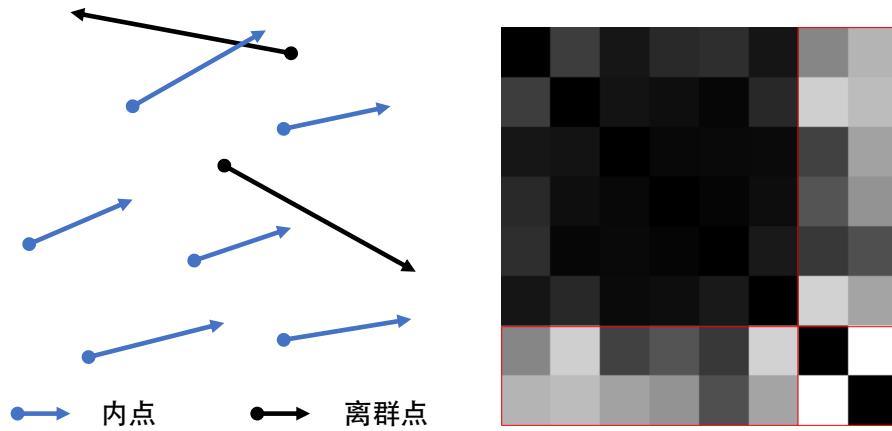


图 3.1 运动一致性度量的图示。左图是一组匹配的向量表示，右图中的矩阵对这些匹配之间的两两之间的运动一致性的度量值进行编码。在右图的每个元素中，白色表示较大的值，黑色表示相反。红线圈起的是涉及离群点的匹配元素，这些离群点通常对应着比内点更大的值。

### 3.2.3 局部运动一致性先验

利用等式 (3.7) 中所描述的节点间兼容性的度量，我们能够在图模型中施加运动一致性以消除错误的匹配。然而，这种先验知识通常会带来全局的约束。在图像中存在严重变形或独立运动的情况下，运动可能只是分段平滑的，此时全局约束是不合适的。为了解决这个问题，我们将运动一致性推广为局部运动一致性。

#### 3.2.3.1 局部交互结构

对于 PGM 来说，边表示节点至今的交互作用，而完全连通图则表示全局的约束或先验。为了施加局部的先验，我们只允许节点之间的局部交互，这意味着对于  $V$  中的每个节点，只有相邻节点可以与之连接。我们将这一点解释为图模型  $G$  的局部相互作用结构。

我们根据向量场采样的空间位置构造每个节点的邻域。特别地，考虑两个节点  $v_i$  和  $v_j$  之间的欧氏距离度量：

$$D(v_i, v_j) = \|\mathbf{p}_i - \mathbf{p}_j\|_2. \quad (3.8)$$

一般来说，不存在一个通用的邻域定义。在我们的方法中，我们使用搜索  $K$  个最近节点作为邻域的策略。

因此，为了对图模型中的局部交互结构进行编码，我们考虑节点之间的距离，并

将每个节点与其相邻节点连接起来，构建了一个  $K$ -最近邻图。因此，对于连接  $v_i$  和  $v_j$  的每个边  $e_k \in E$ ，必须满足  $v_i \in \mathcal{N}_{v_j}$  或  $v_j \in \mathcal{N}_{v_i}$ ，其中  $\mathcal{N}_{v_i}$  和  $\mathcal{N}_{v_j}$  分别表示  $v_i$  和  $v_j$  的  $K$  最近相邻节点集。通过这一点，我们将创建一个稀疏连通图  $G$ ，这同时也减少了存储空间的需求并提高了计算效率。正如我们将在章节 3.2.4 中看到的， $G$  的结构将直接与我们对问题建模时所使用的稀疏矩阵联系起来。

### 3.2.3.2 成对交互模型

边的势函数的模型描述了两个相连通的节点之间的概率意义上的成对交互作用，这是我们所提出的概率图模型的核心。我们使用先前定义的运动一致性度量来描述交互模型（边的势函数），该模型将运动一致性引入到我们对问题的建模之中。

考虑连接  $v_i$  和  $v_j$  的一条边  $e_k$ ，并假设潜在的运动场具有空间一致性，那么使用运动一致性度量，两个节点都是内点的概率可以通过以下公式给出：

$$p(x_i = 1, x_j = 1) \propto e^{-\lambda c_{ij}}, \quad (3.9)$$

其中  $\lambda > 0$  是一个常数。出于概率意义上的严谨性，我们限制所有的条件概率的值之和为 1。令  $\psi(e_k) = \psi_{ij}(x_i, x_j)$ ，其中  $x_i$  和  $x_j$  指代节点  $v_i$  与  $v_j$  所对应的随机变量。一般意义上来说，出于对称性的考虑， $\psi_{ij}(1, 0)$  和  $\psi_{ij}(0, 1)$  应该具有相等的值。最终，对于边  $e_k$  其势函数应定义为：

$$\psi(e_k) = \begin{cases} e^{-\lambda c_{ij}} & x_i = 1, x_j = 1 \\ \frac{1}{d}(1 - e^{-\lambda c_{ij}}) & x_i \neq x_j \\ (1 - \frac{2}{d})(1 - e^{-\lambda c_{ij}}) & x_i = 0, x_j = 0 \end{cases} \quad (3.10)$$

值得注意的是，我们引入了一个参数  $d$  来控制所定义的概率分布，它也起到了施加一种先验知识的作用，如下所述。总的而言，对于边的势函数来说， $v_i$  和  $v_j$  不是同时为内点可以分为三种情况。我们可以简单地假设这三种情况是等价的，即， $d = 3$ 。但是，这种设置通常会使算法容易陷入局部最优。实际上，我们可以观察到真实的匹配更可能出现在成簇的匹配中，而离群点的则更分散（如文章<sup>[69, 82]</sup> 中所讨论的）。这一点说明如果一个节点的一个相邻节点是一个离群点，那么该节点就更有可能是一个离群值，即  $\psi_{ij}(0, 0) > \psi_{ij}(1, 0) = \psi_{ij}(0, 1)$ 。基于这一点，我们将  $d$  设置为略大于 3 的

一个值，可以看出， $d = 3$  为后两种情况赋予了相等的概率，而  $d > 3$  的设置使两者都是离群点的概率比只有一个离群点的概率要大。

在实现局部交互结构和成对交互模型的基础上，在本节中我们引入了局部运动一致性先验。应当说明的是，由于所有节点最终都会通过环得到连接，因此我们所设计的这种图模型的建模方式并不是产生分段平滑解的理想模型，但在实际的任务中通常可以产生这样的解。

### 3.2.4 求解算法

如等式 (3.1) 所述，求解的过程即是一种解码的过程，即找到使所有随机变量的联合概率分布最大化的一种最优配置。为此，我们需要归一化常数  $Z$ 、节点势函数以及边的势函数，如等式 (3.2) 所示。对于节点的势函数，我们简单地假设没有关于节点状态的先验知识，即使用均匀分布来描述， $\forall v_i \in V$ ，对于  $x_i = 0$  与  $1$  我们有  $\psi(v_i) = 0.5$ 。在这种情况下， $Z$  和  $\psi(v_i)$  不会随随机变量的变化而变化，在优化过程中可以忽略。因此，等式 (3.1) 中的解码问题可以重写为：

$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in \{0,1\}^N} \prod_{e_k \in E} \psi(e_k). \quad (3.11)$$

对于离散的二阶图模型来说，解码的问题可以视作一个整数二次规划问题 (Integer Quadratic Programming)。下一步我们来解释如何将解码问题的目标函数转化为关于随机变量  $\mathbf{x}$  的二次函数。

我们将连接  $v_i$  和  $v_j$  的边  $e_k$  等价地记作  $(v_i, v_j)$ 。使用矩阵  $\mathbf{M}_{11} \in \mathbb{R}^{N \times N}$  来编码  $\psi_{ij}(1, 1)$  对应的概率值：

$$\mathbf{M}_{11}(i, j) = \begin{cases} \log(\psi_{ij}(1, 1)), & \text{if } (v_i, v_j) \in E, \\ 0, & \text{otherwise.} \end{cases} \quad (3.12)$$

相似地，我们使用  $\mathbf{M}_{10}$ ， $\mathbf{M}_{01}$  和  $\mathbf{M}_{00}$  来分别编码  $\psi_{ij}(1, 0)$ ， $\psi_{ij}(0, 1)$  和  $\psi_{ij}(0, 0)$  所对应的概率值。最终目标函数可以写作：

$$\log \prod_{e_k \in E} \psi(e_k) = \mathbf{x}^T \mathbf{M}_{11} \mathbf{x} + \mathbf{x}^T \mathbf{M}_{10} (\mathbf{1} - \mathbf{x}) +$$

$$(\mathbf{1} - \mathbf{x})^T \mathbf{M}_{01} \mathbf{x} + (\mathbf{1} - \mathbf{x})^T \mathbf{M}_{00} (\mathbf{1} - \mathbf{x}), \quad (3.13)$$

其中  $\mathbf{1}$  代表元素全为 1 的列向量。

注意到条件  $\mathbf{M}_{10} = \mathbf{M}_{01}$ , 并且忽略常数项, 解码的问题可以写作等价的 IQP 形式:

$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in \{0,1\}^N} J(\mathbf{x}) = \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{c}^T \mathbf{x}, \quad (3.14)$$

其中  $\mathbf{Q} = \mathbf{M}_{11} - 2\mathbf{M}_{10} + \mathbf{M}_{00}$ ,  $\mathbf{c} = 2(\mathbf{M}_{10} - \mathbf{M}_{00})\mathbf{1}$ 。

IQP 问题是一种常见的 NP-hard 问题, 目前还没有高效率的算法来保证全局最优。因此, 一种较为常见的策略是将  $\mathbf{x}$  松弛在凸集中, 即去掉二进制约束, 令  $x_i \in [0, 1]$ 。尽管由于二次函数是非凸的, 松弛也不能保证全局最优解, 但松弛过后的问题通常将允许使用梯度方法(如 FW 算法)等数值优化手段。

在图像特征匹配中, 由于对计算时间的要求往往超越对全局最优性的要求, 因此一般采用启发式算法。并且通常启发式算法所得到的解也是令人满意的。在所提算法中, 通过对整数约束的松弛, 我们开发了一种基于 FW 的优化算法<sup>[149, 150]</sup>, 它不同于其他需要投影到约束条件所定义的空间上的基于梯度的方法, FW 算法可以自动停留在可行域来更新解。具体地说, 在每一次迭代中, FW 算法将考虑目标函数的线性近似, 并朝着该线性函数的极小值移动, 直到收敛。

$$\mathbf{y} = \arg \max_{\mathbf{y} \in \mathcal{D}} \langle \nabla J(\mathbf{x}^{(k)}), \mathbf{y} \rangle, \quad (3.15)$$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)} (\mathbf{y} - \mathbf{x}^{(k)}), \quad (3.16)$$

其中  $\nabla J(\mathbf{x}^{(k)})$  代表目标函数在  $\mathbf{x}^{(k)}$  的梯度:

$$\nabla J(\mathbf{x}^{(k)}) = \mathbf{Q} \mathbf{x}^{(k)} + \mathbf{c}. \quad (3.17)$$

在下文, 我们将讨论如何解决 (3.15) 并且在每次循环中确定  $\alpha^{(k)}$  以完成我们提出的整个算法。

显然, (3.15) 是一个线性规划的问题, 可以使用经典的内点法或者单纯形法来解决。然而这种做法在实际中具有很高的时间消耗。幸运的是, 该问题的约束仅仅是对

**Algorithm 2 MCDM 算法**


---

输入: 初始匹配集合  $\mathcal{S} = \{(\mathbf{u}_i, \mathbf{v}_i)\}_{i=1}^N$ , 参数  $s, d, K, \lambda$

输出: 内点集合  $\mathcal{I}$

- 1: 初始化  $\mathbf{x} = \mathbf{1}$ ;
  - 2: 将  $S$  转化为向量场表示  $F = \{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^N$ ;
  - 3: 在  $F$  中构建一个  $K$  近邻的图模型;
  - 4: 根据 (3.7)与(3.10)计算边的势函数  $\psi(e_k)$ ;
  - 5: **while**  $\mathbf{x}$  未收敛 **do**
  - 6:   根据 (3.15)与 (3.18)来决定最优解  $\mathbf{y}$ ;
  - 7:   根据 Eqs. (3.16)与(3.19)来更新  $\mathbf{x}$ ;
  - 8: **end while**
  - 9: 返回最优解  $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_N^*)$ ;
  - 10: 利用 (3.20)来决定最终的内点集  $\mathcal{I}$ .
- 

于  $\mathbf{x}$  中每个分量  $x_i$  的上界与下界的限制, 这使得该问题具有一种解析形式的解法。具体来说, 令  $\mathbf{G} = \nabla J(\mathbf{x}^{(k)})$ , 则对于  $\forall i = 1, 2, \dots, N$ , (3.15)的最优解为

$$\mathbf{y}_i = \begin{cases} 1, & \mathbf{G}_i \geq 0, \\ 0, & \mathbf{G}_i < 0, \end{cases} \quad (3.18)$$

其中  $\mathbf{G}_i$  代表  $\mathbf{G}$  的第  $i$  个元素。

在算法中, 步长可以依照一种线性搜索的方式来确定<sup>[151]</sup>, 其也具有一种解析形式的最优解。具体来说, 令  $\mathbf{d}^{(k)} = \mathbf{y} - \mathbf{x}^{(k)}$ , 则  $\alpha^{(k)}$  应当是一个抛物线方程的最大值点 (出于简洁的考虑, 下标  $(k)$  被省略):

$$\alpha^* = \arg \max_{\alpha \in [0, 1]} J(\mathbf{x} + \alpha \mathbf{d}) = a\alpha^2 + b\alpha + const, \quad (3.19)$$

其中  $a = \mathbf{d}^T \mathbf{Q} \mathbf{d}$ ,  $b = 2\mathbf{x}^T \mathbf{Q} \mathbf{d} + \mathbf{c}^T \mathbf{d}$ 。基本的代数知识可以告诉我们最大值点必然在以下的三个点中:  $\{0, 1, -\frac{b}{2a}\}$ 。因此,  $\alpha^*$  可以通过为每一个点计算  $J(\mathbf{x} + \alpha \mathbf{d})$  的值并选取其中达到最大值的点来确定。

一旦 FW 算法的迭代过程收敛, 则内点的集合  $\mathcal{I}$  可以通过一个简单的阈值筛选获得:

$$\mathcal{I} = \{(\mathbf{u}_i, \mathbf{v}_i) | x_i^* > 0.5, i = 1, \dots, N\}. \quad (3.20)$$

我们将算法命名为 MCDM(Motion-Consistency Driven Matching), 并将算法的细节记录在算法 2中。

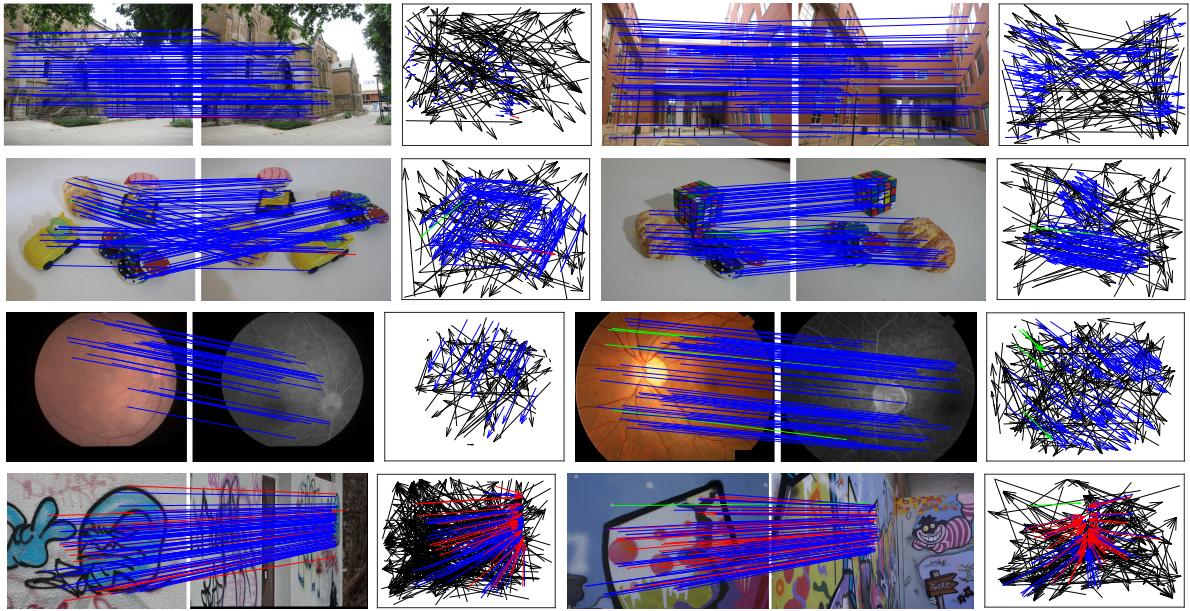


图 3.2 MCDM 在 8 个典型图像对上的特征匹配性能的定性说明。对于每组结果来说，左图为匹配结果，右图为初始匹配的向量场表示，其中每个箭头的头部和尾部分别对应于两个图像中特征点的位置（蓝色 = 真阳性，black = 真阴性，绿色 = 假阴性，红色 = 假阳性）。出于视觉效果的考虑，在左图中，最多展示 100 个随机选择的匹配，同时真阴性没有被显示。

### 3.2.5 计算复杂度分析

MCDM 算法包括三个主要步骤，包括搜索  $K$  近邻、确定边的势函数以及用 FW 算法最大化目标函数。利用 K-D 树算法<sup>[152]</sup> 搜索  $S$  中每个特征点的  $K$  近邻，时间复杂度大约是  $O((K + N) \log N)$ 。为了确定边的势函数，其时间复杂度取决于边的数目。对于  $K$  最近邻图的图模型，边的个数不超过  $KN$ ，因此这一步的时间复杂度为  $O(KN)$ 。在优化阶段，由于我们在每个迭代中都有一个解析形式的更新过程，因此主要的时间需求是涉及  $\mathbf{Q}$  的矩阵运算。显然， $\mathbf{Q}$  中的每个非零元素对应于  $E$  中的一条边，这表示  $\mathbf{Q}$  是稀疏的。简单稀疏矩阵运算的计算复杂度应与  $\mathbf{Q}$  中非零元素的个数成正比，即小于  $O(KN)$ 。因此，优化阶段的时间复杂度为  $O(mKN)$ ，其中  $m$  是迭代次数。通常， $K$  是常量，且  $K \ll N$ 。因此，从总体上来说，我们所提出的 MCDM 算法的总时间复杂度是线性的，可以简单地写成  $O(mN + N \log N)$ 。在实际应用中，我们的基于 FW 的算法在大多数情况下收敛于 10 迭代次数内。对于存储空间的复杂度来说，主要的内存需求是存储  $K$  最近邻和边的势函数，其复杂度为  $O(KN)$ 。考虑到  $K \ll N$ ，MCDM 的总空间复杂度也是线性的，可以简单地写成  $O(N)$ 。

### 3.2.6 实现细节

引导性的邻域构建：在我们建模时， $K$  最近邻是从整个初始匹配的集合  $S$  中搜索得来的。由于受到离群点严重污染的邻域提供的运动一致性信息很少，因此在离去群点占很大比例的情况下，这一策略将受到干扰而不能形成很好的效果。因此，我们采用了一种引导策略来更好地构建邻域<sup>[81]</sup>。特别地，我们最初使用整个节点集  $V$  来构造邻域，过滤错误的匹配，并获得一个内点集  $\mathcal{I}$ 。然后，使用更加干净的子集  $\mathcal{I}$  为每个节点构造一个新的邻域。随后，对整个初始匹配的集合  $S$ ，利用重建的邻域结构，重新运行我们的 MCDM。通过使用这种引导策略，我们发现误匹配去除的性能可以得到显著提升。实际上，这个过程可以一致迭代直到达到某种收敛的状态为止。然而，在实践中，两次迭代已经可以产生令人满意的结果。因此，为了提高效率，我们只使用两次迭代。

参数的设置：在我们的方法 MCDM 中，存在着四个参数： $s$ 、 $d$ 、 $K$  和  $\lambda$ 。参数  $s$  用于使运动一致性度量更加平滑；我们根据经验将其设置为 0.2。参数  $d$  用于修改边的势函数的概率分布，其应该略大于 3，我们根据经验将其设置为 3.05。参数  $K$  确定最近邻构造时的邻域数目，它控制局部和全局运动一致性先验之间的折衷。参数  $\lambda$  表示我们对运动一致性的信心，即  $\lambda$  越大，则运动一致性约束就越强。由于我们使用的引导性的邻域构造方式， $K$  和  $\lambda$  都存在两个设置。在第一次迭代中，我们设置  $K = \min(\text{round}(N/15), 50)$  和  $\lambda = 0.5$ ；而在第二次迭代中，我们设置  $K = 3$  和  $\lambda = 1.5$ 。我们将在实验中对参数设置进行详细的分析。在整个实验过程中，参数值是固定的。此外，在算法处理之前，我们将规范化  $\{\mathbf{u}\}_{i=1}^N$  和  $\{\mathbf{v}\}_{i=1}^N$ ，使其均值和单位方差为零。

## 3.3 实验结果

在本节中，我们对 MCDM 在真实图像的特征匹配中的性能进行评估，并将其应用于基础矩阵估计以及闭合检测任务中。我们使用开源的 VLFeat 实现 K-D 树<sup>[152]</sup>的算法以搜索  $K$  近邻。实验中使用的环境配置是 3.40Ghz 的 Intel Core i5-7500 CPU、8GB 内存以及 MATLAB 代码。

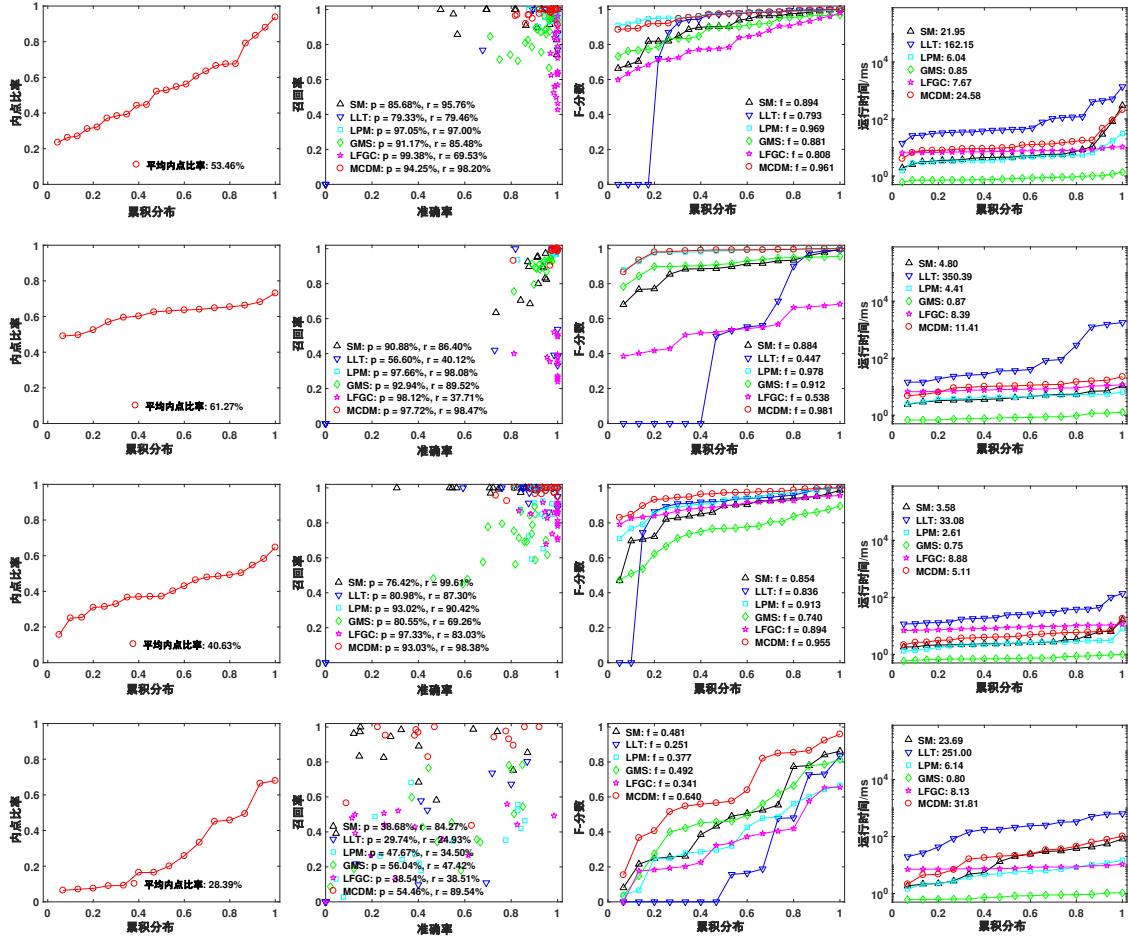


图 3.3 所有算法在四个数据集上的定量比较。从上到下数据集依次为：Adelaide1、Adelaide2、Retina 和 EVD。从左到右依次为：初始匹配内点比率的累积分布、精确率与召回率的统计、 $F$ -分数的累积分布和运行时间的累积分布。累积分布的意义在于，坐标为  $(x, y)$  的曲线上的一点表示有  $(100*x)$  个百分比的图像对，其性能值（即内点数目、 $F$ -分数或运行时间）不超过  $y$ 。对于  $F$ -分数来说，方法的性能越好，其展示的曲线越接近顶部，而对于运行时间则相反。

### 3.3.1 图像特征匹配实验结果

#### 3.3.1.1 数据集以及度量指标

-*Adelaide1*: *Adelaide* 数据集<sup>1</sup>是一个公开的数据集，其在许多先前的工作中已经得到广泛的使用<sup>[153]</sup>。我们将它分为两部分：*Adelaide1* 和 *Adelaide2*。其中 *Adelaide1* 具有 23 对不同建筑的图像，其匹配主要由极线几何关系约束。该数据集使用 SIFT 算法来建立匹配对应关系，并包括人工构建的地真标签。

-*Adelaide2*: *Adelaide2* 来自 *Adelaide* 数据集，其仅包括存在多个运动的图像对。该数据集有 16 幅图像，其中独立运动的数量在 1 到 4 之间。匹配对应关系与地真的标签的构建方式与 *Adelaide1* 相同。

<sup>1</sup><https://cs.adelaide.edu.au/~hwong/doku.php?id=data>

-*Retina*: *Retina* 是一个由 20 对视网膜图像组成的医学图像集。由于匹配图像的模态不同，我们使用 SURF<sup>[39]</sup> 构造匹配对应关系。地真的标签由我们手工建立，由于多模态的场景，该数据集的内点率相对较低。

-*EVD*<sup>2</sup>: 此数据集包含 15 对图像对，图像之间存在极端的视角变换。初始匹配的几何以及地真的单应矩阵由数据集给出。为了构造地真的标签，我们采用了一种简单的策略，即利用给定的单应变换来避免人工构建标签：我们计算每个匹配相对于地真的单应矩阵的变换误差<sup>[3]</sup>，并将阈值设置为 3 个像素来确定内点。由于图像之间存在的极端视角变换，该数据集具有很大的挑战性。

我们采用准确率、召回率和 F-分数来表征性能<sup>[82]</sup>，其中准确率定义为识别的真实内点数目与所有保留的匹配数目的比值，召回率定义为识别的真实内点的数目与所有真实内点数目的比值，F-score 定义为  $2 \cdot \text{Precision} \cdot \text{Recall} / (\text{Precision} + \text{Recall})$  的比值。

### 3.3.1.2 定性研究

为了直观地说明我们方法在图像特征匹配任务中的性能，我们从上述数据集中选择了 8 对典型的图像对，并将结果显示在图 3.2 中。这些图像对具有不同类型的变换，包括极线几何、单应以及独立运动等。这些图像对的初始内点的比值分别为 52.16%，63.49%，65.40%，64.78%，31.43%，37.21%，28.65% 与 31.42%。利用本文提出的误匹配去除的方法之后，8 对图像的准确率、召回率和 F-分数，分别为 (99.25%，100.0%，0.9963)，(100.00%，100.00%，1.0000)，(99.35%，99.35%，0.9935)，(100.0%，98.66%，0.9932)，(100.00%，100.00%，1.0000)，(100.00%，95.00%，0.9500)，(79.31%，100.00%，0.8846) 和 (54.59%，98.26%，0.7019)。从结果可以看出，我们所提出的局部空间一致性先验是通用的，能够处理不同类型的空間变换，甚至可以处理图像场景中存在独立运动的情形（如后两个例子）。测试数据中的优异结果也证明了所提出的 MCDM 的有效性。

### 3.3.1.3 数据集定量研究

接下来，我们在公开的数据集上对所提方法进行定量评估，并与六种近年来提出的方法进行了比较，即 SM<sup>[53]</sup>、LLT<sup>[79]</sup>、LPM<sup>[81]</sup>、GMS<sup>[82]</sup> 和 LFGC<sup>[133]</sup>。这些竞争方法在文献中分别对应着不同的类别，因此可以在一定程度上很好地代表图像特征匹

<sup>2</sup><http://cmp.felk.cvut.cz/wbs/>

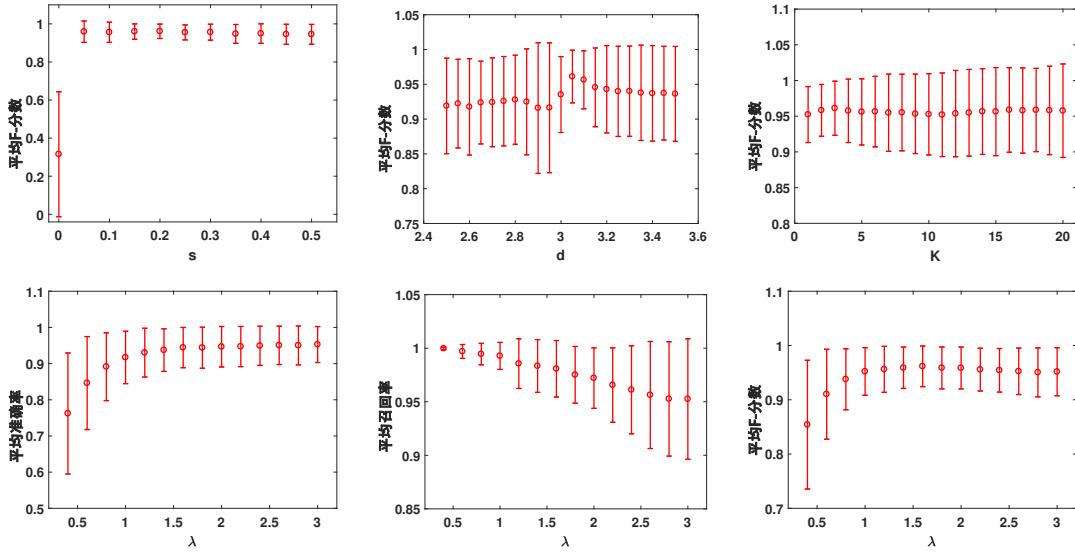


图 3.4 在 Adelaide1 数据集上的参数分析结果，我们使用带有标准差的平均 F-分数来分析其性能变化。对于参数  $\lambda$  来说，平均准确率和召回率也展示在图中。

配领域。所有的方法均基于公开的代码实现，并且我们尽可能地将参数调优，然后在整个实验过程中固定不变。对于 SM，我们使用所有匹配的平均置信度作为阈值来确定内点和离群点。对于 LLT，我们使用其非刚性的版本，它本质上可以看做 VFC<sup>[69]</sup> 的改进算法。

图 3.3 给出了定量的比较结果，在每一行，我们提供某一数据集的内点率、成对的准确率与召回率、F-分数以及运行时间的统计信息。内点率的波动范围较大，平均值较低，因此使得图像特征匹配任务具有较大的挑战性。在这四个数据集中，初始匹配的数目的平均值分别为 392.6, 263.7, 159.2 and 471.3。在准确率-召回率的统计图中，每个点的坐标表示某一图像对上的一个准确率-召回率数据。此外，在图例中我们给出了统计平均值的结果，即平均准确率、召回率和 F-分数。在运行时间的统计图中，我们给出了平均运行时间作为效率评估的一个直接标准。可以看出，SM 在所有四个数据集的精确度和召回率数据上均较所提出的 MCDM 方法要差。GMS 是最快速的匹配方法，但其只能给出粗略的匹配结果。LLT 采用一个全局变换的模型，不适用于多个运动的情况，这一点在 Adelaide2 数据集上可以看出。LFGC 作为一种基于深度学习的方法，受训练数据的限制，泛化能力相对较差。具体而言，在 Adelaide1 和 Retina 中，LFGC 的准确率较高，但召回率很低。此外，这一算法也不能处理复杂的场景，例如多个运动和极端的视角变化，这使得它在 Adelaide2 和 EVD 数据集上的性能较差。在其它的方法中，LPM 是主要的竞争算法，但 MCDM 仍有一些明显的优势。在 Adelaide1

和 *Adelaide2* 中, LPM 和我们的 MCDM 具有相当的性能, 但是, 在更具挑战性的情况下, 例如低内点比率以及初始匹配较为稀疏时, 我们所提出的 MCDM 的性能要优于 LPM, 这一点在 *Retina* 和 *EVD* 上的结果可以看出。在运行时间方面, 我们可以看到除了 LLT 以外, 其他的所有方法均具有高效率的特点, 并且差异并不显著。

### 3.3.1.4 参数分析

在图 3.4 中, 我们给出所提方法中四个参数的参数分析, 即  $s$ 、 $d$ 、 $K$  和  $\lambda$ 。为简单起见, 我们只使用 *Adelaide1* 数据集进行评估, 并且当测试一个参数时, 所有其他参数都是固定的。参数  $s$  用于对运动一致性度量进行平滑。从图 3.4 中左上角的图可以看出, 缺少这一参数 (即  $s = 0$ ) 会使得运动一致性度量受到影响并导致较差的性能。相比之下, 在  $s > 0$  的情况下, 我们的 MCDM 始终保持较高的匹配精度。参数  $d$  用于微调边的势函数中的概率分布。如前所述, 我们的结论是  $d$  应略大于 3 作为某种先验知识的体现。图 3.4 中, 上排中间的图的结果验证了这一点。从图 3.4 中右上角的图中, 我们可以观察到, 随着参数  $K$  的增加, 平均的 F-分数先呈现增加的趋势然后呈现降低的趋势, 这表明  $K$  取一个相对较小的值时, 效果最好。注意到  $K$  平衡着局部和全局的运动一致性的先验, 因此这一结果也说明了局部运动一致性先验的优势。同时, 较小的  $K$  也代表着更少的计算消耗。

图 3.4 的第二行展示了使用 F-分数, 准确率以及召回率的参数  $\lambda$  的参数分析结果。在所提的 MCDM 方法中,  $\lambda$  的提高代表对不连续性施加一种更强的惩罚, 即其实际上控制着约束的强弱性。图中的结果也证实了我们的解释, 当  $\lambda$  变高时, 准确率会提高而召回率会降低。因此实际上  $\lambda$  控制着准确率与召回率的折衷, 因此提高了所提的 MCDM 方法的灵活性。

### 3.3.1.5 算法分析

我们对于所提的 MCDM 方法也进行了其他方面的深入分析, 包括关于引导性的邻域构建策略, 基于 FW 的优化策略以及对于离群点的敏感程度。因此, 我们在所有的数据集的集合上进行实验, 将数据展示在 Fig. 3.5 的左子图, 并在右子图中展示使用合成数据所得来的对于离群点比率敏感度的分析结果。在合成数据中, 我们将数据集中的内点抽离出来, 并且人工地通过随机连接两幅图像的两个点来添加一部分离群点以

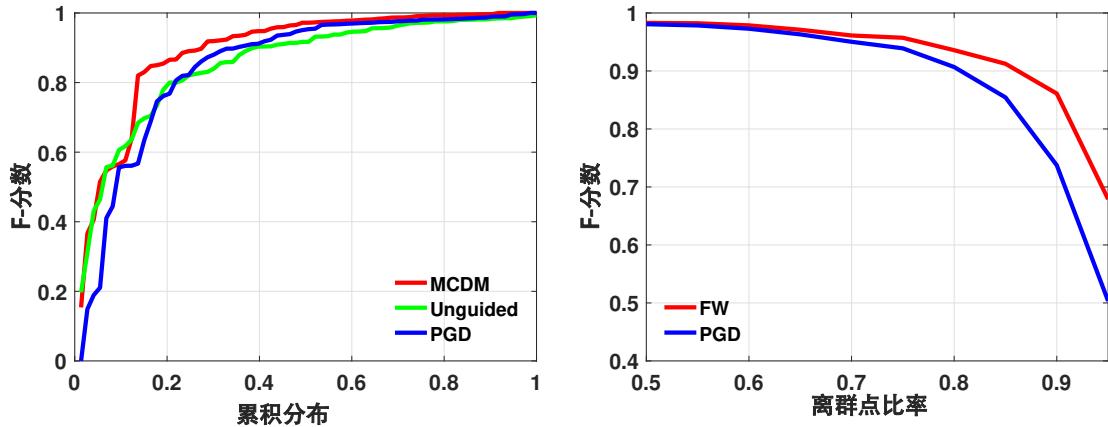


图 3.5 对 MCDM 在三个方面的分析，包括引导性的邻域构建策略，基于 FW 的优化策略以及对于离群点的敏感性。左图展示了 F-分数的累积分布，其中 Unguided 代表所提的 MCDM 无引导性邻域重建的情况，PGD 代表使用投影梯度下降代替基于 FW 的算法。在右图中展示了所提的 MCDM 的性能随着离群点比率的性能变化，其中 FW 代表使用基于 FW 的优化策略，而 PGD 表示使用投影梯度下降的优化策略。对于左图的累积分布来说，曲线中的一个点  $(x, y)$  代表有  $(100 \cdot x)\%$  的图像对的 F-分数不大于  $y$ 。算法的性能越好，则曲线越靠近顶部。

控制离群点的比率。我们将基于 FW 的优化策略与投影梯度下降 (projected gradient descent, PGD) 的方法在解 (3.14) 的问题时相比较。对于 PGD 算法来说，我们优化调整了学习率并设置为 0.001。我们使用 F-分数作为度量的指标。

在左子图中，我们展示了关于 F-分数的累积分布。可以看到，由于引导性的邻域构建策略使得算法抵抗离群点的能力增强，其带来了显著的性能的提升。在右子图中，我们研究了所提的 MCDM 方法对于离群点比率的敏感性。虽然 MCDM 的性能随着离群点的增多将会下降，然而其抵抗离群点的能力也得到了体现，平均的 F-分数的值在存在 90% 的离群点的情况下也可以达到 0.8。并且，我们所使用的基于 FW 的优化策略相对于 PGD 的优势也得到了论证。这是由于 FW 可以在迭代的过程中自适应的决定最优的步长。

### 3.3.2 基础矩阵估计实验结果

对于图像特征匹配的任务来说，一个最为重要的应用就是用来重建场景的三维结构。在这一过程中，匹配常与鲁棒的估计子（如 RANSAC）一起使用来获得基础矩阵的估计（对于未标定的相机）或者本质矩阵的估计（对于已标定的相机）。这一估计的结果毫无疑问对整个系统的性能有着重要的影响。在下面我们将测试所提的方法在这一下游应用中的表现，并使用估计的精度作为度量标准。

我们的实验在<sup>[154]</sup> 中所提出的测试基准中进行。在该基准中介绍了四个数据集，

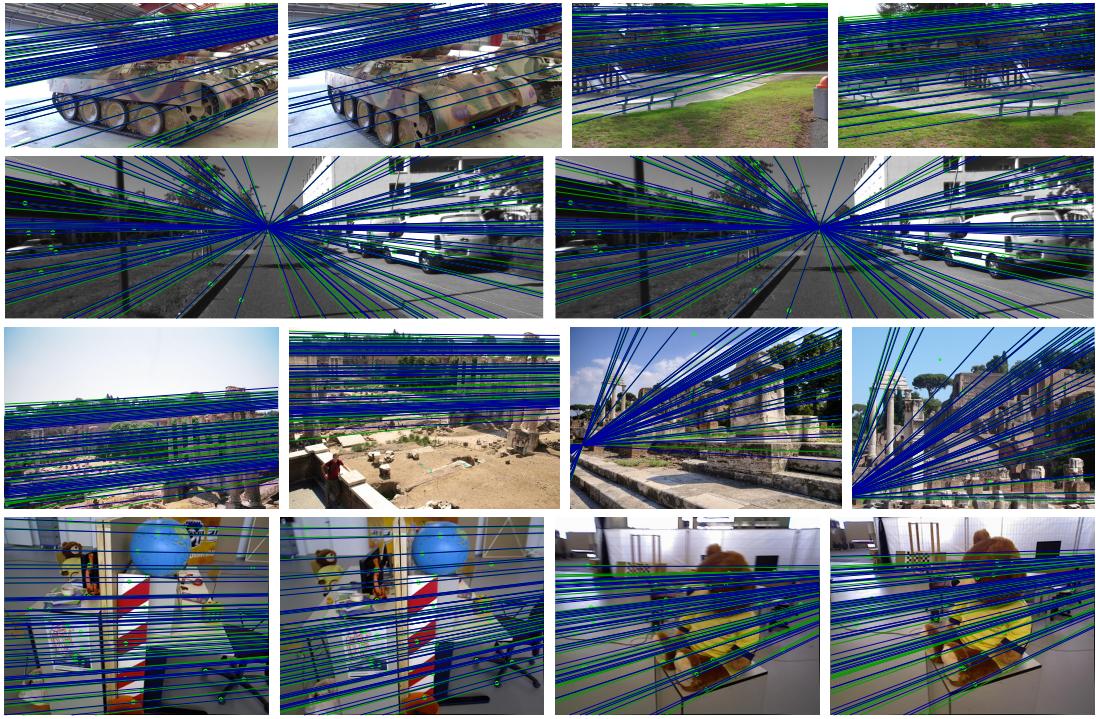


图 3.6 所提出的 MCDM 方法在基础矩阵估计中的定性的结果。其中蓝色的极线由地真的基础矩阵产生，而绿色的线由估计值产生。第一行代表在 Tanks and Temples 数据集中的结果，第二行代表在 KITTI 数据集中的结果，第三行代表在 CPC 数据集中的结果，第四行代表 TUM 数据集中的结果。

包括 (i) *TUM SLAM* 数据集<sup>[155]</sup>: 其是一个室内场景的数据集，包含着窄基线的图像对，其分辨率为  $480 \times 640$ 。(ii) *KITTI* 里程计数据集<sup>[156]</sup>: 这是一个驾驶场景的数据集，图像之间不同视角的运动是前向的，包含分辨率为  $370 \times 1,226$  的窄基线图像对。(iii) *Tanks and Temples (T&T)* 数据集<sup>[157]</sup>: 该数据集提供了许多用于三维重建的对场景或物体的不同视角的拍摄结果，并且是宽基线的。其分辨率为  $1,080 \times 2,048$  或  $1,080 \times 1,920$ 。(iv) *Community Photo Collection (CPC)* 数据集<sup>[158]</sup>: 该数据集提供了使用未知的相机在未知的时间所拍摄的世界上的著名景点的无序图像，因此图像对是宽基线的并且分辨率是多样化的。

在所有可以构成匹配的图像对中，对每个数据集随机挑选了 1,000 对图像来对算法进行测试。地真值为测试基准所提供的基础矩阵。其中，在 *TUM* 与 *KITTI* 中，基础矩阵可以由数据集中给出的相机的内参与外参中得出。对于 *T&T* 与 *CPC* 而言，地真的基础矩阵由利用 COLMAP<sup>[16]</sup> 所给出的三维重建结果得出。所估计得来的基础矩阵使用<sup>[159]</sup> 所给出的 *symmetric geometry distance (SGD)* 来衡量<sup>[154]</sup>。考虑到不同图像的分辨率所带来的影响，我们在实验中使用 Normalized SGD<sup>[154]</sup>，其可以理解为

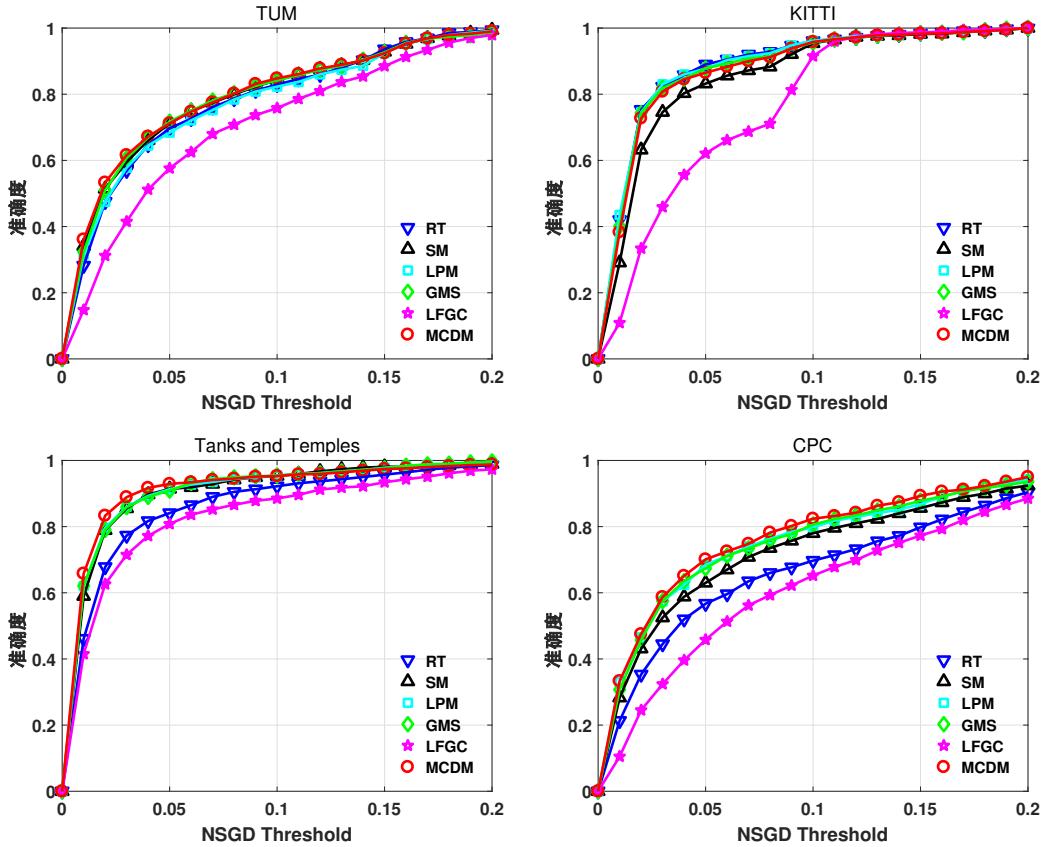


图 3.7 关于 *Normalized SGD* 在 TUM, KITTI, T&T 与 CPC 数据集中的累积分布函数。算法越精确，则曲线越靠近顶部。

SGD 在使用图像的对角线长度归一化之后的结果。

在我们的实验中，如同文章<sup>[154]</sup>，SIFT 算法被用来建立初始的匹配，并使用 ratio test 来滤除一部分的离群点。然而，不同于<sup>[154]</sup> 中直接将 ratio test 使用跟其他的误匹配去除的方法代替的做法，我们认为将误匹配去除的方法在 ratio test 之后使用才能达到最佳的效果。这是由于 ratio test 与一般的误匹配去除算法利用的是不同的信息，因此应当被同时使用。出于这种考虑，我们将 SIFT 建立的匹配使用 0.8 的 ratio test 来滤除一部分匹配作为误匹配去除方法的输入。我们也测试了仅仅基于 ratio test 的结果，记作 RT。在最终的基础矩阵估计的过程中，我们采用了最近的处于领先水平的 MAGSAC++ 算法<sup>[98]</sup>，并将其参数设置为与原文<sup>[98]</sup> 中一致。我们对其施加了一个 1 秒钟的时间限制来控制算法的结束。除 LLT<sup>[79]</sup> 由于其时间消耗过大以外，所有的对比方法均得到了测试。

一些定性的基础矩阵的估计结果在图 3.6 中被给出。为了直观地看出估计结果的准确性，我们在图像中展示了地真基础矩阵与估计值在图像中所产生的极线。所得到

表 3.1 所有方法在四个数据集中基础矩阵估计的精度统计结果。准确与否使用 *Normalized SGD* 衡量，门限为 0.05。黑体代表最好的结果。

Dataset	RT	SM	LPM	GMS	LFGC	MCDM
<i>TUM</i>	69.70	71.20	68.40	<b>71.70</b>	57.6	71.20
<i>KITTI</i>	<b>89.00</b>	83.20	87.70	87.80	62.1	86.50
<i>T&amp;T</i>	84.10	91.40	91.30	90.90	80.8	<b>92.90</b>
<i>CPC</i>	56.70	63.00	68.50	67.40	45.8	<b>70.10</b>

的估计越精确，则其与地真基础矩阵的极线越接近。关于 Normalized SGD 误差的累积分布函数在图 3.7 中给出。作为一个更直观的度量指标，我们也在表格 3.1 中提供了准确度（与文章<sup>[154]</sup> 所使用的召回率意义相同），即准确的估计占所有的估计结果的比值，的统计结果。所得来的基础矩阵的估计简单地被一个阈值分为准确与不准确的两种情况，与文章<sup>[154]</sup>，我们将这个阈值设置为 0.05。从结果中可以看到，LFGC 在所有的数据集中均具有较差的性能，这是由于作为一种数据驱动的方法，其泛化能力不能得到有效保证。对于窄基线的数据集来说，除 LFGC 之外的所有的方法的表现均类似。对于宽基线的数据集 *T&T* 与 *CPC* 来说，情况有所不同。除 LFGC 以外，所有的误匹配去除的方法均相较于 RT 带来了明显的性能提升。其中，MCDM 由于其使用的概率图模可以更好地利用场景的几何信息，因此达到了最优的结果。

### 3.3.3 闭环检测实验结果

闭环检测是视觉 SLAM 的一个核心的问题之一，该任务的核心是识别环境中曾经到达的场景。对于一个 SLAM 系统来说，闭环检测可以有效地减少轨迹计算中的由于累积误差所造成的轨迹计算中的漂移，从而得出一个全局一致的地图。我们将一个说明性的例子展示于图 3.8 中。在本节中，我们研究图像特征匹配的算法在基于外观的闭环检测中的应用。在这一技术路线中，图像之间的相似度由准确匹配的数目决定，并用来识别曾经到达的场景。传统的算法一般基于 RANSAC 来去除离群点，本节我们研究基于误匹配去除的方法来完成这一任务。RANSAC 算法则被用做一个基准算法，并使用 1 个像素作为内点与离群点的区分门限，0.99 作为置信值，5000 作为最大的循环次数。

在我们的实验中，为了降低时间的需求我们采用了一种两步的策略。在第一步，闭环的候选帧由快速的 bag-of-words (BoW)<sup>[160]</sup> 方法选取。在第二步，我们在图像之间建立匹配关系并使用误匹配去除来检测闭环事件。为了测试方法的性能，我们选取了

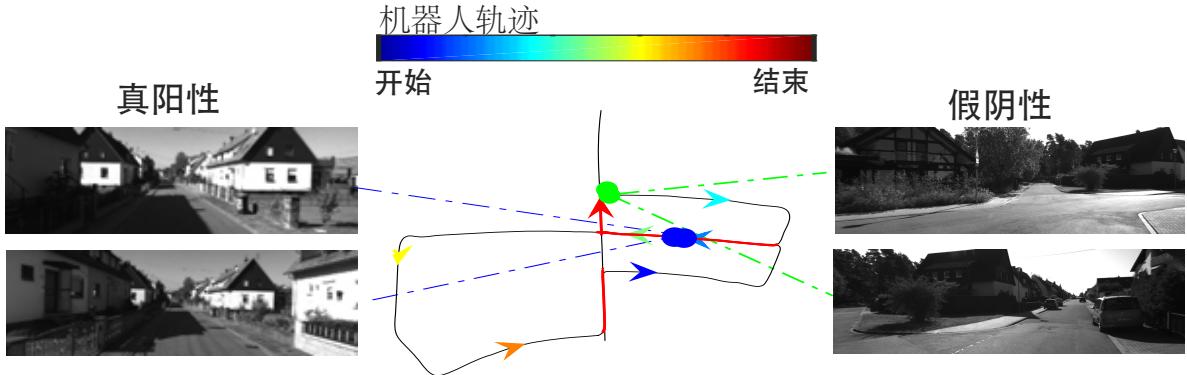


图 3.8 在 KITTI 05 数据集中构建的一个闭环检测说明性的例子。黑色的轨迹是由相机的姿态转化而来的，其中彩色的箭头以 colorbar 为基准指示着时间推进的方向。红色的点代表一个闭环事件的发生，并且具有代表性的真阳性与假阴性的例子被突出显示。

表 3.2 所有算法 100% 准确率下的最大召回率以及平均运行时间的比较。黑体代表最好的结果。

Dataset	RANSAC	SM	LPM	GMS	MCDM
<i>Lip6Outdoor</i>	80.25/212.7	45.27/ <b>1.2</b>	80.76/5.8	62.69/2.3	<b>81.09</b> /2.5
<i>KITTI 05</i>	87.77/186.2	83.62/3.4	<b>93.01</b> /7.6	88.21/ <b>2.0</b>	91.92/9.5
<i>Euroc MH 05</i>	60.04/255.1	67.54/4.3	77.73/9.3	<b>81.22</b> / <b>1.1</b>	80.93/7.9

三个公开的数据集, *Lip6Outdoor*<sup>[161]</sup>, *KITTI 05*<sup>[156]</sup>, *Euroc MH 05*<sup>[162]</sup>。*Lip6Outdoor* 数据集和 *KITTI 05* 数据集代表着室外环境, 其分辨率为  $240 \times 192$  与  $1,241 \times 376$ 。*Euroc MH 05* 数据集代表着室内环境, 其分辨率为  $752 \times 480$ 。

对于闭环检测任务来说, 由于假阳性对于 SLAM 系统有着非常严重的负面影响, 甚至可能会产生完全错误的地图, 因此其结果必须保证假阳性的情况尽可能的少。在这种情况下, 测试闭环检测算法的性能常用的指标是在 100% 准确率的情况下最大的召回率。我们也同时汇报了算法平均的运行时间, 结果展示于表格 3.2 中。可以看到, RANSAC 作为基准的算法, 在 *Lip6Outdoor* 与 *KITTI 05* 数据集中性能一般。然而, 在室内数据集 *Euroc MH 05* 中 RANSAC 近乎完全失败。对于其他的误匹配去除算法来说, 其计算效率一般较 RANSAC 要快, 性能表现一般也较好。相较于其他的方法, 所提的 MCDM 性能表现最为稳定, 并且展示出最好的在三个数据集中的平均性能。

### 3.4 本章小结

在本章中, 我们研究了概率图模型在图像特征匹配任务中的应用, 并提出一种快速的误匹配去除算法来应对离群点。出于计算效率的考虑, 我们考虑了最初等的二阶

图模型的结构，利用匹配数据两两之间的相关性关系作为算法的基础。针对图像特征匹配这一问题，我们设计了一种局部运动一致性的交互模型，并将其应用于概率图模型中作为去除离群点的先验。最终，我们将所建立的概率图模型进一步抽象为一种特殊的整数二次规划的形式，并提出一种基于 Frank-Wolfe 方法的快速求解算法。在图像鲁棒特征匹配，基础矩阵估计以及闭环检测的实验中，我们证明了本章所提出的方法具有较强的鲁棒性，甚至可以对图像中存在多种运动模式的场景进行处理，与处于领先水平的现有算法进行对比，也显示出了本章所提方法的优越性。

## 4 基于子空间恢复的鲁棒几何模型估计方法研究

### 4.1 引言

图像特征匹配是计算机视觉领域一个基本的问题，在诸多的视觉任务中均起着重要的作用。在 3D 视觉中，图像特征匹配是将 2D 图像与 3D 场景联系起来的重要手段。在前人研究的基础上，经典的多视图几何的视角将现实生活中一大类的图像看做是真实世界的 3D 场景在摄像机所决定的平面上的一种投影<sup>[3]</sup>。基于这种观点，我们就可以由理论上导出一系列从图像中获知相机的运动，或者从图像中重构真实世界的 3D 点云的方法，而图像特征匹配，就是这一类方法的核心。具体来说，图像的特征匹配中包含着关键的几何信息，根据不同的情况，被定义为称作基础矩阵或者称作单应变换的几何模型。

正如前文所说，由于特征匹配一般会受到噪声以及更难以处理的离群点的影响，从初始匹配的集合中得出几何模型的精确估计是一个非平凡的问题。在现有的文献中，经典的 RANSAC 方法<sup>[85]</sup> 占据了鲁棒几何模型估计方法的主流。然而，RANSAC 方法由于其随机采样的特性，对于离群点较多的情况效果并不理想。在另一类方法中，许多学者希望通过构建确定性的优化模型来解决这一问题。然而，由于这一问题的 NP-hard 的特性，构建优化模型常常会受到局部最优解的困扰，尤其在需要考虑计算效率的情况下更是如此。

在本章中，我们将图像特征匹配与其在 3D 视觉中的应用直接联系起来，研究一种新的几何模型估计的算法。从几何模型的数学定义与背景出发，我们创新性地从一个新的角度来考虑基于优化的鲁棒几何模型估计的方法，并在保证高效性的同时尽可能地摆脱局部最优解。具体来说，我们将更加鲁棒的  $l_1$  损失介绍到鲁棒几何模型估计的问题中，并对传统的直接线性变换方法<sup>[3]</sup>(Direct Linear Transformation, DLT) 做出了有效的改进。通过这样一种思路，我们将问题转化为一种新的数学形式，并和鲁棒子空间恢复<sup>[163]</sup>(Robust Subspace Recovery, RSR) 相联系起来提出了一系列新的算法。在实验中，我们验证了所提方法的高效性与准确性。

## 4.2 基于子空间恢复的鲁棒几何模型估计方法

### 4.2.1 经典的 DLT 算法

假设给定一个存在一定的离群点的初始匹配的集合  $S = \{(\mathbf{x}_i, \mathbf{x}'_i)\}_{i=1}^N$ , 其中  $\mathbf{x}_i = (x_i, y_i, 1)^T$  以及  $\mathbf{x}'_i = (x'_i, y'_i, 1)^T$  是代表两幅图像中的特征点的齐次坐标的列向量。我们的目的则是得出其中所编码的潜在的几何模型, 例如基础矩阵  $\mathbf{F} \in \mathbb{R}^{3 \times 3}$  或者单应变换  $\mathbf{H} \in \mathbb{R}^{3 \times 3}$ , 其对于 3D 视觉具有巨大的重要性。

首先我们对于经典的 DLT 算法给出一个简单的总结, 其利用问题的线性结构给出了几何模型估计的一种非常快速的解法。这里所说的匹配数据的线性结构指, 几何模型的估计可以通过求解一个由数据中导出的线性系统来获得。在下面, 我们将分别讨论基础矩阵  $\mathbf{F}$  与单应变换  $\mathbf{H}$  的基于 DLT 的解法。

基础矩阵  $\mathbf{F}$  代表两视图之间一种广泛存在的极线约束。这种约束可以表达为:

$$\mathbf{x}'_i^T \mathbf{F} \mathbf{x}_i = 0. \quad (4.1)$$

在特征点分布于一个近似平面或者相机的运动是纯旋转时, 图像特征匹配之中潜在着单应变换的几何模型。这一变化可以通过向量的外积来表示:

$$\mathbf{x}'_i^T \times \mathbf{H} \mathbf{x}_i = \mathbf{0}. \quad (4.2)$$

对于  $\mathbf{F}$  与  $\mathbf{H}$  的估计, DLT 的解法可以视作求解一种超定的线性系统:

$$\mathbf{Mz} = \mathbf{0}, \quad (4.3)$$

其中  $\mathbf{M} \in \mathbb{R}^{K \times D}$  代表从匹配中导出的数据矩阵, 而  $\mathbf{z} \in \mathbb{R}^D$  代表未知参数形成的列向量。对于基础矩阵估计而言, 我们有  $\mathbf{M} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]^T$ , 并且数据的嵌入形式为:

$$\mathbf{a}_i = (x'_i x_i, x'_i y_i, x'_i, y'_i x_i, y'_i y_i, y'_i, x_i, y_i, 1)^T. \quad (4.4)$$

对于单应变换的估计而言，我们有  $\mathbf{M} = [\mathbf{b}_1^T, \mathbf{b}_2^T, \dots, \mathbf{b}_n^T]^T$ ，并且数据的嵌入形式为：

$$\mathbf{b}_i^T = \begin{bmatrix} \mathbf{0}^T & -\mathbf{x}_i^T & y_i' \mathbf{x}_i^T \\ \mathbf{x}_i^T & \mathbf{0}^T & -x_i' \mathbf{x}_i^T \end{bmatrix}. \quad (4.5)$$

最终的解即是矩阵  $\mathbf{M}$  的对应着最小的奇异值的右奇异向量。

### 4.2.2 鲁棒推广

在不存在离群点的情况下，DLT 可以给出近似最优的解。因此，一个自然的问题是，我们可以把这样一种框架推广到存在离群点的情况吗？在本节中，我们将给出一个乐观的答案。

首先，我们考虑几何模型估计的理想情况，即匹配中不存在任何的噪声。在这种情况下，我们可以考虑离群点的存在，对 DLT 做出改动：

$$\min_{\mathbf{z} \in \mathbb{R}^D} \|\mathbf{Mz}\|_0, \quad s.t. \quad \mathbf{z} \neq \mathbf{0}. \quad (4.6)$$

这一基于  $\ell_0$  的泛函  $\|\mathbf{Ez}\|_0$  的值在本质上代表着与求得的线性结构不兼容的数据点的数目。由于离群点之中不存在任何线性的结构，因此解 (4.6) 就代表着正确的估计结果。然而，对于  $\ell_0$  目标函数的最小化是几乎不可实现的，因此我们将其使用如下的  $\ell_1$  最小化的问题来替代：

$$\min_{\mathbf{z} \in \mathbb{R}^D} \|\mathbf{Mz}\|_1, \quad s.t. \quad \|\mathbf{z}\|_2 = 1. \quad (4.7)$$

在这种意义上，几何模型估计的任务变得可解，并且同时具有了抵抗噪声的能力。问题 (4.7) 与 DLT 算法之间的关系是很清楚的，即经典的 DLT 算法所给出的结果等同于在 (4.7) 中使用一种基于  $\ell_2$  损失的目标函数。这也在同时解释了 DLT 的局限性，即  $\ell_2$  损失对于离群点是极其敏感的。

在数学的概念上，问题 (4.6) 与问题 (4.7) 可以视作一种超平面拟合的问题。实际上，诸如 (4.7) 形式的问题在近年来的鲁棒子空间恢复的文献中的得到了研究 [164, 165]，其中超平面拟合可以理解为子空间维度  $d = D - 1$  的一种特例。对于这种基于  $\ell_1$  损失的目标函数，其鲁棒的性质在理论上得到了证明。简单来说，在某种对于离群点的分布的假设中，几何模型估计的任务 (4.7) 甚至可以忍受  $O(m^2)$  的离群点，其中  $m$  代表

内点的数目。

注意到 (4.7) 在本质上是一个非凸 (由于解的可行域是一个球体) 且不连续 (由于基于  $\ell_1$  的目标函数) 的问题, 因此求解算法是不平凡的。幸运的是, 对于问题 (4.7) 而言, 在现有文献中已经有一些求解算法被提出。在文章<sup>[164]</sup> 中, 问题 (4.7) 被松弛为求解一系列的线性规划的问题, 其可以保证在有限的迭代次数中收敛到全局的最优解。然而, 这一方法具有较高的计算复杂度。在文章<sup>[164]</sup> 中, 一种迭代加权最小二乘的方法被提出, 其更加高效但没有理论上的严谨保证。一种投影次梯度下降的算法在文章<sup>[165]</sup> 中被提出, 这一算法中只需要计算矩阵以及向量的乘积, 因此更加高效。

由于在几何模型估计的算法中, 降低计算的需求常常比理论上的最优保证更为重要, 我们采用投影次梯度下降的算法来解决这一问题。需要明确的是, 由于松弛策略的采用, (4.7) 的最优解往往并非是存在噪声时理想的解。实际上在文章<sup>[166]</sup> 中证明, 该问题的最优解将偏离事实上的最优解, 且偏离程度与噪声的强度成正比 (但仍可以忍受  $O(m^2)$  的离群点)。因此在这种问题下所得到的解往往是粗糙的, 并且需要后处理来进行改善质量, 关于后处理我们将在章节 4.2.4 中进行讨论。我们将所提出的基于问题 (4.7) 的几何模型估计算法总结在算法 3 中。

### 4.2.3 对于线性结构的进一步探索

虽然在问题 (4.7) 中的鲁棒的建模方式引出了许多对于几何模型估计来说有效的算法, 但在面对现实中所遇到的受到严重污染的数据来说, 仍存在一些待解决的问题。由于问题本身所具有的非凸的特性, 不能保证全局性的算法很可能被困在较差的局部最优解中, 特别是在数据被强噪声以及过多的离群点污染的情况下。这是我们对问题进行进一步的探索的动机。

#### 4.2.3.1 仿射相机中的线性结构

首先, 我们从最简单的仿射相机中来思考这一问题。如果给定的两个视图均来自于仿射相机, 那么相匹配的两个特征点将由仿射变换联系起来:

$$\mathbf{x}'_i = \mathbf{A}\mathbf{x}_i, \quad (4.8)$$

**Algorithm 3** 基于超平面拟合的几何模型估计

---

**输入:** 初始匹配集合  $\mathcal{S}$   
**输出:** 估计出的几何模型  $\mathbf{s}$

- 1: 将匹配映射为嵌入形式  $\mathbf{a}_i$  或  $\mathbf{b}_i$  以得出数据矩阵  $\mathbf{M}$ 。
- 2: 初始化  $\mathbf{z}$  为矩阵  $\mathbf{M}$  对应于最小特征值的奇异向量。
- 3: **while** 算法未收敛 **do**
- 4: 计算次梯度:  $\mathbf{g} = \mathbf{M}^T \text{sign}(\mathbf{M}\mathbf{z})$ .
- 5: 更新优化算法的步长  $\mu$ 。
- 6: 次梯度下降:  $\mathbf{z} = \mathbf{z} - \mu\mathbf{g}$ .
- 7: 球体投影:  $\mathbf{z} = \mathbf{z}/\|\mathbf{z}\|_2$ .
- 8: **end while**
- 9: 计算每个匹配相对于  $\mathbf{z}$  的残差。
- 10: 通过对残差取阈值来决定内点集  $\mathcal{I}$ 。
- 11: 在内点集  $\mathcal{I}$  中进行后处理来得到  $\mathbf{s}$ .
- 12: **return**  $\mathbf{s}$ .

---

其中

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \quad (4.9)$$

代表仿射矩阵。

与单应变换的 DLT 估计方法相似, 一个直接的利用这种结构的方法是将其转化为一个超平面拟合的问题, 其嵌入形式为:

$$\mathbf{c}_i^T = \begin{bmatrix} \mathbf{x}_i^T & \mathbf{0}^T & -x'_i \\ \mathbf{0}^T & \mathbf{x}_i^T & -y'_i \end{bmatrix}. \quad (4.10)$$

给定  $n \geq 3$  个匹配时, 这一问题可以参考问题 (4.7) 简单地得到求解, 其中数据矩阵为  $\mathbf{M} = [\mathbf{c}_1^T, \mathbf{c}_2^T, \dots, \mathbf{c}_n^T]^T$ , 而  $\mathbf{z} = [a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}, 1]^T$  编码仿射的参数。

注意到问题 (4.7) 中定义的求解方法在实际上是一种在数据受到噪声以及离群点影响的情况下子空间恢复的问题。在子空间学习的理论中, 一个广为人知的事实是相对维度  $d/D$ , 即数据的本质维度与数据本身的维度的比值, 对于子空间学习的问题来说具有非常重要的特性。从一般来说, 在相对维度较小的情况下, 子空间学习的问题将更容易求解。因此, 下面我们将提出另一种建模的方式来更小的相对维度之下利用这种线性的结构。

由于仿射变换对于匹配数据来说仅仅包含着线性的项, 我们考虑如下的嵌入形式:

$$\mathbf{d}_i = [x_i, y_i, x'_i, y'_i, 1]^T. \quad (4.11)$$

这一嵌入形式的结构在以下的由 (4.8)推导而来的公式中得到揭示:

$$\mathbf{A}'\mathbf{d}_i = 0, \quad (4.12)$$

在  $\forall i = 1, 2, \dots, n$  时成立, 其中

$$\mathbf{A}' = \begin{bmatrix} a_{11} & a_{12} & -1 & 0 & a_{13} \\ a_{21} & a_{22} & 0 & -1 & a_{23} \end{bmatrix} \quad (4.13)$$

由于  $\mathbf{A}'$  的秩显然为 2, 这意味着在我们研究的问题中 5 维的嵌入形式的数据存在于一个本质维度不超过 3 的线性子空间中。

这一点启发我们转而解决如下的一种本质维度为 3 的子空间恢复的问题:

$$\min_{\mathbf{v} \in \mathbb{R}^{D \times 2}} \sum_i \|\mathbf{d}_i^T \mathbf{v}\|_2 = \|\mathbf{M}\mathbf{v}\|_{1,2}, \quad s.t. \quad \mathbf{v}^T \mathbf{v} = \mathbf{I}, \quad (4.14)$$

其中  $\mathbf{M} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n]^T$ ,  $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2]$  代表包含两个相互正交的单位向量,  $\mathbf{I}$  代表单位矩阵, 而  $\|\cdot\|_{1,2}$  表示输入矩阵的所有行的范数的和。此时对于 (4.14)来说, 相对维度的值为  $3/5$ , 远远小于 (4.10) 所代表的超平面拟合问题  $6/7$  的相对维度。这使得问题更加容易求解。

在 (4.14) 中所包含的思路可以简述为找到内点所在的线性子空间的补空间点两个正交基底。正如在综述论文<sup>[163]</sup> 中所指出的, 这一问题可以使用较为常规的鲁棒子空间恢复的方法来解决, 例如文章<sup>[167]</sup>。在本章中, 我们采用了一种更加高效率的方法来迭代地找出两个正交的基。在第一次迭代中, 问题等同于一个超平面拟合的算法来找到第一个基底。在第二次迭代中, 算法总体上与超平面拟合的算法类似, 但多出一个投影步骤。这个额外的投影步骤保证了在第二次迭代中所找到的基底与第一个基底是正交的。具体来说, 假设我们得到了第一个基底  $\mathbf{v}_1$ , 那么补空间的投影矩阵应为  $\mathbf{I} - \mathbf{v}_1 \mathbf{v}_1^T$ , 因此第二个基底  $\mathbf{v}_2$  应该被投影为  $\mathbf{v}_2 = (\mathbf{I} - \mathbf{v}_1 \mathbf{v}_1^T) \mathbf{v}_2 = \mathbf{v}_2 - \mathbf{v}_1 \mathbf{v}_1^T \mathbf{v}_2$ 。我们将解决问题 (4.14) 的算法在算法 4 中给出。

**Algorithm 4** 基于线性嵌入的内点检测

---

**输入:** 初始匹配集合  $\mathcal{S}$

**输出:** 估计得出的内点集合  $\mathcal{I}$

- 1: 将匹配映射到嵌入形式中  $\mathbf{d}_i$  以获得数据矩阵  $\mathbf{M}$ 。
  - 2: 初始化  $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2]$  为矩阵  $\mathbf{M}$  对应于最小的特征值的右奇异向量。
  - 3: **while** 未收敛 **do**
  - 4: 计算子梯度:  $\mathbf{g}_1 = \mathbf{M}^T \text{sign}(\mathbf{M}\mathbf{v}_1)$ 。
  - 5: 更新优化的步长  $\mu$ 。
  - 6: 次梯度下降:  $\mathbf{v}_1 = \mathbf{v}_1 - \mu_i \mathbf{g}_1$ 。
  - 7: 球体投影:  $\mathbf{v}_1 = \mathbf{v}_1 / \|\mathbf{v}_1\|_2$ 。
  - 8: **end while**
  - 9: 正交投影:  $\mathbf{v}_2 = \mathbf{v}_2 - \mathbf{v}_1 \mathbf{v}_1^T \mathbf{v}_2$ 。
  - 10: 球体投影:  $\mathbf{v}_2 = \mathbf{v}_2 / \|\mathbf{v}_2\|_2$ 。
  - 11: **while** 未收敛 **do**
  - 12: 计算次梯度:  $\mathbf{g}_2 = \mathbf{M}^T \text{sign}(\mathbf{M}\mathbf{v}_2)$ 。
  - 13: 更新优化的步长  $\nu$ 。
  - 14: 次梯度下降:  $\mathbf{v}_2 = \mathbf{v}_2 - \nu_j \mathbf{g}_2$ 。
  - 15: 正交投影:  $\mathbf{v}_2 = \mathbf{v}_2 - \mathbf{v}_1 \mathbf{v}_1^T \mathbf{v}_2$ 。
  - 16: 球体投影:  $\mathbf{v}_2 = \mathbf{v}_2 / \|\mathbf{v}_2\|_2$ 。
  - 17: **end while**
  - 18: 计算每个匹配相对于  $\mathbf{v}$  的残差。
  - 19: 通过对残差取阈值来决定内点集  $\mathcal{I}$ 。
  - 20: **return**  $\mathcal{I}$ .
- 

#### 4.2.3.2 线性结构的推广

根据上面的讨论，我们可以得出结论，在仿射相机的假设下，问题将具有一种更加简单的基于嵌入形式  $\mathbf{d}_i$  的解法。因此，一个自然的问题是，算法 4 能够用于更广义的仿射相机的假设没有满足的问题中吗？下面我们尝试给出一个乐观的回答。

为了回应这个问题，第一步应该明确在广义的情境中，算法 4 的解应该具有什么特点？这一问题有着不同的解释方法，我们将在下面进行一一的解释。

第一种解释<sup>[3]</sup> 中章节 6.3 的一些重要的结论。我们使用  $\mathbf{x}_p$  表示使用一个有限投影相机测量得到的图像特征点坐标，使用  $\mathbf{x}_a$  表示虚构的另一个图像特征点，其可以理解为一个虚构的相机在无穷远处所得到的相同的 3D 点的图像坐标。则我们可以导出  $\tilde{\mathbf{x}}_p$  与  $\tilde{\mathbf{x}}_a$  由以下的公式相关联：

$$\tilde{\mathbf{x}}_a - \tilde{\mathbf{x}}_p = \frac{\Delta}{d_0} (\tilde{\mathbf{x}}_p - \tilde{\mathbf{x}}_0), \quad (4.15)$$

其中  $\mathbf{x}_0$  代表主点， $\tilde{\mathbf{x}}_a$ ,  $\tilde{\mathbf{x}}_p$  与  $\tilde{\mathbf{x}}_0$  代表非齐次的列向量坐标， $\Delta$  与  $d_0$  可以简单理解为图像场景中的深度差以及平均深度。假设我们给定  $(\mathbf{x}_p, \mathbf{x}'_p)$  作为来自通常相机的一对匹配，那么相对应的线性嵌入的形式为  $\mathbf{d} = [\tilde{\mathbf{x}}_p, \tilde{\mathbf{x}}'_p, 1]^T$ 。根据 (4.15)， $\mathbf{d}$  可以从另一种

**Algorithm 5** 基于线性嵌入形式的单应变换估计

---

**输入:** 初始匹配集合  $\mathcal{S}$

**输出:** 得到的几何模型  $s$

- 1: 在  $\mathcal{S}$  中使用算法 4 来找到内点集合  $\mathcal{I}$ 。
  - 2: 在集合  $\mathcal{I}$  中进行后处理以得到最终的几何模型的估计  $s$ .
  - 3: **return**  $s$ .
- 

角度来理解:  $\mathbf{d} = [\tilde{\mathbf{x}}_a + \epsilon, \tilde{\mathbf{x}}'_a + \epsilon', 1]^T$ , 其中  $\epsilon = \frac{\Delta}{d_0}(\tilde{\mathbf{x}}_0 - \tilde{\mathbf{x}}_p)$ ,  $\epsilon' = \frac{\Delta}{d_0}(\tilde{\mathbf{x}}'_0 - \tilde{\mathbf{x}}'_p)$  表示正比于  $\frac{\Delta}{d_0}$  的噪声项。由于  $\mathbf{x}_a$  与  $\mathbf{x}'_a$  存在一种仿射变换的关系, 并且考虑到  $\ell_1$  损失的对噪声鲁棒的性质, 只要  $\frac{\Delta}{d_0}$  足够小, 我们就可以使用算法 4 以及线性嵌入的形式来利用数据中的这一结构。在这种情况下, 假如相机距离 3D 的场景较远的情况下 ( $d_0$  较大), 大部分内点可以由算法 4 检测出来。在其他情况下, 得到的内点将一般在靠近一个平面分布 ( $\Delta$  较小), 并且可能只包含一部分的内点。

在另一种观点下, 我们可以从模型本身来解释得来的解。即便仿射变换不能完全代表场景中平面对平面的变换, 其可以理解为非线性的单应模型的线性近似。因此仿射的模型至少在局部是有效的。在这种情况下, 算法 4 可以检测得到空间中临近分布的内点。

总结上述的讨论来说, 我们可以说算法 4 至少可用来检测得到内点集合的一个子集。这是由于基于  $\ell_1$  的损失函数对于近似带来的误差是不敏感的。然而, 在内点与一个高质量的模型中仍存在着差距。在本章中, 我们将使用一种后处理的方法来找到最优的模型。这引出了如下的我们算法的扩展:

(1) 单应变换估计: 正如上文所述, 当成像的场景服从广义上的单应变换时, 算法 4 并非是理想的算法。然而, 我们可以确认至少一个内点的子集可以被检测到。在这种情况下, 我们选择使用后续的一个局部优化的步骤来找到更多的内点, 并计算得出真实的单应变换。我们将在章节 4.2.4 讨论后处理过程的细节。我们将一些直观的例子在章节 4.3.1 中给出。算法的完整步骤总结于算法 5。

(2) 基础矩阵估计: 对于基础矩阵任务而言, 仅仅具有由仿射或者单应变换所关联的一些匹配是不可能得到精确的估计的<sup>[94]</sup>。因此, 正如序贯性的 RANSAC 算法<sup>[168]</sup>, 我们提出一种检测两组内点的方法。在第一次迭代中, 我们使用算法 4 来检测一组内点。在第二次迭代中, 我们将第一组内点去除并检测第二组的内点。最终, 基础矩阵的估计可以通过两组内点的并集得到。我们将一些直观的例子在章节 4.3.1 中给出。算法的细节在算法 6 给出。

---

**Algorithm 6** 基于线性嵌入形式的基础矩阵估计

---

**输入:** 初始匹配集合  $\mathcal{S}$

**输出:** 得到的几何模型  $s$

- 1: 在  $\mathcal{S}$  中使用算法 4 来得到一个内点集合  $\mathcal{I}_1$ 。
  - 2: 从  $\mathcal{S}$  去除  $\mathcal{I}_1$  并得到集合  $\mathcal{S}'$ 。
  - 3: 在  $\mathcal{S}'$  中使用算法 4 以获得内点集合  $\mathcal{I}_2$ 。
  - 4: 在集合  $\mathcal{I}_1 \cup \mathcal{I}_2$  中利用后处理来决定最终的几何模型估计结果  $s$ .
  - 5: **return**  $s$ .
- 

#### 4.2.4 实现细节

为了提高数值稳定性，在所有的匹配数据在映射到嵌入形式后，我们将其归一化为单位模长作为算法的输入。

后处理过程：在获得了内点集合  $\mathcal{I}$  后，由于所得到的内点仍旧有可能含有一定的离群点，因此为了达到更高的准确度，我们对内点集合  $\mathcal{I}$  使用固定的 100 个采样并使用 DLT 算法来得出几何模型（单应变换，基础矩阵）的估计。每一个得出的模型均在原有的匹配数据中进行评价，并且我们对每一个目前最好的模型使用局部优化<sup>[93]</sup>。最终，我们采用具有最大的一致集的最好的模型，来作为输出的估计结果。与基于随机采样的方法类似，这种后处理的方式同样需要一个预先定义的区分内点与离群点的阈值。

### 4.3 实验结果

在本节中，我们在真实图像的几何模型估计任务中研究所提方法的性能表现。我们将所提方法命名为 SRE(subspace recovery estimator)。我们将一些直观的例子展示于图 4.1 中。

单应变换的估计：我们选取得到广泛应用的 homogr<sup>1</sup> 以及 EVD<sup>2</sup> 数据集进行实验。其中 homogr 数据集包含 16 对基线相对较窄的图像，EVD 数据集包含 15 对具有极端视角变换的图像。在这两个数据集中，均提供了一定数量的人工标注的正确匹配用于模型的评估。此外，我们也收集了 20 对遥感图像并制作了 RS 数据集。由于在遥感的任务中，成像设备距离实际的场景具有非常远的距离，图像之间可以认为服从一种仿射变换。RS 数据集的特点是极高的离群点比例（平均高于 80%），因此可以用于评

<sup>1</sup><http://cmp.felk.cvut.cz/data/geometry2view/>

<sup>2</sup><http://cmp.felk.cvut.cz/wbs/>

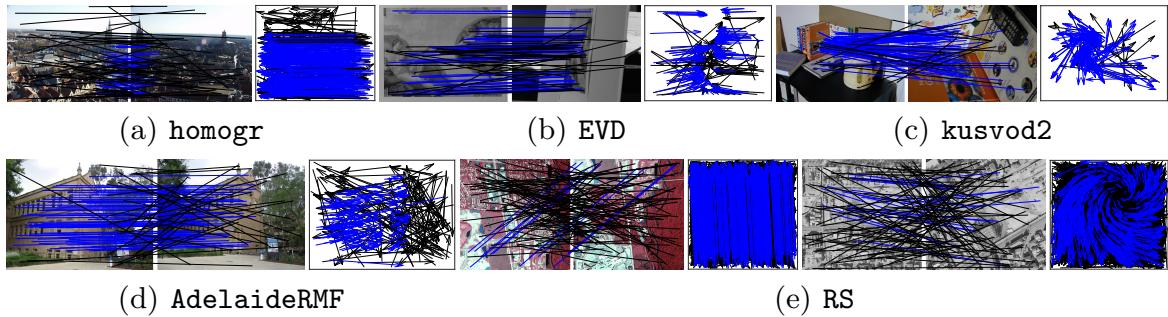


图 4.1 一些直观的应用所提 SRE 方法的例子。对于一对图像而言，出于视觉效果的考虑，我们只可视化 100 个随机挑选得出的匹配。对于每一对图像，我们也将匹配的运动场显示出来，其中箭头的头部与尾部分别代表特征点在两幅图像中的坐标。算法识别得到的内点显示为蓝色，而离群点显示为黑色。

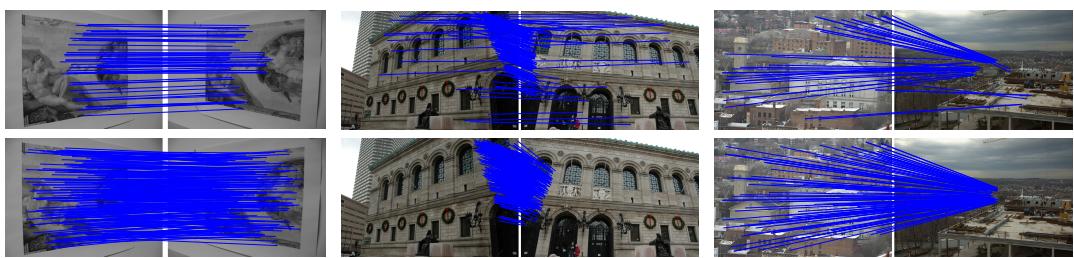


图 4.2 一些来自 homogr 数据集的单应变换估计的直观例子。第一行代表算法 4 所检测得到的内点，第二行代表后处理步骤后所得到的内点。

估方法在极端情况下的鲁棒性。这一数据集中的正确匹配（即内点）是人工标注的结果。

基础矩阵估计：对于这一任务，我们选取了得到广泛应用的 kusvod2<sup>3</sup>与 AdelaideRMF<sup>4</sup>数据集。kusvod2 数据集包含 16 对存在弱透视或者强透视效应的图像，以及一定数量的用于评估模型质量的正确匹配。AdelaideRMF 数据集包含一系列的校园建筑图像，并且给出了人工标注的匹配。我们使用该数据集的 19 对不存在多个物体的不连续运动的子集。由于相机通常距离建筑比较远，这些图像对通常具有弱透视效应。

#### 4.3.1 线性嵌入形式的定性分析

根据所利用的匹配数据嵌入形式的不同，即 (4.4), (4.5)与 (4.11)，我们所提出的 SRE 方法包含数种不同的算法。其中 (4.4)与 (4.5)可以视作经典的 DLT 算法的派生，因此理论基础较强，然而线性嵌入形式 (4.11)还需要进一步的研究分析。

我们将对于利用线性嵌入形式的单应变换估计的示例，即算法 5，展示于图 4.2 中。

<sup>3</sup><http://cmp.felk.cvut.cz/data/geometry2view/>

<sup>4</sup><https://cs.adelaide.edu.au/~hwong/doku.php?id=data>

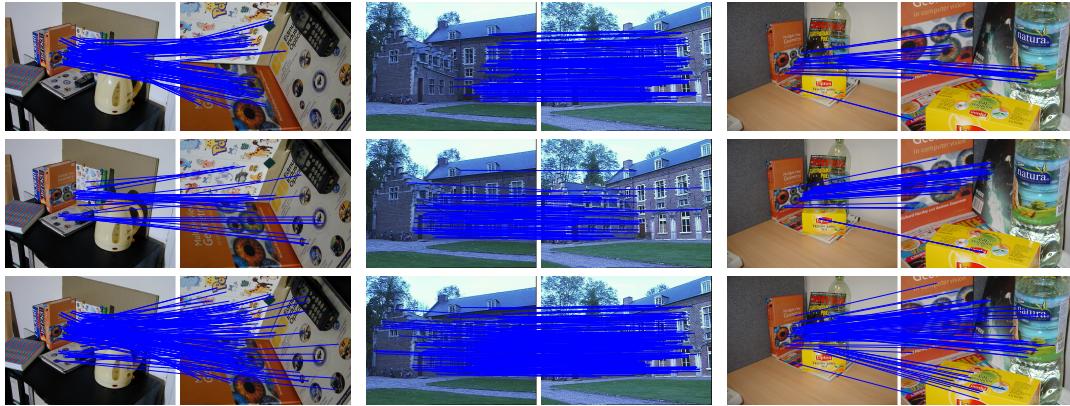


图 4.3 一些来自 kusvod2 数据集的基础矩阵估计的直观的例子。前两行代表使用算法 4 所检测得到的第一组以及第二组内点。最后一行代表后处理步骤后所得到的内点。

可以看到，所检测得到的内点可能只是真实内点的一部分，甚至包含一定数目的离群点。然而在经过包含局部优化步骤的后处理过程后，所有的内点均可以被找到，同时真实的单应变换也得到还原。我们将利用线性嵌入形式进行基础矩阵估计的直观示例，即算法 6 展示于图 4.3 中。可以看到，运行算法 4 一次所得到的内点通常分布于一个平面的周围。然而，如果迭代地使用算法 4 来找到第二组内点并将两组内点合并，则最终得到的内点集合将足够使基础矩阵得到准确的估计。值得一提的是，所检测得到的内点并不一定分布于一个自然场景物体中的平面，实际上也可能分布于一个虚拟的平面周围。

### 4.3.2 基础矩阵与单应变换的估计

在所提的 SRE 方法中，由 DLT 衍生出来的基础矩阵与单应变换估计的嵌入形式，即 (4.4) and (4.5)，可以在算法 3 中使用。而线性嵌入的形式，即 (4.11) 可以在算法 4 中得到使用。另一个值得研究的嵌入方式是 (4.10)，其可以以一种相似的方式在算法 3 使用。我们将这些变种分别记作 SRE-F, SRE-H, SRE-A 以及 SRE-At。对于 SRE-F 来说，检测得到的内点可以直接用于基础矩阵和单应变换的估计过程中。对于 SRE-H, SRE-A 以及 SRE-At，我们在单应变换估计中使用一次迭代，在基础矩阵估计中使用两次迭代。

我们将所提出的 SRE 方法与基准方法 RANSAC<sup>[85]</sup>，以及处于领先水平的 USAC<sup>[95]</sup>, MAGSAC<sup>[97]</sup> 相比较。其参数的设置参考了原文。我们将 RANSAC, MAGSAC 以及 UASC1 的最大迭代次数设置为 5,000，将 USAC2 的最大迭代次数设置为 50,000。对

于我们所提出的 SRE 方法来说，我们经验性地将算法中的门限设置为平均的残差值的 1/5。对于所有的方法来说，我们将内点与离群点的判别阈值设置为 2 个像素。

我们使用所估计的模型相对于给定的内点的平均几何误差作为度量指标。由于一个完全错误的模型通常将引入不稳定的统计信息，因此我们在计算平均误差时 (**e**) 排除了失败的测试结果并同时汇报了得出错误的模型的比例 (**f**)。我们将 Sampson distance 作为几何误差。为了判断是否是一个错误模型，我们使用 5 个像素 ( $e_1, f_1$ ) 与 10 个像素 ( $e_2, f_2$ ) 两个阈值。假如所估计的模型得出的几何误差大于这一阈值，我们将其视作一个失败的模型。对于每个方法而言，出于对随机性的考虑，我们汇报的是对于每个图像对的 100 次试验所得出的统计平均结果。

表 4.1 定量评价实验。所用的数据集，估计问题的类型，图像对的数目以及度量指标在前三列分别给出。其他列展示 100 次试验所得出的平均几何误差 (**e**) 以及失败比例 (**f**)。其中下角标 1 和 2 分别表示判断失败模型的阈值为 5 和 10 个像素的结果。平均的运行时间（以毫秒为单位）**t** 以及所有数据集中的概括性的统计信息，即 all 也在表格中给出。黑体表示最好的结果。

Alg.			RANSAC	USAC1	USAC2	MAGSAC	SRE-F	SRE-H	SRE-At	SRE-A
homogr	$H_{16}$	$e_1$	1.73	1.41	1.40	1.69	1.14	1.15	<b>1.12</b>	1.15
		$f_1$	0.055	0.141	0.100	0.210	0.078	0.163	0.148	<b>0.036</b>
		$e_2$	1.92	1.56	1.53	1.87	<b>1.31</b>	1.37	1.34	<b>1.31</b>
		$f_2$	0.016	0.116	0.077	0.181	0.046	0.124	0.108	<b>0.001</b>
		$t$	207.1	<b>20.9</b>	37.3	185.2	69.0	70.6	66.2	68.4
EVD	$H_{15}$	$e_1$	1.91	1.04	1.02	1.31	1.04	1.00	<b>0.93</b>	1.02
		$f_1$	0.255	0.734	0.733	<b>0.208</b>	0.263	0.367	0.506	0.216
		$e_2$	2.30	1.06	<b>1.02</b>	1.50	1.08	1.34	1.39	1.08
		$f_2$	0.189	0.733	0.733	<b>0.181</b>	0.259	0.334	0.471	0.208
		$t$	343.9	<b>29.3</b>	196.9	213.8	54.1	58.5	58.3	50.4
kusvod2	$F_{16}$	$e_1$	1.65	1.56	1.52	1.03	0.81	0.80	<b>0.76</b>	0.81
		$f_1$	0.143	0.094	0.085	0.151	0.139	0.132	0.137	<b>0.084</b>
		$e_2$	1.94	1.77	1.70	1.32	1.13	0.95	<b>0.85</b>	0.87
		$f_2$	0.096	0.056	<b>0.053</b>	0.109	0.094	0.112	0.125	0.076
		$t$	13.8	<b>16.4</b>	16.8	338.9	32.2	36.7	34.6	38.9
Adelaide	$F_{19}$	$e_1$	0.76	0.65	0.63	0.89	0.63	0.62	0.58	<b>0.55</b>
		$f_1$	0.002	0.001	<b>0.000</b>	0.084	0.068	0.117	0.094	<b>0.000</b>
		$e_2$	0.78	0.66	0.63	1.17	0.83	0.71	0.66	<b>0.55</b>
		$f_2$	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	0.041	0.040	0.106	0.084	<b>0.000</b>
		$t$	41.8	<b>26.2</b>	37.8	290.5	33.3	35.3	30.4	33.0
RS	$H_{20}$	$e_1$	1.65	<b>0.94</b>	1.15	1.95	1.68	2.11	1.96	2.22
		$f_1$	0.349	0.603	0.227	0.763	0.618	0.818	0.805	<b>0.085</b>
		$e_2$	2.17	<b>1.00</b>	1.18	2.90	1.96	2.26	2.23	1.59
		$f_2$	0.182	0.599	0.221	0.709	0.590	0.812	0.791	<b>0.000</b>
		$t$	945.4	<b>23.6</b>	92.7	862.0	136.2	151.1	145.2	117.4
all		$e_1$	1.48	1.12	1.13	1.25	0.98	0.94	<b>0.90</b>	0.94
		$f_1$	0.163	0.312	0.215	0.300	0.245	0.335	0.349	<b>0.079</b>
		$e_2$	1.87	1.22	1.21	1.57	1.18	1.13	1.23	<b>1.09</b>
		$f_2$	0.096	0.299	0.203	0.259	0.217	0.314	0.328	<b>0.051</b>
		$t$	330.2	<b>23.4</b>	74.2	399.5	67.2	72.9	69.3	62.5

我们将对于所有的方法的评价结果在表格 4.1 给出。从中可以看到，RANSAC 较 USAC 以及 MAGSAC 而言鲁棒性要更强，因此失败的比例较低。然而，对于几何模

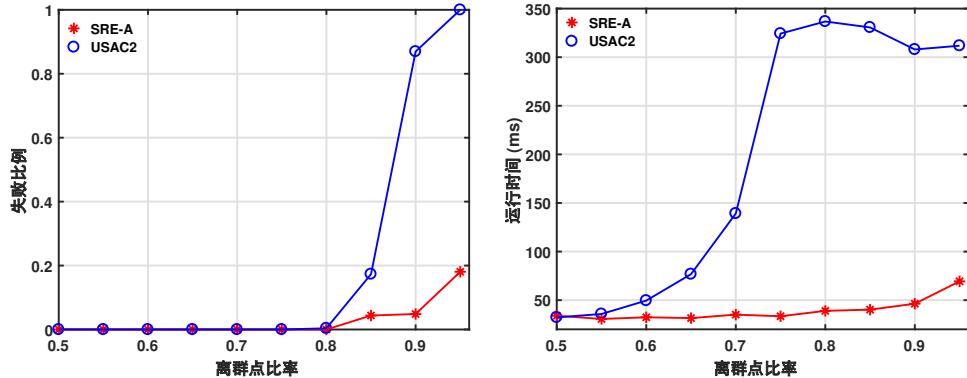


图 4.4 在合成数据中的 SRE-A 与 USAC2 的比较结果。左子图给出两个方法失败比例的结果，而右子图给出运行时间的结果。

型估计的任务来说，该方法通常准确度不够高，而且非常耗时。USAC 综合了许多随机采样方法的重要进展，因此在速度与精度上都有提高。然而其鲁棒性不够强，特别是对于具有挑战性的数据集 EVD 以及 RS 来说。MAGSAC 作为最新提出出来的一种方法，其拥有无需实现设置内点与离群点的阈值的优势。在性能方面，其在 EVD 数据集中表现良好，然而在其他的数据集中则一般，并且在具有挑战性的 RS 数据集中近乎完全失败。对于所提出的 SRE 方法的各种变种来说，其具有高效率的优势，与 USAC 相差不多。具体来说，SRE-F 与 SRE-H 要更加精确，但鲁棒性不占优势。SRE-A 的性能表现远超 SRE-At，这是由于其代表了一种更加容易求解的子空间恢复问题的建模方式。因此，总的来说，表现最好的方法是 SRE-A，其具有最高的鲁棒性，并且效率与 USAC 相当。

总结而言，SRE-A 是所提出的 SRE 方法最好的一个类型。与处于领先水平的几何模型估计算法相比，其鲁棒性以及准确性均有较大的优势。其一个重要的性质是对离群点的鲁棒性，在最具挑战性的 RS 数据集中也取得了很好的效果。

### 4.3.3 离群点比率的敏感度

对于几何模型的鲁棒估计问题而言，一个引人关注的问题是其对于离群点比率的敏感程度。为研究这一点，我们使用所有内点均得到标定的 AdelaideRMF 数据集进行实验。我们随机对两幅图中的随机点进行匹配来制造离群点，并将其与内点匹配混合来控制数据的离群点比率。在实验中，对所有方法在所有的图像对中均进行 100 次重复实验来得出统计平均的结果。我们使用 SRE-A 与最高迭代次数为 50,000 的 USAC2 进行比较。我们使用选取 5 个像素作为门限的失败比例以及算法的运行时间作为度量

指标，并将结果展示于图 4.4 中。

从图 4.4 中可以看出，所提出的 SRE-A 方法在鲁棒性以及速度方面均超过了 USAC。SRE-A 可以在存在极端的离群点比率的情况下（高达 95%）得出有效的结果，然而 USAC2 在此情况下将失效。同时，我们所提出的方法几乎不受离群点比率的影响，这一特点也是 USAC2 所不具备的，其计算复杂度随着离群点的比率指数增长。

## 4.4 本章小结

在本章中，我们研究了图像特征匹配在三维视觉中的一个重要应用，鲁棒几何模型估计。我们从几何模型的数学结构出发，由理论上对于估计算法的性质进行了分析，并给出了传统方法的鲁棒性推广。特别地，我们创新性地将  $\ell_1$  损失引入到几何模型估计这一任务中，在本质上提高了算法对离群点的抵抗能力。在求解方面，我们将问题进行了新的诠释，并与鲁棒子空间恢复理论相联系以提出一系列的高效率的求解算法。我们在鲁棒几何模型估计的公开数据集中对于本章所提算法以及处于领先水平的现有算法进行对比，结果显示本章所提出方法在鲁棒性以及效率方面的显著优势。

## 5 总结与展望

### 5.1 全文总结

图像特征匹配技术是计算机视觉领域一个基础且关键的问题，其主要的目标是检测两幅图像中的特征点并建立物理意义相同的点之间的匹配关系。在多种视觉任务中，图像特征匹配作为底层视觉通往高层视觉的纽带，以及实现信息识别与整合、从低维图像恢复高维结构的有效途径，均起着重要的作用，例如图像拼接，目标追踪，三维重建以及同步建图与定位等等。因此图像特征匹配这一课题是实现智能视觉的关键，有着重要的理论研究意义与实际应用价值。

出于计算效率与算法广泛适用性的考虑，图像的稠密匹配通常仅仅在特定的场景下使用，而图像特征匹配最常用的方式是稀疏特征点匹配，即首先在图像中检测显著而稳定的特征点，再在检测得到的两个特征点集中建立匹配关系。在进行匹配时，有着直接匹配与简化匹配两种策略。直接匹配面临的是一个极其复杂的组合优化问题，对于理想化的特征点的数目均为  $N$  的情况，其涉及到的所有可能的匹配情况可以达到  $O(N!)$  种。因此，同样出于计算效率与算法稳定性的考虑，简化匹配的策略得到了最广泛的应用，其首先对于特征点提取描述子作为局部图像信息的一种抽象，然后根据描述子之间的欧式距离建立匹配关系。然而，由于现实场景的复杂性，以这种方式所建立的匹配可能并不可靠，在绝大多数情况下会包含一定数目的离群点。本文主要对于基于简化策略的图像特征匹配技术进行研究，对于离群点问题提出了多种鲁棒算法。

在图像特征匹配问题的现有研究中，存在着多种类别的鲁棒算法对于离群点问题进行解决。这些方法所体现的主要思想大概有三类，基于全局性的平滑函数恢复，基于局部性的误匹配去除以及基于几何模型估计的随机采样算法。这三类算法不仅在实现细节上具有较大的差别，在实际的应用中也具有不同的优势。本文立足于图像特征匹配中的离群点这一问题，根据不同的思想以及理论提出了三种创新性的算法。具体来说，论文的主要贡献如下：

(1) 提出了一种基于紧凑傅里叶基表示的平滑函数恢复的方法，并将其与图像特征匹配的应用结合起来。在这一方法中，我们探索了一种新的平滑函数的表达方式，并将其在一种贝叶斯的框架下建模，使其在数据受到离群点污染的情况下仍能够得出有效的结果。所提出的基于傅里叶基的建模方式由于其紧凑性，仅需很少的一部分参数就可以对于平滑函数进行建模。相比于传统的正则化理论，该方法在不损失精度的情况下，可以将效率提升一个数量级以上。在概率模型的基础上，我们研究了一种基于 EM 算法的快速求解方法。在图像特征匹配以及点集配准的实验中，证明了所提方法的有效性，其可以作为一种解决图像特征匹配中离群点问题的方法，对图像匹配的变换关系进行非刚性的建模，适应于多种多样的复杂场景。

(2) 提出了一种基于运动一致性与概率图模型的误匹配去除的方法。在这一方法中，我们使用机器学习中最具灵活性的图模型来对于图像特征匹配问题进行建模，并进一步利用概率图模型的方式来显式地去除错误匹配。在建模过程中，我们根据图像特征匹配任务的特性，研究了一种局部运动一致性的先验，作为概率图模型中区分正确匹配与错误匹配的依据。不同于其他局部性的算法，所提出的基于概率图模型的方法使用所有局部相关性的叠加作为总的目标函数，达到了局部性与全局性的平衡，因此具有更好的效果。通过将目标函数转化为一个整数二次规划的问题，我们提出了一种快速的求解方法。在图像特征匹配的实验中，证明了所提方法的有效性，其作为一种带有局部性的算法，可以处理数据中存在多个运动模式的情形。

(3) 提出了一种基于子空间恢复的鲁棒几何模型估计算法。在这一方法中，我们考虑图像特征匹配在三维视觉中的应用，研究从数据中鲁棒地估计几何模型的问题。不同于现有的随机性或者确定性的几何模型估计算法，所提方法由优化的角度出发，给出了一种快速并且鲁棒的估计算法。具体来说，考虑几何模型的数学形式，我们将鲁棒的  $\ell_1$  损失引入进来，对于传统的 DLT 算法进行改进以实现在存在离群点的情况下 的几何模型估计。此外，我们将所转化的问题与鲁棒子空间恢复的理论联系起来，提出了一系列快速的基于次梯度下降的求解算法。我们在多种公开数据集中进行的实验表明，所提的方法在鲁棒性以及效率方面均超越了处于领先水平的算法。

## 5.2 未来展望

图像特征匹配是一个久经研究的开放性课题，其是许多视觉任务的基础，但同时也是制约这些任务的性能进一步提升的一个难题。在本文中，以一种简化的匹配策略的角度，我们对于图像特征匹配中的离群点问题进行了研究，并针对性地提出了一系列的高效率且鲁棒性的算法。本文所提出的算法涵盖了数个机器学习的不同领域，在不同层次对图像特征匹配这一问题提出了多种解决方案，并且在充分的定性与定量实验的基础上证明了所提出方法的有效性。然而由于图像特征匹配问题本身的复杂性，本文在理论上和应用层面仍存在许多的不足之处，有待于未来进一步的研究。我们将一些可能的展望方向列举如下：

(1) 如前文所述，在已经分别提取两幅图像中的特征点的前提下，存在直接匹配与间接匹配两种策略。然而对于直接匹配的策略来说，虽然其最大程度上利用了给定的数据，但由于其计算复杂度较高因此在实际中未能得到应用。简化的策略则利用描述子之间的相似度建立了一个初始匹配的集合，因而将问题转化为一个相对简单的处理离群点的问题。然而，简化的策略并不能发掘出所给定数据的全部潜力，其在简化的过程中将不可避免的失掉一部分真实匹配，这一点在许多任务，例如三维重建中是一种严重的损失。因此，要发掘数据的全部潜力，应当着眼于突破简化策略的固有限制，从而提高匹配方法的性能。

(2) 图像特征匹配是多种视觉任务的核心技术，然而这也成为其发展的一种制约。对图像特征匹配本身的研究固然是一个有价值的课题，但要扩展其影响力，将其推向更高的层次，应该与下游的视觉任务，如图像拼接、目标追踪、三维重建等相结合。本文中所提出的基于子空间恢复的鲁棒几何模型估计的方法是对这一方向的一种探索，将图像特征匹配放在三维视觉中几何信息提取的角度来考虑问题。在以后的工作中，应重点关注图像特征匹配与各种视觉任务的联系，增强其实践意义。

(3) 深度学习的兴起近年来在计算机视觉的几乎所有领域均掀起了一波革命性的浪潮，图像特征匹配也不例外。在最近的工作中，许多学者提出了一系列的针对图像特征匹配的深度学习方法，作为传统方法的替代，在一个或多个步骤中实现了对以往解决方案的性能超越。然而，由于图像特征匹配本身的复杂性，这种转变的进程还远远未达到结束的程度。因此，在未来中，使用深度学习方法作为新的技术手段，将其与应用，与传统的解决问题的思想结合起来，也是一个富有前景的方向。

(4) 本文所研究的离群点问题以图像特征匹配任务为依托，然而在实际的计算机视觉研究中，离群点问题是广泛存在的。例如在三维视觉中，点云配准，Triangulation 等任务均可能受到离群点的影响。本文所研究的方法在广义上来说，具有理论上的普适性，抛开细节层面不谈，其也可以视作解决广义的离群点问题的一种探索。因此，在后续的工作中，将所提出的解决离群点问题的思路，推广到其他的应用方向，乃至发展出一种统一的对离群点鲁棒的理论框架，也是一个极有探索意义的方向。

图像特征匹配面对真实世界的自然场景，是一个开放性的课题，同时具有较强的理论研究意义与实际应用价值。本文从该任务中涉及到的离群点问题入手，提出了一系列高效率的鲁棒算法。然而由于任务本身所具有的复杂度，图像特征匹配仍存在许多的问题有待于进一步的研究，因此图像特征匹配是一个充满活力的方向，我们相信随着时代的进一步发展，这一领域将得到更广泛的应用，从而推动智能化社会的进步。

## 参考文献

- [1] 马佳义. 基于非参数模型的点集匹配算法研究 [D]. 武汉：华中科技大学, 2014.
- [2] Ma J, Jiang X, Fan A, et al. Image matching from handcrafted to deep features: A survey[J]. International Journal of Computer Vision, 2021, 129(1) : 23–79.
- [3] Hartley R, Zisserman A. Multiple view geometry in computer vision[M]. [S.l.] : Cambridge University Press, 2003.
- [4] Pratt W K. Digital image processing john wiley & sons[J]. Inc., New York, 1991.
- [5] Viola P, Wells iii W M. Alignment by maximization of mutual information[J]. International Journal of Computer Vision, 1997, 24(2) : 137–154.
- [6] Barnea D I, Silverman H F. A class of algorithms for fast digital image registration[J]. IEEE Transactions on Computers, 1972, 100(2) : 179–186.
- [7] Zitova B, Flusser J. Image registration methods: a survey[J]. Image and vision computing, 2003, 21(11) : 977–1000.
- [8] Sotiras A, Davatzikos C, Paragios N. Deformable medical image registration: A survey[J]. IEEE Transactions on Medical Imaging, 2013, 32(7) : 1153–1190.
- [9] Baker S, Scharstein D, Lewis J, et al. A database and evaluation methodology for optical flow[J]. International Journal of Computer Vision, 2011, 92(1) : 1–31.
- [10] Hirschmuller H, Scharstein D. Evaluation of cost functions for stereo matching[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2007: 1–8.
- [11] Jin Y, Mishkin D, Mishchuk A, et al. Image matching across wide baselines: From paper to practice[J]. International Journal of Computer Vision, 2021, 129(2) : 517–547.
- [12] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2) : 91–110.
- [13] Detone D, Malisiewicz T, Rabinovich A. Superpoint: Self-supervised interest point

- detection and description[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018 : 224–236.
- [14] Szeliski R. Image alignment and stitching: A tutorial[J]. Foundations and Trends® in Computer Graphics and Vision, 2006, 2(1) : 1–104.
- [15] Sattler T, Maddern W, Toft C, et al. Benchmarking 6dof outdoor visual localization in changing conditions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018 : 8601–8610.
- [16] Schonberger J L, Frahm J-M. Structure-from-motion revisited[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016 : 4104–4113.
- [17] Mur-ortal R, Montiel J M M, Tardos J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2015, 31(5) : 1147–1163.
- [18] Yang H, Shi J, Carlone L. Teaser: Fast and certifiable point cloud registration[J]. IEEE Transactions on Robotics, 2020.
- [19] Le H M, Chin T-J, Eriksson A, et al. Deterministic approximate methods for maximum consensus robust fitting[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.
- [20] Radke R J, Andra S, Al-kofahi O, et al. Image change detection algorithms: a systematic survey[J]. IEEE Transactions on Image Processing, 2005, 14(3) : 294–307.
- [21] Zheng L, Yang Y, Tian Q. SIFT meets CNN: A decade survey of instance retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(5) : 1224–1244.
- [22] Ma J, Ma Y, Li C. Infrared and visible image fusion methods and applications: A survey[J]. Information Fusion, 2019, 45 : 153–178.
- [23] Fuentes-pacheco J, Ruiz-ascencio J, Rendón-mancha J M. Visual simultaneous localization and mapping: a survey[J]. Artificial Intelligence Review, 2015, 43(1) : 55–81.

- [24] Fan B, Kong Q, Wang X, et al. A Performance Evaluation of Local Features for Image Based 3D Reconstruction[J]. arXiv preprint arXiv:1712.05271, 2017.
- [25] Goyette N, Jodoin P-M, Porikli F, et al. Changelogdetection. net: A new change detection benchmark dataset[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2012: 1–8.
- [26] Zhang J, Wang C, Liu S, et al. Content-aware unsupervised deep homography estimation[C]. Proceedings of the European Conference on Computer Vision. 2020: 653–669.
- [27] Smeulders A W, Chu D M, Cucchiara R, et al. Visual tracking: An experimental survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 36(7): 1442–1468.
- [28] Chen J, Tian J, Lee N, et al. A partial intensity invariant feature descriptor for multimodal retinal image registration[J]. IEEE Transactions on Biomedical Engineering, 2010, 57(7): 1707–1718.
- [29] Radenović F, Tolias G, Chum O. CNN image retrieval learns from BoW: Unsupervised fine-tuning with hard examples[C]. Proceedings of the European Conference on Computer Vision. 2016: 3–20.
- [30] Chin T-J, Cai Z, Neumann F. Robust fitting in computer vision: Easy or hard?[C]. Proceedings of the European Conference on Computer Vision. 2018: 701–716.
- [31] Beaudet P R. Rotationally invariant image operators[C]. Proc. 4th Int. Joint Conf. Pattern Recog, Tokyo, Japan, 1978. 1978.
- [32] Smith S M, Brady J M. SUSAN—a new approach to low level image processing[J]. International Journal of Computer Vision, 1997, 23(1): 45–78.
- [33] Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002(4): 509–522.
- [34] Moravec H P. Techniques towards automatic visual obstacle avoidance[J], 1977.
- [35] Harris C G, Stephens M, others. A combined corner and edge detector.[C]. Alvey vision conference: Vol 15. 1988: 10–5244.

- [36] Rosten E, Drummond T. Machine learning for high-speed corner detection[C]. Proceedings of the European conference on computer vision. 2006: 430–443.
- [37] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]. Proceedings of the IEEE International Conference on Computer Vision. 2011: 2564–2571.
- [38] Lindeberg T. Feature detection with automatic scale selection[J]. International Journal of Computer Vision, 1998, 30(2): 79–116.
- [39] Bay H, Tuytelaars T, Van gool L. Surf: Speeded up robust features[C]. Proceedings of the European Conference on Computer Vision. 2006: 404–417.
- [40] Alcantarilla P F, Bartoli A, Davison A J. KAZE features[C]. Proceedings of the European conference on computer vision. 2012: 214–227.
- [41] Alcantarilla P F, Solutions T. Fast explicit diffusion for accelerated features in nonlinear scale spaces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 34(7): 1281–1298.
- [42] Schmid C, Mohr R, Bauckhage C. Evaluation of interest point detectors[J]. International Journal of Computer Vision, 2000, 37(2): 151–172.
- [43] Tuytelaars T, Mikolajczyk K, others. Local invariant feature detectors: a survey[J]. Foundations and trends® in computer graphics and vision, 2008, 3(3): 177–280.
- [44] Mukherjee D, Wu Q J, Wang G. A comparative experimental study of image feature detectors and descriptors[J]. Machine Vision and Applications, 2015, 26(4): 443–466.
- [45] Uchida Y. Local feature detectors, descriptors, and image representations: A survey[J]. arXiv preprint arXiv:1607.08368, 2016.
- [46] Krig S. Interest point detector and feature descriptor survey[G]. Computer vision metrics. [S.l.]: Springer, 2016: 187–246.
- [47] Chui H, Rangarajan A. A new point matching algorithm for non-rigid registration[J]. Computer Vision and Image Understanding, 2003, 89: 114–141.
- [48] Myronenko A, Song X. Point Set Registration: Coherent Point Drift[J]. IEEE

- Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(12): 2262–2275.
- [49] Poggio T, Torre V, Koch C. Computational vision and regularization theory[J]. Readings in Computer Vision, 1987: 638–643.
- [50] Jian B, Vemuri B C. Robust point set registration using gaussian mixture models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(8): 1633–1645.
- [51] Levi G. A note on the derivation of maximal common subgraphs of two directed or undirected graphs[J]. Calcolo, 1973, 9(4): 341.
- [52] Babai L. Groups, graphs, algorithms: The graph isomorphism problem[C]. Proc. ICM : Vol 3. 2018: 3303–3320.
- [53] Leordeanu M, Hebert M. A Spectral Technique for Correspondence Problems Using Pairwise Constraints[C]. Proceedings of the IEEE International Conference on Computer Vision. 2005: 1482–1489.
- [54] Liu H, Yan S. Common Visual Pattern Discovery via Spatially Coherent Correspondence[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2010: 1609–1616.
- [55] Zhao Q, Karisch S E, Rendl F, et al. Semidefinite programming relaxations for the quadratic assignment problem[J]. Journal of Combinatorial Optimization, 1998, 2(1): 71–109.
- [56] Almohamad H, Duffuaa S O. A linear programming approach for the weighted graph matching problem[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(5): 522–525.
- [57] Zhou F, De la Torre F. Factorized graph matching[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(9): 1774–1789.
- [58] Dym N, Maron H, Lipman Y. DS++ a flexible, scalable and provably tight relaxation for matching problems[J]. ACM Transactions on Graphics (TOG), 2017, 36(6): 1–14.
- [59] Tian Y, Yan J, Zhang H, et al. On the convergence of graph matching: Graduated

- assignment revisited[C]. Proceedings of the European Conference on Computer Vision. 2012: 821–835.
- [60] van Wyk B J, van Wyk M A. A pocs-based graph matching algorithm[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(11): 1526–1530.
- [61] Ke Y, Sukthankar R, others. PCA-SIFT: A more distinctive representation for local image descriptors[J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2004, 4: 506–513.
- [62] 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.
- [63] Dong J, Soatto S. Domain-size pooling in local descriptors: DSP-SIFT[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 5097–5106.
- [64] Calonder M, Lepetit V, Strecha C, et al. Brief: Binary robust independent elementary features[C]. Proceedings of the European Conference on Computer Vision. 2010: 778–792.
- [65] Leutenegger S, Chli M, Siegwart R Y. BRISK: Binary robust invariant scalable keypoints[M]. [S.l.]: IEEE, 2011.
- [66] Alahi A, Ortiz R, Vandergheynst P. Freak: Fast retina keypoint[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2012: 510–517.
- [67] Yang X, Cheng K-T T. Local difference binary for ultrafast and distinctive feature description[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 36(1): 188–194.
- [68] Li X, Hu Z. Rejecting Mismatches by Correspondence Function[J]. International Journal of Computer Vision, 2010, 89(1): 1–17.
- [69] Ma J, Zhao J, Tian J, et al. Robust Point Matching via Vector Field Consensus[J]. IEEE Transactions on Image Processing, 2014, 23(4): 1706–1721.
- [70] Tikhonov A N, Arsenin V I. Solutions of ill-posed problems : Vol 14[M]. [S.l.]: Vh

Winston, 1977.

- [71] Boyd S, Vandenberghe L. Convex optimization[M]. [S.l.]: Cambridge university press, 2004.
- [72] Ma J, Zhao J, Tian J, et al. Robust Estimation of Nonrigid Transformation for Point Set Registration[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 2147–2154.
- [73] Ma J, Zhao J, Tian J, et al. Regularized vector field learning with sparse approximation for mismatch removal[J]. Pattern Recognition, 2013, 46(12): 3519–3532.
- [74] Ma J, Zhao J, Ma Y, et al. Non-rigid visible and infrared face registration via regularized Gaussian fields criterion[J]. Pattern Recognition, 2015, 48(3): 772–784.
- [75] Ma J, Qiu W, Zhao J, et al. Robust  $L_2E$  Estimation of Transformation for Non-Rigid Registration[J]. IEEE Transactions on Signal Processing, 2015, 63(5): 1115–1129.
- [76] Wang G, Wang Z, Chen Y, et al. A robust non-rigid point set registration method based on asymmetric gaussian representation[J]. Computer Vision and Image Understanding, 2015, 141: 67–80.
- [77] Wang G, Wang Z, Chen Y, et al. Context-Aware Gaussian Fields for Non-rigid Point Set Registration[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 5811–5819.
- [78] Wang G, Zhou Q, Chen Y. Robust Non-Rigid Point Set Registration Using Spatially Constrained Gaussian Fields[J]. IEEE Transactions on Image Processing, 2017, 26(4): 1759–1769.
- [79] Ma J, Zhou H, Zhao J, et al. Robust Feature Matching for Remote Sensing Image Registration via Locally Linear Transforming[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(12): 6469–6481.
- [80] Ma J, Zhao J, Jiang J, et al. Non-Rigid Point Set Registration with Robust Transformation Estimation under Manifold Regularization[C]. AAAI. 2017: 4218–4224.

- [81] Ma J, Zhao J, Jiang J, et al. Locality preserving matching[J]. International Journal of Computer Vision, 2019, 127(5): 512–531.
- [82] Bian J, Lin W-Y, Matsushita Y, et al. Gms: Grid-based motion statistics for fast, ultra-robust feature correspondence[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4181–4190.
- [83] Lin W-Y, Wang F, Cheng M-M, et al. CODE: Coherence based decision boundaries for feature correspondence[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(1): 34–47.
- [84] Lipman Y, Yagev S, Poranne R, et al. Feature matching with bounded distortion[J]. ACM Transactions on Graphics (TOG), 2014, 33(3): 1–14.
- [85] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 6(6): 381–395.
- [86] Ni K, Jin H, Dellaert F. Groupsac: Efficient consensus in the presence of groupings[C]. Proceedings of the IEEE International Conference on Computer Vision. 2009: 2193–2200.
- [87] Chum O, Matas J. Matching with PROSAC - Progressive Sample Consensus[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2005: 220–226.
- [88] Torr P H, Nasuto S J, Bishop J M. Napsac: High noise, high dimensional robust estimation-it's in the bag[C]. Proceedings of the British Machine Vision Conference (BMVC). 2002.
- [89] Chum O, Matas J. Optimal randomized RANSAC[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(8): 1472–1482.
- [90] Matas J, Chum O. Randomized RANSAC with Td, d test[J]. Image and Vision Computing, 2004, 22(10): 837–842.
- [91] Torr P H, Zisserman A. MLESAC: A new robust estimator with application to estimating image geometry[J]. Computer Vision and Image Understanding, 2000, 78(1): 138–156.

- [92] Chum O, Matas J, Kittler J. Locally optimized RANSAC[C]. Proceedings of the Joint Pattern Recognition Symposium. 2003 : 236 – 243.
- [93] Lebeda K, Matas J, Chum O. Fixing the locally optimized ransac–full experimental evaluation[C]. Proceedings of the British Machine Vision Conference. 2012 : 1 – 11.
- [94] Chum O, Werner T, Matas J. Two-view geometry estimation unaffected by a dominant plane[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2005 : 772 – 779.
- [95] Raguram R, Chum O, Pollefeys M, et al. USAC: a universal framework for random sample consensus[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(8) : 2022 – 2038.
- [96] Barath D, Matas J. Graph-cut ransac[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018 : 6733 – 6741.
- [97] Barath D, Matas J, Noskova J. MAGSAC: marginalizing sample consensus[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019 : 10197 – 10205.
- [98] Barath D, Noskova J, Ivashechkin M, et al. MAGSAC++, a fast, reliable and accurate robust estimator[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020 : 1304 – 1312.
- [99] Campbell D, Petersson L, Kneip L, et al. Globally-optimal inlier set maximisation for simultaneous camera pose and feature correspondence[C]. Proceedings of the IEEE International Conference on Computer Vision. 2017 : 1 – 10.
- [100] Li H. Consensus set maximization with guaranteed global optimality for robust geometry estimation[C]. Proceedings of the IEEE International Conference on Computer Vision. 2009 : 1074 – 1080.
- [101] Bazin J-C, Seo Y, Hartley R, et al. Globally optimal inlier set maximization with unknown rotation and focal length[C]. Proceedings of the European Conference on Computer Vision. 2014 : 803 – 817.
- [102] Bazin J-C, Li H, Kweon I S, et al. A branch-and-bound approach to correspon-

- dence and grouping problems[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(7) : 1565–1576.
- [103] Cai Z, Chin T-J, Koltun V. Consensus Maximization Tree Search Revisited[C]. Proceedings of the IEEE International Conference on Computer Vision. 2019 : 1637–1645.
- [104] Enqvist O, Ask E, Kahl F, et al. Tractable algorithms for robust model estimation[J]. International Journal of Computer Vision, 2015, 112(1) : 115–129.
- [105] Purkait P, Zach C, Eriksson A. Maximum Consensus Parameter Estimation by Reweighted  $\ell_1$  Methods[C]. Proceedings of the International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition. 2017 : 312–327.
- [106] Cai Z, Chin T-J, Le H, et al. Deterministic consensus maximization with biconvex programming[C]. Proceedings of the European Conference on Computer Vision. 2018 : 685–700.
- [107] Verdie Y, Yi K, Fua P, et al. Tilde: A temporally invariant learned detector[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015 : 5279–5288.
- [108] Zhang X, Yu F X, Karaman S, et al. Learning discriminative and transformation covariant local feature detectors[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017 : 6818–6826.
- [109] Savinov N, Seki A, Ladicky L, et al. Quad-networks: unsupervised learning to rank for interest point detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017 : 1822–1830.
- [110] Barroso-laguna A, Riba E, Ponsa D, et al. Key. net: Keypoint detection by hand-crafted and learned cnn filters[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019 : 5836–5844.
- [111] Simo-serra E, Trulls E, Ferraz L, et al. Discriminative learning of deep convolutional feature point descriptors[C]. Proceedings of the IEEE International Conference on Computer Vision. 2015 : 118–126.

- [112] Balntas V, Riba E, Ponsa D, et al. Learning local feature descriptors with triplets and shallow convolutional neural networks.[C]. Proceedings of the British Machine Vision Conference : Vol 1. 2016 : 3.
- [113] Tian Y, Fan B, Wu F. L2-net: Deep learning of discriminative patch descriptor in euclidean space[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 661–669.
- [114] Mishchuk A, Mishkin D, Radenovic F, et al. Working hard to know your neighbor's margins: Local descriptor learning loss[J]. arXiv preprint arXiv:1705.10872, 2017.
- [115] Tian Y, Yu X, Fan B, et al. Sosnet: Second order similarity regularization for local descriptor learning[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 11016–11025.
- [116] Ebel P, Mishchuk A, Yi K M, et al. Beyond cartesian representations for local descriptors[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 253–262.
- [117] Luo Z, Shen T, Zhou L, et al. Geodesc: Learning local descriptors by integrating geometry constraints[C]. Proceedings of the European Conference on Computer Vision. 2018: 168–183.
- [118] Luo Z, Shen T, Zhou L, et al. Contextdesc: Local descriptor augmentation with cross-modality context[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2527–2536.
- [119] Yi K M, Trulls E, Lepetit V, et al. Lift: Learned invariant feature transform[C]. Proceedings of the European Conference on Computer Vision. 2016: 467–483.
- [120] Noh H, Araujo A, Sim J, et al. Large-scale image retrieval with attentive deep local features[C]. Proceedings of the IEEE international Conference on Computer Vision. 2017: 3456–3465.
- [121] Ono Y, Trulls E, Fua P, et al. LF-Net: Learning local features from images[J]. arXiv preprint arXiv:1805.09662, 2018.
- [122] Dusmanu M, Rocco I, Pajdla T, et al. D2-net: A trainable cnn for joint detection and description of local features[J]. arXiv preprint arXiv:1905.03561, 2019.

- [123] Revaud J, Weinzaepfel P, De souza C, et al. R2D2: repeatable and reliable detector and descriptor[J]. arXiv preprint arXiv:1906.06195, 2019.
- [124] Sarlin P-E, Detone D, Malisiewicz T, et al. SuperGlue: Learning feature matching with graph neural networks[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 4938–4947.
- [125] Yu T, Wang R, Yan J, et al. Learning deep graph matching with channel-independent embedding and hungarian attention[C]. International Conference on Learning Representations. 2019.
- [126] Fey M, Lenssen J E, Morris C, et al. Deep graph matching consensus[J]. arXiv preprint arXiv:2001.09621, 2020.
- [127] Wang R, Yan J, Yang X. Learning combinatorial embedding networks for deep graph matching[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 3056–3065.
- [128] Rolínek M, Swoboda P, Zietlow D, et al. Deep graph matching via blackbox differentiation of combinatorial solvers[C]. Proceedings of the European Conference on Computer Vision. 2020: 407–424.
- [129] Brachmann E, Krull A, Nowozin S, et al. Dsac-differentiable ransac for camera localization[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 6684–6692.
- [130] Brachmann E, Rother C. Neural-guided RANSAC: Learning where to sample model hypotheses[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 4322–4331.
- [131] Ranftl R, Koltun V. Deep fundamental matrix estimation[C]. Proceedings of the European conference on computer vision (ECCV). 2018: 284–299.
- [132] Qi C R, Su H, Mo K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 652–660.
- [133] Yi K M, Trulls E, Ono Y, et al. Learning to find good correspondences[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

2018: 2666–2674.

- [134] Zhang J, Sun D, Luo Z, et al. Learning two-view correspondences and geometry using order-aware network[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 5845–5854.
- [135] Sun W, Jiang W, Trulls E, et al. Acne: Attentive context normalization for robust permutation-equivariant learning[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11286–11295.
- [136] 赵键. 点模式匹配算法研究 [D]. 长沙: 国防科学技术大学, 2012.
- [137] 宋智礼. 图像配准技术及其应用的研究 [D]. 上海: 复旦大学, 2010.
- [138] 于伟. 基于卷积神经网络特征的图像匹配研究 [D]. 哈尔滨: 哈尔滨工业大学, 2017.
- [139] 柳成荫. 基于点特征的多模与多视角图像非刚性配准算法研究 [D]. 武汉: 华中科技大学, 2018.
- [140] Girosi F, Jones M, Poggio T. Regularization theory and neural networks architectures[J]. Neural Comput., 1995, 7(2): 219–269.
- [141] Evgeniou T, Pontil M, Poggio T. Regularization networks and support vector machines[J]. Advances in computational mathematics, 2000, 13(1): 1–50.
- [142] Ma J, Zhao J, Tian J, et al. Regularized vector field learning with sparse approximation for mismatch removal[J]. Pattern Recognit., 2013, 46(12): 3519–3532.
- [143] Grebenkov D S, Nguyen B-T. Geometrical structure of Laplacian eigenfunctions[J]. SIAM Review, 2013, 55(4): 601–667.
- [144] Courant R, Hilbert D. Methods of Mathematical Physics: Partial Differential Equations[M]. [S.l.]: John Wiley & Sons, 2008.
- [145] Reed M. Methods of modern mathematical physics: Functional analysis[M]. [S.l.]: Elsevier, 2012.
- [146] Sullivan T J. Introduction to uncertainty quantification: Vol 63[M]. [S.l.]: Springer, 2015.
- [147] Jiang X, Jiang J, Fan A, et al. Multiscale Locality and Rank Preservation for Robust Feature Matching of Remote Sensing Images[J]. IEEE Transactions on

- Geoscience and Remote Sensing, 2019, 57(9) : 6462–6472.
- [148] Koller D, Friedman N, Bach F. Probabilistic graphical models: principles and techniques[M]. [S.l.] : MIT press, 2009.
- [149] Frank M, Wolfe P. An algorithm for quadratic programming[J]. Naval Research Logistics Quarterly, 1956, 3(1-2) : 95–110.
- [150] Jaggi M. Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization.[C]. Proceedings of the International Conference on Machine Learning. 2013 : 427–435.
- [151] Goldstein A A. On steepest descent[J]. Journal of the Society for Industrial and Applied Mathematics, Series A: Control, 1965, 3(1) : 147–151.
- [152] Bentley J L. Multidimensional binary search trees used for associative searching[J]. Commun. ACM, 1975, 18(9) : 509–517.
- [153] Wong H S, Chin T-J, Yu J, et al. Dynamic and hierarchical multi-structure geometric model fitting[C]. Proceedings of the International Conference on Computer Vision. 2011 : 1044–1051.
- [154] Bian J-W, Wu Y-H, Zhao J, et al. An evaluation of feature matchers for fundamental matrix estimation[J]. arXiv preprint arXiv:1908.09474, 2019.
- [155] Sturm J, Engelhard N, Endres F, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems. 2012 : 573–580.
- [156] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2012 : 3354–3361.
- [157] Knapitsch A, Park J, Zhou Q-Y, et al. Tanks and temples: Benchmarking large-scale scene reconstruction[J]. ACM Transactions on Graphics (ToG), 2017, 36(4) : 1 – 13.
- [158] Wilson K, Snavely N. Robust global translations with 1dsfm[C]. Proceedings of the European Conference on Computer Vision. 2014 : 61–75.
- [159] Zhang Z. Determining the epipolar geometry and its uncertainty: A review[J].

- International Journal of Computer Vision, 1998, 27(2) : 161 – 195.
- [160] Baeza-yates R, Ribeiro-neto B, others. Modern information retrieval : Vol 463[M]. [S.l.] : ACM press New York, 1999.
- [161] Angeli A, Filliat D, Doncieux S, et al. Fast and incremental method for loop-closure detection using bags of visual words[J]. IEEE Transactions on Robotics, 2008, 24(5) : 1027–1037.
- [162] Burri M, Nikolic J, Gohl P, et al. The EuRoC micro aerial vehicle datasets[J]. International Journal of Robotics Research, 2016, 35(10) : 1157–1163.
- [163] Lerman G, Maunu T. An overview of robust subspace recovery[J]. Proceedings of the IEEE, 2018, 106(8) : 1380–1410.
- [164] Tsakiris M C, Vidal R. Dual principal component pursuit[J]. Journal of Machine Learning Research, 2018, 19(1) : 684–732.
- [165] Zhu Z, Wang Y, Robinson D, et al. Dual principal component pursuit: Improved analysis and efficient algorithms[C]. Advances in Neural Information Processing Systems. 2018 : 2171 – 2181.
- [166] Ding T, Zhu Z, Ding T, et al. Noisy dual principal component pursuit[C]. Proceedings of the International Conference on Machine Learning. 2019 : 1617 – 1625.
- [167] Lerman G, Mccoy M B, Tropp J A, et al. Robust computation of linear models by convex relaxation[J]. Foundations of Computational Mathematics, 2015, 15(2) : 363 – 410.
- [168] Vincent E, Lagani  re R. Detecting planar homographies in an image pair[C]. Proceedings of the International Symposium on Image and Signal Processing and Analysis. 2001 : 182 – 187.

## 攻读硕士学位期间发表的学术论文

1. Fan Aoxiang, Jiang Xingyu, Wang Yang, Jiang Junjun, Ma Jiayi. Geometric Estimation via Robust Subspace Recovery[C]. European Conference on Computer Vision(**ECCV**), 2020, DOI : 10.1007/978-3-030-58542-6\_28.(CCF B 类, 计算机视觉领域顶级会议, 对应本文第四章)
2. Ma, Jiayi and Jiang, Xingyu, **Fan, Aoxiang** and Jiang, Junjun and Yan, Junchi. Image Matching from Handcrafted to Deep Features: A Survey[J]. International Journal of Computer Vision(**IJCV**), 2020, DOI : 10.1007/s11263-020-01359-2. (SCI 一区, IF=5.698, 计算机视觉领域顶级期刊)
3. Jiang Xingyu, Jiang Junjun, **Fan Aoxiang**, Wang Zhongyuan, Ma Jiayi. Multi-scale Locality and Rank Preserving for Robust Feature Matching of Remote Sensing Images[J]. IEEE Transactions on Geoscience and Remote Sensing(**TGRS**), 2019, DOI : 10.1109/TGRS.2019.2906183. (SCI 二区, IF=5.855)
4. Jiang Xingyu, Ma Jiayi, **Fan Aoxiang**, Xu Haiping, Lin Geng, Lu Tao, Tian Xin. Robust Feature Matching for Remote Sensing Image Registration via Linear Adaptive Filtering[J]. IEEE Transactions on Geoscience and Remote Sensing(**TGRS**), 2019, DOI : 10.1109/TGRS.2019.2906183. (SCI 二区, IF=5.855)
5. **Fan Aoxiang**, Jiang Xingyu, Ma Yong, Mei Xiaoguang, Ma Jiayi. Smoothness-Driven Consensus Based on Compact Representation for Robust Feature Matching. IEEE Transactions on Neural Networks and Learning Systems(**TNNLS**), 2021. (二审中, SCI 一区, IF=8.793, 对应本文第二章)
6. **Fan Aoxiang**, Ma Jiayi, Ling Haibin. Efficient Deterministic Search with Robust Loss Functions for Geometric Model Fitting. IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**), 2021. (二审中, SCI 一区, IF=17.861, 计算机视觉领域顶级期刊)

## 致    谢

在完成硕士学位毕业论文的此时此刻，不觉已经是在武大求学的第七年。七年的时间里，我在武大经历了人生中不可多得的一段缤纷多彩的时光，令我从傻傻的少年时期迈入了青春年代的末尾，同时也来到了一段更广阔的旅程的开始。如今，身在电子信息学院东科技楼的实验室中，回想起刚刚过去的三年硕士生涯，真诚地觉得自己取得了超乎预料的收获，心中也充满了谢意。

作为我攻读硕士学位以及踏上科研之路最重要的人物，首先我特别需要感谢的是我的导师马佳义教授。马老师是一位年轻有为的学者，仅仅长我十岁，但除此之外，他更是一个真正意义上为学生着想的教师职业的几近完美的代表。在本科临近毕业之时，面临保研导师选择的我，凭借为数不多的经验以及一种微妙的直觉首先联系了马老师，并迅速得到了正面的回应。在如今看来，当时的选择实在是得到了运气的眷顾。在其时我对科研有一种涉世未深时理想化的想象，然而在马老师这里，我竟真正觉得曾经的幻想，曾经发自内心想要得到的生活得到了实现。马老师衷心为学生考虑，一直努力给予我们一种自由的学习环境，并以身作则，用自身严谨的治学态度以及忘我的工作热情给我们做出榜样。最终，作为马老师的弟子，我们得益于组内一种自由愉快且不断进取的氛围，都获得了长足的进步，为今后踏入科研道路或者社会工作打下了坚实的基础。如今的我选择继续攻读博士学位，以及未来可能的为取得教职而努力的目标，也得益于马老师的激励。在此希望马老师阖家幸福，事业继续高升，做出更多有影响力的成绩。

特别感谢实验室马泳教授、黄珺老师、梅晓光老师、樊凡老师、周华兵老师、李皞老师，他们与马老师一起是多谱视觉处理实验室的存在基础，组成了一个奋发向上、朝气蓬勃的科研团队，令无数学生从中得到了巨大的成长。感谢他们为我们所身处的科研环境所做的贡献，希望各位老师事业蒸蒸日上，阖家幸福美满。特别感谢图森未来科技有限公司的赵季老师。赵季老师是马老师在博士期间亲密合作的战友，在和他相处的过程中，我获得了许多科研上以及人生中的经验，增长了见识也得到了充分的锻炼，希望赵老师事业顺利，在未来能够实现自己所定下的目标。

感谢同课题组一起学习科研、一起玩耍的全体同学们，在和大家的相处过程中，真实感受到了身处一个友爱的集体的开心与快乐，也是他们教会了我诚以待人，教会了我不必压抑自己，释放自我。在这里，我十分幸运结识了同组的非常优秀的来来往往的许多人，高一级的最早做出优秀科研成果的余威、梁鹏伟师兄，篮球高手兼开山博士大弟子彭程里师兄，实验室的元老级灵魂人物，十分爱护师弟师妹的蒋兴宇带师兄；同一年级的优雅有气质的学姐王歆雅，一如既往的优秀的学姐徐涵和爱笑又洒脱的潘尔婷；次一级的是科研能手但不太老实的张浩，友爱但油腻的乐朱亮，自信但憨憨的王阳，和两位可可爱爱的师妹，东风千金张恺宁以及爱挑事的高文静；以及再次一级的总是欺负师兄的唐霖峰，不是很胖的公子李梓卓以及卷王大佬邓宇新。虽然我记性不太好，但是和大家吃饭吹牛、打羽毛球篮球乒乓球网球、玩密室逃脱狼人杀吃鸡、出去踏青，以及占用时间最多的在实验室一起学习科研的时光，将被我好好保存在记忆中不会忘记。祝愿大家前程似锦天天开心！

最后，感谢这么多年来一直爱护我支持我的亲人们。感谢我的父母，他们虽然文化程度不高但是非常关心我的成长，尽最大努力为我提供一个好的学习环境。感谢我的姥姥姥爷以及姨母舅舅们，他们善良可爱，努力工作，同样是我最亲的亲人。感谢我很少叫姐姐的姐姐，我的同样很帅的弟弟，以及许多其他的弟弟妹妹们，他们是我最信任的人。由衷希望我的家人健康快乐，一切安好。

虽然前路仍存在许多未知，但我确定我将一直充满希望地乐观地走下去。

范翱翔 2021年04月于武汉大学

# 武汉大学学位论文使用授权协议书

(一式两份，一份论文作者保存，一份留学校存档)

本学位论文作者愿意遵守武汉大学关于保存、使用学位论文的管理办法及规定，即：学校有权保存学位论文的电子版和印刷本，并提供文献检索与阅览服务；学校可以采用影印、缩印、数字化或其它复制手段保存论文；在以教学与科研服务为目的的前提下，学校可以向本校认证用户公布部分及全部内容。

- 1、 在本论文提交当年，同意向本校认证用户以及中国高等教育文献保障系统（CALIS）之高校学位论文系统提供查询及前十六页浏览服务。
- 2、 在本论文提交口当年/口一年/口两年/口三年以后，允许本校认证用户在线浏览全文，学校可以为存在馆际合作关系的图书馆提供全文文献传递服务。

论文作者（签名）：范翱翔

学 号：2018202120098

学 院：电子信息学院

日期：2021 年 5 月 18 日