GNSS拒止下的无人机视觉辅助融合导航定位技术研究

作者姓名 吴永云

指导教师姓名、职称 刘贵喜 教授

申请学位类别 工学硕士

学校代码 10701　　学　号 20041211913
分 类 号 TN82　　密　级 公开

西安电子科技大学

硕士学位论文

GNSS 拒止下的无人机视觉辅助融合导航定位技术研究

作者姓名：吴永云

一级学科：控制科学与工程

二级学科（研究方向）：控制理论与控制工程

学位类别：工学硕士

指导教师姓名、职称：刘贵喜　教授

学　　院：机电工程学院

提交日期：2023 年 5 月

Research on Vision-Aided Navigation and Localization Technology for UAV Under GNSS Denial

A thesis submitted to

XIDIAN UNIVERSITY

in partial fulfillment of the requirements

for the degree of Master

in Control Science and Engineering

By

Wu Yongyun

Supervisor: Liu Guixi Title: Professor

May 2023

摘要

高精度的定位与导航是实现无人机自主飞行、高效侦查与精确打击的关键技术之一，现有的无人机导航系统大多采用惯性导航系统（INS，Inertial Navigation System）与全球卫星导航系统（GNSS，Global Navigation Satellite System）组合导航的方式，但是在赛博限制或电子战攻击中，GNSS 无法稳定连接从而进入拒止状态，此时仅能依靠 INS 导航，但其存在如随机游走、零偏不稳定等性质，因此长时间工作下累计误差不可忽略。为了保障无人机在 GNSS 拒止条件下的自主可靠飞行，基于图像的视觉辅助导航方案得到了广泛研究，其有着设备结构简单、信息来源稳定、定位精度高、潜在风险小等优点。

然而视觉辅助无人机导航定位方案仍存在诸多问题和挑战，例如：1）受时间变化影响，无人机飞行过程中实时采集的地面影像与预先搭载的数字影像地图必然存在较大差异；2）无人机导航要求具备时效性，这给视觉定位算法提出了较高要求；3）基于图像匹配的视觉位姿结果并不具备连续性，错误匹配关系和明显振荡的位姿数据对无人机的定位精度和鲁棒性带来了极大影响。基于此，本文进行了深入分析和研究，主要开展了以下几个方面的工作：

首先，针对数字影像地图和无人机实时采集的地面影像差异大的问题，本研究提出了两种基于点特征的异源影像匹配方法。根据两幅异源图像间显著视觉特征的比例，本文将异源图像分为两类，在异源图像视觉特征显著的情况下，本文采用了特征提取和匹配分离的方案，通过提取历史遥感影像地图的先验特征并进行稀疏增强的策略大大减少了图像匹配所需的时间，其具有速度快精度高的优势；在异源图像视觉特征不显著的情况下，本文采用了密集特征匹配的方式，通过构造四维空间来搜索一致的匹配模式，能够在特征稀缺或特征重复度高的异源图像间实现稳定匹配。

其次，研究了基于匹配点对的无人机位姿估计方案，针对图像匹配完成后的结果，通过单应性矩阵进行约束，以筛选掉异源图像间不在同一平面上的匹配点对，在此基础上构建 PnP-MAGSAC 模型，将筛选后的匹配点对在模型上进行最小二乘迭代，获取最优模型从而解算无人机的视觉位姿估计结果。此外，考虑到低空无人机视野范围狭小视觉定位易失败的问题，文本还提出了一种高低空协同导航的方案，以保障低空无人机的长时间稳定定位。

最后，本研究提出了一种基于卡尔曼的多源信息融合导航方案，以解决在 PnP 模型下，图像匹配的像素级误差因为相机距地高度而被放大导致的视觉定位位姿结果明显振荡问题。相比于仅使用视觉位姿估计，融合导航方案能够显著提升定位的精度，减少纯视觉位姿解算的不定性，保障 GNSS 拒止情况下无人机长时间导航的稳定性。

关 键 词：GNSS 拒止，视觉定位，无人机导航，异源图像匹配，多传感器融合

ABSTRACT

High-precision positioning and navigation is one of the key technologies for achieving autonomous flight, efficient reconnaissance, and accurate strikes of UAV. Existing navigation systems mostly use a combination of inertial navigation system (INS) and global navigation satellite system (GNSS) for navigation. However, in cyber restrictions or electronic warfare attacks, GNSS loses its signal and become unreliable. In such cases, UAV can only rely on INS for navigation, but INS has certain characteristics such as random walk and unstable zero drift, and cumulative errors cannot be ignored over long periods of work. To ensure the autonomous and reliable flight of UAV under GNSS denial conditions, a vision-based navigation scheme based on image matching is a powerful auxiliary tool.

However, there are still many problems and challenges in visual assisted UAV navigation and positioning solutions, such as: 1) Due to the influence of time changes, there must be significant differences between the real-time ground images collected during UAV flight and the pre loaded digital image maps; 2) UAV navigation requires real-time, which puts higher demands on visual positioning algorithms; 3) The visual pose results based on image matching do not have continuity and are limited by constantly changing ground information. Image matching is not always successful, and mis-matching point pairs and obviously oscillating pose estimation results greatly affect the positioning and navigation accuracy and robustness of the UAV. Based on this, this dissertation conducted in-depth analysis and research, mainly focusing on the following aspects:

Firstly, to address the problem of large differences between digital image maps and ground images captured in real time by unmanned aerial vehicles (UAVs), this study proposes two point feature-based methods for cross-source image matching. Based on the proportion of significant visual features between two cross-source images, this study categorizes them into two types. For cross-source images with significant visual features, a feature extraction and matching separation-based sparse feature enhancement scheme is adopted to greatly reduce the time required for image matching by using prior feature extraction and sparse feature enhancement strategies from historical remote sensing image maps, thus having the advantages of fast speed and high accuracy. For cross-source images without significant visual features, a dense feature matching approach is used to achieve stable matching

between cross-source images with sparse features or high feature redundancy by searching for neighborhood consistency patterns.

Next, this study investigated an unmanned aerial vehicle (UAV) pose estimation method based on matched point pairs obtained from image matching. By constraining the matched point pairs between heterogeneous images that are not on the same plane using a homography matrix, the PnP-MAGSAC model was built to perform least-squares iterations on the selected matched point pairs to obtain the optimal model and thereby obtain the UAV's visual pose estimation result. In addition, considering the problem of narrow visual positioning range and easy failure in low-altitude unmanned aerial vehicles, this study also proposes a high-low altitude cooperative navigation solution to ensure stable navigation and positioning of low-altitude unmanned aerial vehicles over a long period of time.

Finally, although the accuracy of image matching can reach the pixel level, the slight error in image matching under the 3D PnP model can also be amplified due to the camera-to-ground height. To address this issue, this study proposes a multi-source information fusion navigation solution based on the Kalman filter, utilizing INS measurement data and altimeter data. Compared to using only visual pose estimation, fusion navigation can significantly improve positioning accuracy, reduce the oscillation of pure visual pose estimation, and ensure the stability of long-term navigation of unmanned aerial vehicles in the absence of GNSS.

Keywords: GNSS Denial, Visual Localization, UAV Navigation, Heterogeneous Image Matching, Multi-source Sensor Fusion

插图索引

表格索引

符号对照表

符号	符号名称
$O_w - X_w Y_w Z_w$	世界坐标系
$O_b - X_b Y_b Z_b$	机体坐标系
$O_c - X_c Y_c Z_c$	相机坐标系
$O_1 - XY$	成像平面坐标系
$O_2 - UV$	图像坐标系
ψ	偏航角
φ	俯仰角
θ	滚转角
f_x	相机 x 方向焦距
f_y	相机 y 方向焦距
d_x	相机 x 方向偏移量
d_y	相机 y 方向偏移量
$\mathbf{K}$	相机内参矩阵
$\mathbf{R}$	相机位姿旋转矩阵
$\mathbf{T}$	相机位姿平移矩阵
$H(\cdot)$	单应性变换矩阵
N_h	重定位帧间隔数

缩略语对照表

缩略语	英文全称	中文对照
AGNN	Attentional Graph Neural Network	图注意力神经网络
CNN	Convolutional Neural Network	卷积神经网络
DLT	Direct Linear Transformation	直接线性变换
GNSS	Global Navigation Satellite System	全球卫星导航系统
GRD	Ground Range Detected	地距多视影像
IMU	Inertial Measurement Unit	惯性测量单元
INS	Inertial Navigation System	惯性导航系统
kNN	k-Nearest Neighbor	最近邻
MAE	Mean Absolute Error	平均绝对误差
MLP	Multilayer Perceptrons	多层感知机
MMA	Mean Matching Accuracy	匹配准确率
MT	Matching Time	匹配时间
NCNet	Neighbourhood Consensus Networks	邻域共识网络
NMS	Non-Maximum Suppression	非极大值抑制
RANSAC	Random Sample Consensus	随机一致性抽样
RMSE	Root Mean Squard Error	均方根误差
SAR	Synthetic Aperture Radar	合成孔径雷达
SIFT	Scale-invariant Feature Transform	尺度不变特征转换
SLAM	Simultaneous Localization And Mapping	同步定位与建图
SLC	Single Look Complex	单视复数影像
SLEA	Spot Extended Area	光斑扩展区域
SSC	Suppression via Square Covering	方形抑制
UAV	Unamanned Aerial Vehicle	无人机

目录

第一章　绪论

1.1 研究背景与研究意义

21 世纪以来，无人机相关产业蓬勃发展，基于无人机的多种优点，如体积小、成本低、灵活度高、安全风险系数小等，其在各领域都得到了广泛应用。美国国防部至今已发布 8 版《无人系统综合路线图》，路线图中指出未来无人机应具备自主飞行、高效侦察与精确打击能力[1]。最新版中又提到无人系统包括无人机、无人车等在 GPS 拒止环境中易受赛博限制和电子战攻击，因此为了保证信息的完整性、实用性和安全性，需要不断加强集成赛博防御、电子战防护技术[2]。“快速轻量自主（FLA，Fast Lightweight Autonomy）”计划就是美国国防预先研究计划局（DARPA）于 2015 年启动的相关项目，该项目旨在开发一种新的导航、感知、控制和规划算法，以保证无人系统能够在复杂环境中自主稳定飞行。

一般情况下，无人机大多使用包括全球卫星导航系统如 GPS、北斗系统、伽利略系统等，以及惯性导航系统和高度计等构成的组合导航系统进行导航定位。但在卫星信号弱或信号遭到封锁等情况下，无人机无法从卫星系统获取位置信息，此时只能依靠惯性导航系统，尽管 INS 相比其他导航系统有着自主实时、高隐蔽且不受外界干扰的众多独特优点，但其本身不与外界进行信息交换，因此误差会随时间不断积累，仅能在短时间内可靠工作。

视觉导航是一种在 GNSS 拒止条件下的强有力的辅助工具，其可以增强无人机的自主性和可靠性。根据工作环境和场景，现有的无人机视觉定位方案可以分为如下两类：a）基于同步定位与建图（SLAM，Simultaneous Localization And Mapping）的室内/室外小型场景视觉导航方案，经典 SLAM 框架包括视觉里程计、后端优化、回环检测与建图，其主要应用在没有环境先验信息的情况下，载体于运动过程中建立环境模型的同时估计自己的运动。b）基于图像匹配的室外大场景视觉导航，通过搭载的离线地图数据来与无人机获取的实时地面影像进行相似性匹配，估计无人机位置及姿态进而实现自身定位和导航。

视觉信息辅助定位方案相比其他附加传感器的辅助定位方案具有以下优点：与基于雷达和激光雷达的方案相比，其拥有独立的被动信息采集，在敌对地区自我暴露的风险较低；与基于无线电的方案相比，其在复杂环境下的鲁棒性高，很难被电磁干扰或屏蔽；与线性扫描方案相比，其表面观测能力强，允许使用更复杂和更强大的信息处理技术；与使用高程信息的方案相比，丰富的高精度遥感图像资源更易于获取[3]。

本文主要研究的是无人机的视觉辅助导航方案，在获取具有明显特征的数字影像

地图数据的基础上，使用无人机搭载的摄像机等传感器获取地面影像进行匹配，再结合 INS、高度传感器给出的无人机位姿、高度等数据融合解算出无人机的实时位置和姿态从而辅助定位。整个过程中，信息来源稳定、定位精度高、潜在风险小，可以保证无人机的自主性与全局性。

1.2 国内外研究现状

本文的视觉辅助导航方案核心是图像匹配技术，该技术对于整个导航方案的精度和稳定性有着重要影响。因此，本小节首先介绍图像匹配技术的相关研究，再进一步介绍基于图像匹配的无人机视觉导航技术。

1.2.1 图像匹配技术相关研究

图像匹配技术经过几十年的发展，大体上可以分为基于区域的匹配方法、基于变换域的匹配方法和基于特征的匹配方法。其中基于区域的匹配方法是指根据局部像素灰度信息来最小化信息差异进行匹配，如互相关法、相关系数度量法等。基于变换域的方法通过将图像映射到其它域进行匹配，如沃尔什变换、相位相关法等，其在刚性变换效果较好，相关研究也已经趋于成熟。然而，基于区域和变换域的方法对于成像条件、图像形变和图像信噪比有着极为严苛的要求，且往往运算复杂度较高，这大大限制了其在实际生产生活中的应用。为了解决上述问题，基于特征的图像匹配技术得到了广泛的研究。

基于特征的匹配算法从图像中提取显著的结构特征，例如特征点、特征线或边缘以及具有显著性的形态区域，由于特征是对整张图像的精简表达，相对减少了很多不必要的计算，同时也减少了噪声、畸变等因素的影响，是目前图像匹配的主要形式，也是本研究选择的主要方案。

（1）基于局部邻域的传统特征

在传统的特征检测方法中，基于邻域计算的特征表现出了较好的性能，例如二阶偏导数方法，其基于尺度空间理论的拉普拉斯-高斯（LoG，Laplacian of Gaussian）函数[4]，通过在多尺度空间搜索极值来作为特征点，具有一定的尺度不变特性。Lowe 等人[5]在此基础上提出了著名的 SIFT 特征检测算法，利用高斯差分（DoG，Difference of Gaussians）函数来逼近对数滤波器，大大加快了计算速度。随后 Mikolajczyk 等人[6]将 Harris 和 Hessian 检测器与 Laplacian 和 Hessian 矩阵相结合用于尺度和仿射特征检测，称为 DoH（Difference of Hessian）。Bay 等人[7]提出了 SURF 算法，其通过 Haar 小波计算近似与 Hessian 矩阵的检测器以及积分图像来加速 SIFT，从而简化二阶差分模板的构造。后来相继有人提出了一些基于 SIFT 和 SURF 的改进，包括使用双边滤

波来近似 Laplacian 计算的 ASIFT[8]，使用分段三角形滤波器来逼近 DoH 的 DART[9]，使用余弦高斯调制滤波器的 SIFT-ER[10]等。2020 年，Divya 等人[14]在针对合成孔径雷达（SAR，Synthetic Aperture Radar）图像的匹配中提出一种基于结构张量的 SIFT 算法，作者认为标准 SIFT 通过高斯滤波来生成尺度空间，但同样模糊了图像中的信息，因此利用张量扩散技术来构建尺度空间并提取特征。

与圆形的高斯响应函数不同，KAZA 检测器[11]采用非线性偏微分方程，通过非线性扩散滤波进行点特征搜索。随后作者又通过在金字塔框架中嵌入快速显式扩散从而开发出了一种加速版本 AKAZA[12]，以加快非线性尺度空间中的特征检测。另外一种方法是 WADE[13]，它通过波传播函数来实现非线性特征检测。

（2）基于有监督学习的手工特征

随着计算机资源的发展，近些年来深度学习方法在对图像特征上表现出了优越的学习和表达能力，在图像匹配领域内取得了巨大成效。通过训练多个分段线性回归模型，Verdie 等人[15]提出了能够在光照或剧烈成像变化下检测可重复关键点的 TILDE，其训练样本使用 DoG 算子从训练集收集，然后从同一视点的多幅图像中检测良好的候选特征点，并训练一般回归因子预测得分图，其非极大值抑制（NMS，Non-Maximum Suppression）后的最大值被视为期望的特征点。

DetNet（Lenc 等人，2016 年）[16]是第一个学习通用局部变换特征公式的网络，它将检测任务转换为一个回归问题，推导出一种协方差约束来学习稳定的特征。Savinovetal 等人[17]提出的 Quad-net 使用单实值响应函数实现了变换不变性分数排序下的关键点检测，使其可以通过优化可重复性的排序从头开始学习检测器。随后 Zhang 等人[18]提出了一种相似的检测器，其将这种排序损失与峰值损失结合起来，训练了一个可重复的检测器。

不同于前文所述的仅为特征检测而设计的深度学习方法，越来越多的研究者选择将特征检测和描述集成在一个网络中。2016 年，Kwang 等人[19]提出了一种深度学习框架 LIFT，算法思想来源于传统算法 SIFT，包含特征检测、方向估计和描述符三个模块，各模块均基于卷积神经网络（CNN，Convolutional Neural Network）实现，训练所使用的特征点来自 SIFT-SfM（SIFT-Structure from Motion）算法三维重建的结果。Pautrat 等人[20]认为当前特征描述的局限性在于不变性和区分性之间的取舍，为此使用 LISRD 网络来为 SIFT 检测到的特征生成描述符，其可以根据当前环境（主要是有无旋转和光照之间组合）自动选择拥有最好不变性特征的描述符。

考虑到大多数区域匹配任务需要最近邻搜索，Tian 等人[21]基于 CNN 模型提出了一种 L2-Net 描述符，希望可以直接在欧几里得空间学习高性能描述子。算法采用全卷积结构，通过优化批次描述符之间的相对距离来训练。同年，Szkocka 研究小组的 Anastasiya Mishchuk 等人[22]受 SIFT 启发，引入一种度量学习所用的损失函数以最大

化训练样本中最近正负样本间的距离，结合该损失函数与 L2-Net 结构提出了描述符 HardNet。

2022 年，Zhang 等人[23]为了解决 SAR 图像和可见光图像匹配的问题，其首先利用着色网络和生成对抗网络将 SAR 图像生成一幅类可见光图像，考虑到乘性相干斑噪声对二阶导数的影响，作者使用了多尺度的 Harris 算法构建尺度空间并检测特征点，在此基础上，使用 GLOH 方法[24]生成描述符、最近邻（kNN，k-Nearest Neighbor）方法进行匹配、随机抽样一致性（RANSAC，Random Sample Consensus）算法筛除错误匹配点对。

（3）基于弱/自/无监督学习的邻域特征

受传统特征提取方案的影响，基于卷积神经网络的特征检测的一般解决方案是构建响应图并在有监督的环境中搜索兴趣点。尽管有监督的方法已经显示出了使用“锚点”（如 SIFT 特征）来指导其训练的好处，但这会使深度学习的性能很大程度上受到锚点构造方法的影响，因为特征本质上很难定义，在图像附近没有锚点时训练集会阻止网络提出新的特征[25]。因此众多学者开始研究摆脱锚点限制的解决方案，由三维重建得到的 GT（Ground Truth）信息更多地作为预训练或是测试数据，而不再作为网络主要的输入信息。

2018 年，DeTone 等人[26]提出了一种基于自监督框架，用于训练兴趣点检测器和描述器的完全卷积模型 SuperPoint，网络能够同时输出像素级的特征点位置及其描述符。One 等人[27]提出的 LF-Net 使用了相机的运动结构估计参数和三维数据（即深度图）训练模型。在此基础上 Shen 等人[28]创建了一个由 LF-Net 结构改进的端到端匹配框架 RF-Net，其提出了构造接受特征图以产生更有效的关键点检测，此外还引入了邻域掩模的方式来促进训练过程中图像对的选择以增强描述符的稳定性。2020 年 Luo 等人[29]提出的 ASLFeat 同样采用联合学习特征检测和描述的方式来提高检测精度。另外一种与检测相关的基于深度学习的方法是方向估计，此外空间变换网络[30]（STN，Spatial Transformer Network）也是一种基于深度学习的旋转不变性检测器的重要参考。

2019 年，Jerome 等人[31]认为可重复提取的兴趣点未必有较高的匹配分数，例如对于重复纹理程度较高的摩天大楼的窗户，因此提出了 R2D2 特征检测和描述网络，其仅在具有高置信度的区域学习检测和描述，通过自监督训练可以同时输出稀疏、可重复且可靠的关键点。针对当前图像匹配算法算法无法处理差异大的图像对的问题，Dusmanu 等人[32]提出了同时进行检测和描述的 D2Net 算法，作者认为当前先进行特征检测再描述的方式有着根本性的缺陷，即特征点由很小区域的低级信息得到，而描述符通常包含较大区域信息，导致检测结果不稳定，因此 D2Net 算法采用密集提取的方式来计算描述符并检测具有不同描述符的像素来作为特征点。

（4）基于端对端网络的语义特征

随着图像匹配数据集的广泛增长以及端到端深度学习的兴起，能够理解语义信息的端到端网络在图像匹配任务上展现出了出众的效果。2016 年 Choy 等人[33]提出了 UCN 模型，该模型使用深度度量学习来直接学习保留几何或语义相似性的特征空间。相似的，NCN（Rocco 等人，2018 年）[34]基于使用半局部约束相处特征匹配的经典思想，通过在全局几何模型中分析邻域共识的模式来搜索空间上一致的匹配集，而无需点对点进行任何人工注释，Han 等人[35]-[39]也提出了类似的方法。

此外，为了实现亚像素级匹配，2020 年 Zhou 等人[40]提出了 PatchPix 网络，通过 patch-level 语义粗匹配+pixel-level 细化匹配的两阶段匹配方式来获取更精准的匹配结果。2021 年 Jiang 等人[41]也采用了一种两阶段匹配方法并命名为 LoFTR，相比于在视差空间搜索对应关系的稠密匹配方法，LoFTR 利用自注意力机制和交叉注意力机制来对两幅图像的特征进行编码，其提供的全局感受野使算法能够在弱纹理区域产生密集匹配。

1.2.2 无人机视觉导航技术相关研究

视觉辅助下的无人机导航研究大多基于图像匹配进行，如基于区域相似性度量的匹配、基于特征点的匹配、基于语义形状的匹配等。无人机影像与参考地图影像经过对齐后可以计算出图像变换矩阵，再考虑相机参数和其他传感器数据，可以解算出无人机的姿态以及绝对坐标。但无人机影像具有实时性、复杂性，同时在参考地图中搜寻到一幅小图的计算量是庞大的，此外根据匹配关系来解算无人机的姿态也颇具挑战，很多研究学者提出了自己的导航方案。

2017 年，Mo 等人[42]-[43]提出了一种 GPS 拒止下的视觉辅助导航框架，在估计无人机的起始位置后通过光流（OF，Optical Flow）预测无人机在后续帧中的位置，利用运动场和单应分解的方式来计算帧间平移，随后在无人机影像和地图间使用 HOG（Histogrms of Oriented Gradients）特征进行匹配，最后采用粒子滤波算法（PF，Particle Filter）对无人机进行逐步搜索定位。

为了实现不同时期的地图与无人机影像间的变化匹配，2019 年 Bhavit 等人[44]提出了一种基于渲染后的地图影像定位方法，其采用归一化信息距离（NID，Normalization Information Distance）来增加对光照或季节等变化下的鲁棒性。与使用 NID 不同，JUREVIČIUS 等人[45]采用粒子滤波的方式来确定无人机位置，通过对一组粒子（粒子被假设为无人机位置）进行概率分布采样来实现，并通过测量连续帧的运动来减少粒子滤波搜索空间。

2020 年，Chen 等人[46]提出了一种基于均值漂移聚类的显著性分析算法来粗提取地图中的路径点。首先对输入图像进行颜色空间变换和高斯低通滤波预处理，利用均

值漂移聚类将图像划分为子区域，作为提取候选路径点的基本单元，利用像素颜色值与整幅图像平均颜色值之间的欧氏距离来表示显著性，并将图像中的显著性区域作为候选路径点。不同于使用传统方案，悉尼大学的 Anastasiia 等人[47]提出了一种基于深度分割网络的自主视觉导航系统，通过对图像分割得到的语义特征进行描述，并与先验建立的分割数据库进行匹配来确定无人机在 GPS 拒止环境中的位置。此外江苏自动化所的 Zhao 等人[48]提出了一种精确定位时间敏感目标（快速移动）的策略。其包含了一种两阶段的图像配准算法，第一阶段对航空图像执行图像缝合，以便生成的图像可以包括场景的更多结构信息，在第二阶段将得到的图像与来自图像数据库的地理参考图像对齐。

2021 年，Tian 等人[49]认为现有的方法忽略了卫星影像和无人机影像的直接几何空间对应关系，因此提出了一种集成了交叉视图综合模块和地理定位模块的端到端跨视图匹配方法，其首先将交叉视图下的无人机影像转换为垂直视图，随后使用条件生成对抗网络将其转换为卫星影像的风格，最后再执行匹配算法。相似的，Jiang 等人[50]针对无人机获取的低光图像提出了一种基于 Retinex 理论的新型全卷积网络，将其用于对低光图像进行噪声抑制和光照补偿。

作为辅助导航的一部分，无人机需要实时得到视觉定位相关结果进而实现系统导航，但无人机搭载的计算平台资源有限，因此如何实现快速定位也是图像匹配的重要处理步骤。实现先验地图的预处理，将在线处理步骤离线处理完成，进而构建合适的数据库，可以大大加快在线处理的速度。Donald[51]在其毕业论文中提出了一种地标数据库构建策略，首先通过先验地图构建数据库，然后在运行过程中通过利用无人机获取地面图像并依据一种权重策略不断更新库中数据。Meng 等人[52]为了实现行星精密定点着陆，提出了一种图像导航数据库构建方法，该数据库特征选择的核心也是计算各图像特征在整个导航系统中的贡献。

1.3 研究内容及章节安排

本研究为综合型研究，提出了一个 GNSS 拒止条件下的视觉辅助的无人机导航方案的框架并针对其原理作出可行性验证。其中涉及了异源遥感影像的稳健匹配技术，无人机视觉特征数据库构建技术，单目视觉位姿估计技术和多传感器数据融合技术。

1.3.1 研究内容

（1）研究了异源遥感影像的鲁棒匹配技术。根据定位任务的复杂性，将异源图像分为两类，分别提出了相应的图像匹配方案，能够完成包括可见光图像、红外图像、SAR 图像、地理影像等异源影像之间的稳健匹配，并针对无人机具体任务场景如视

角、光照等变化条件下具有良好的精度和稳定性。

（2）研究了无人机视觉特征数据库构建技术与实时视觉位姿更新技术。针对无人机视觉特征提取耗时的问题，对遥感地图影像进行先验处理，并提出了一种特征增强技术提取更多“优质”特征，随后对其进行稀疏化，在保证定位精度和鲁棒性的同时提高视觉定位速度。

（3）研究了基于视觉辅助的多源传感器融合导航技术。针对视觉位姿结果频率低且不稳定的问题，基于视觉、IMU 和高度计传感器数据，使用卡尔曼数据融合算法建立了一个较为完整的视觉辅助的无人机导航定位框架，以完成在 GNSS 拒止条件下视觉辅助无人机进行长时间导航定位的精度和鲁棒性验证。

（4）研究了基于高低空无人机协同的视觉定位技术，针对复杂环境下导航定位的需求，提出了一种高低空无人机协同定位的方案，以保证在复杂场景下视觉辅助的无人机导航系统能具备较高的鲁棒性。

1.3.2 章节安排

本文的章节安排如图 1.1 所示，下面进行简要介绍。

第一章 绪论，对本文的研究背景和研究意义做出说明，随后重点介绍了关于图像匹配和视觉辅助无人机导航的国内外研究现状，最后对研究内容与章节安排进行阐述。

第二章 基于点特征的异源影像匹配技术，是本研究的核心内容之一，首先介绍了异源图像的定义和分类，随后对两类匹配难度不同的异源图像，针对性的提出了两种不同的图像匹配方案，重点对图像匹配结果的匹配速度和稳定性做出改进。

第三章 基于异源影像匹配的无人机位姿估计技术，在异源影像匹配的基础上，介绍了如何根据图像匹配结果估计无人机的视觉位姿。此外，在该框架之上，结合高低空无人机的优缺点，介绍了高低空视觉协同的方案来应对无人机在部分场景可能出现的视觉长时间匹配失败的问题。

第四章 基于卡尔曼的多源信息融合导航处理技术，由于图像匹配结果并不总是可靠的，为此简单介绍了惯性导航系统，然后结合惯性导航系统的数据、高度计数据和视觉位姿解算数据建立了卡尔曼数据融合系统，最后针对两种不同图像匹配方案以及高低空协同导航方案进行了实验验证。

第五章 软件设计与实现，介绍了本研究对整个研究内容进行可视化显示所设计的 Qt 界面，包括软件开发环境、界面设计和功能。

第六章 总结与展望。

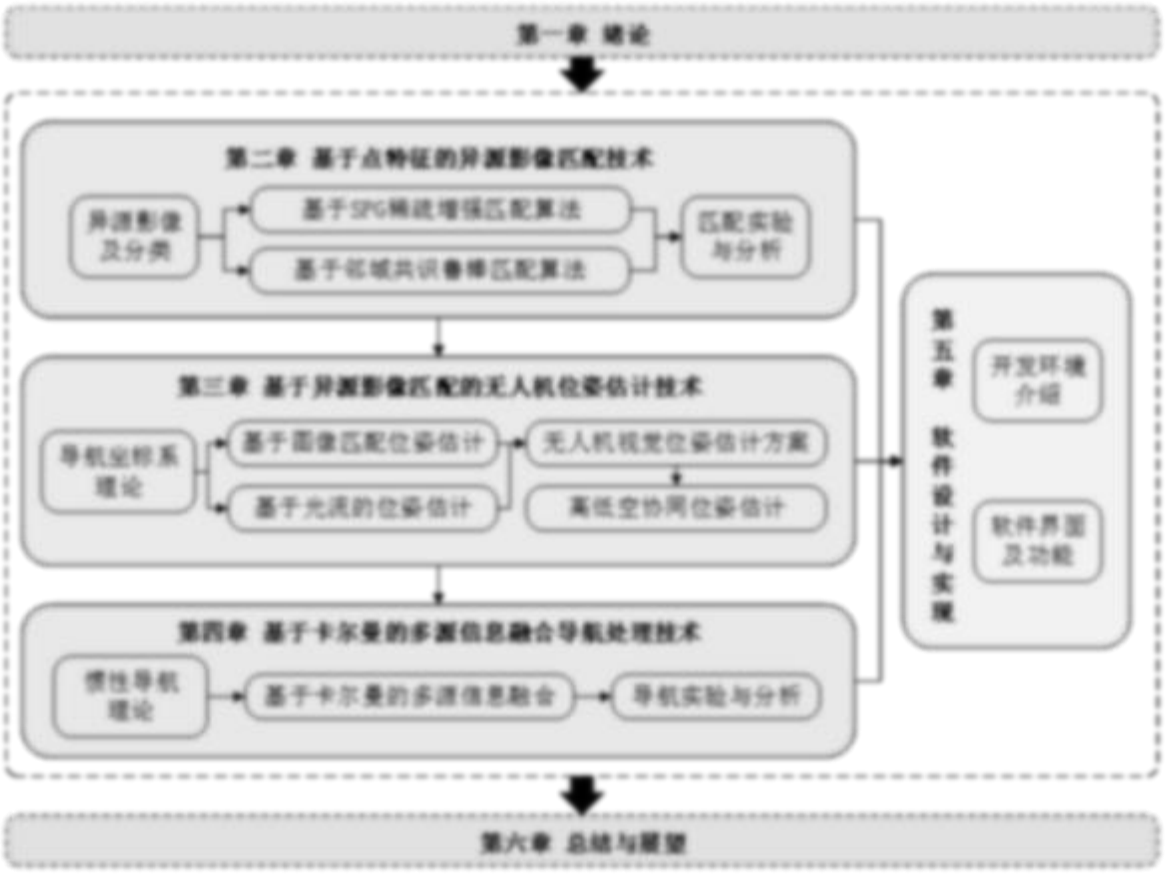


图1.1 本文章节安排结构图

第二章　基于点特征的异源影像匹配技术

图像匹配技术是 GNSS 拒止条件下视觉辅助无人机实现自主导航定位的核心技术，无人机在飞行过程中通过传感器获取地面影像，随后通过图像匹配技术与预先搭载的卫星地图进行匹配，从而实现无人机自身的定位。在这个过程中，一方面卫星地图相对于实时采集的影像总是“过时”的，另一方面，在不同任务需求下，为了避免自然环境影响，无人机使用的图像传感器可以是可见光、红外或微波等，这些因素使得两幅待匹配图像间存在巨大差异，在这里将这种具备较大差异的两幅或多幅图像统称为异源影像。为了解决异源图像间的匹配问题，本文提出了两种基于点特征的异源影像匹配算法。针对不同任务需求，下面首先介绍两类图像匹配网络。

基于点特征的图像匹配网络根据有无中间过程可分为两类：端到端的图像匹配模型和非端到端的图像匹配模型。在图像匹配的任务中，输入是两幅图像，输出是两幅图像间的匹配关系，中间过程包含特征检测、特征描述、特征匹配和外点筛除等多个部分。端到端模型将整个任务全部集成在一个网络里，两幅图像同时进行处理，网络不输出中间过程（甚至无需再为每个特征生成对应的描述符），如 LoFTR[41]等网络。而非端到端的模型是针对图像匹配过程中某一个或多个具体任务来进行的，如仅用于特征检测的 FAST 算法[53]，同时针对特征检测和描述两个任务的 SIFT[5]、ORB[54]和 SuperPoint[26]算法，仅针对特征匹配的最近邻 kNN 匹配算法，同时针对特征匹配和外点滤除的 GMS[55]、SuperGlue[56]算法，仅针对外点滤除的 RANSAC[57]算法等，如图 2.1 所示。

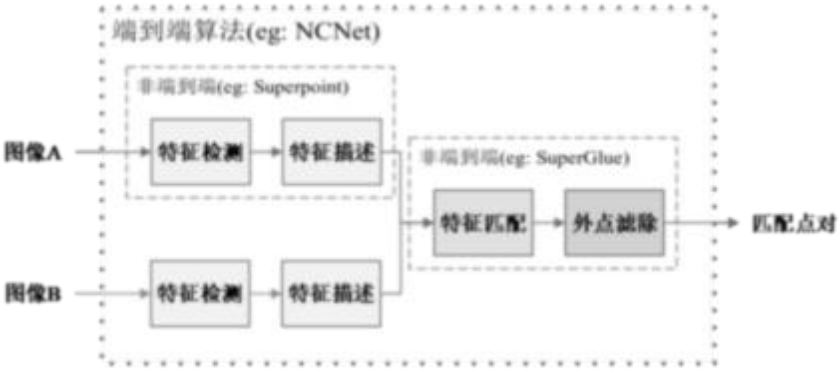


图2.1　端到端匹配算法与非端到端匹配算法

由于端到端匹配算法是针对整个图像匹配任务来进行的，网络可以充分关注两幅输入图像的数据，对有效像素及其邻域做出评估；而非端到端算法在各阶段均有其独立的目标，如 SuperPoint 算法，其是对单幅图像进行处理的，目的是找到更加稳定且独特的特征点，在应对具有大量重复纹理的图像时，则只能交给后续的特征匹配算法。

因此端到端算法相对于非端到端算法在针对具有大量重复纹理、点特征稀疏的区域时具有更加稳定的匹配效果，而后者由于更关注图像细节，相比于前者往往具有更高的匹配精度。

考虑到端到端算法和非端到端算法的优劣，针对异源影像匹配的挑战，本文提出了两种分别适用于不同场景下的鲁棒性匹配算法，分别是针对 I 类异源影像的基于 SPG 的稀疏增强匹配技术（非端到端算法），以及针对 II 类异源影像的基于邻域共识的鲁棒性匹配技术（端到端算法），本章将首先介绍异源影像的两类分类，随后对两种算法进行详细描述。

2.1 异源影像类别介绍

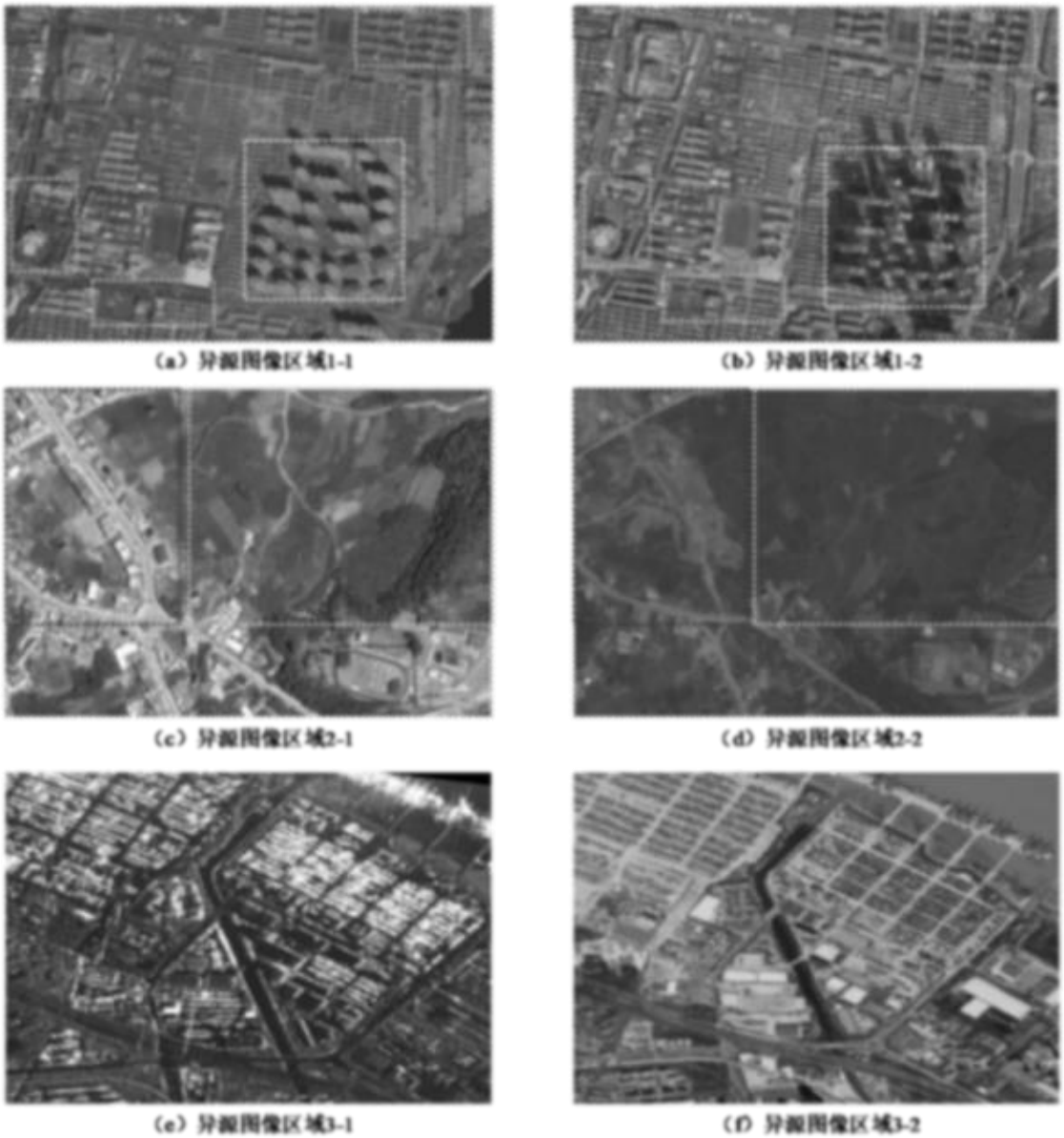


图2.2 异源图像差异性对比

由于以下各方面因素给基于图像匹配的无人机定位带来了严重的挑战，在这里将因为这些因素导致的成像变化统称为异源影像，通常是两幅同一地理区域的影像：

a) 环境因素如时间气候的变化（包括一天中的光照差异、一年中的季节变化）；

b) 人类活动等带来的成像内容的变化（如建筑物的拆建、地形地貌等改变）；

c) 气流、无人机飞控不稳或有意控制下带来的成像视角的变化;

d) 成像设备不同带来的成像质量或成像数据的变化（如红外成像、雷达成像、可见光成像等）。

如图 2.2 所示显示了两个区域在不同时间不同成像设备下的成像情况，可以看出其存在明显的视觉差异。其中（a）、（b）两图是城市区域，存在大量的人工建筑物和少量的绿化景观，（c）、（d）两图是乡村区域，存在部分建筑物和大片的森林田地等区域。黄色框区域标注了无人机成像位置或视角的变化下拍摄到的建筑外观变化情况，红色框区域主要是建筑、道路或地形等地的拆建变化情况，而蓝色框区域主要是绿植、森林、田地等在不同时间季节的外观变化情况，通常在不考虑成像技术的情况下绝大部分的成像差异都是由这三种变化引起的。而（e）、（f）两图分别是 SAR 成像和可见光成像，SAR 图像仅有地物目标反射形成的明暗纹理，而可见光具有丰富的色彩等数据信息。

可见，无论是哪种变化，异源图像间均显示出了较大差异，这也为无人机视觉导航的带来了极大的挑战。为了能使用更加适合的算法来应对各种异源图像带来的匹配挑战，基于两幅图像间是否具备显著相似点特征本文将以上异源图像分为两类：

a) I 类异源图像

该类图像主要是指两幅图像中地物目标点特征显著且未曾大量改变的区域，此类场景通常出现在城市建筑群系，主要包含大量的建筑、道路、桥梁等人工造物，换句话说，I 类异源影像代表了异源图像中特征最为显著的场景。如图 2.2 中（a）和（b）异源图像对，尽管存在季节、相机视角、部分道路拆建等变化，但绝大部分区域建筑物的拐点、角点等都稳定存在，并不受环境太大影响。此外需要说明的是，由于不同成像方式对同一目标的描述数据具有很大不同，例如对于图 2.2 中的（e）SAR 图像和（f）可见光图像，但按照特征显著性标准其也属于 I 类异源影像。

b) II 类异源图像

该类图像主要是指非 I 类异源图像，即点特征稀少或两幅图像间点特征相似性差异较大的图像，包括不同成像方式下的两幅同地理区域图像。如图 2.2 中（c）和（d）异源图像对，尽管单幅图像存在大量的点特征，且林地区域（蓝色框内）纹理细节丰富，但该场景纹理随时间季节变化大，特征不具备辨识度，会导致大量错误匹配产生，此外在建筑区域（红色框内）表面建筑物拆建严重，因此被归类为 II 类异源图像。

可见，I 类异源图像特征显著，场景更具有针对性，相对来说匹配更为容易，因

此更追求匹配速度和匹配精度；而 II 类异源图像更加复杂多变，对传统的特征处理方式提出了巨大的挑战，因此本研究尽量保证其匹配的鲁棒性。基于此，针对两类异源图像，本文提出了一种基于 SPG 的异源 I 类影像稀疏匹配算法和基于邻域共识的 II 类异源影像匹配算法，主要工作如下：

1）针对 I 类异源影像场景引入 SuperPoint 特征提取+SuperGlue 特征匹配，减少异源图像中光照、视角等因素变化的影响。并提出了一种先验特征数据库的构建方式，包含先验数据特征的预处理和特征稀疏化部分，在保证精度的情况下进一步提升算法速度；

2）针对 II 类异源影像场景引入了邻域共识网络来处理密集特征，以避免复杂场景下图像内容变化大导致的匹配失败问题，从而提升匹配的鲁棒性；

3）对算法在不同场景下的匹配效果设计了不同的实验，并与其它具有代表性的算法进行了对比。

2.2 基于 SPG-SE 的 I 类异源影像匹配算法

针对 I 类异源图像的匹配需求——速度快、精度高，本研究提出了一种基于 SPG（SuperPoint+SuperGlue）算法的稀疏增强匹配算法，其基本框架如图 2.3 所示，算法采用了如下方法：

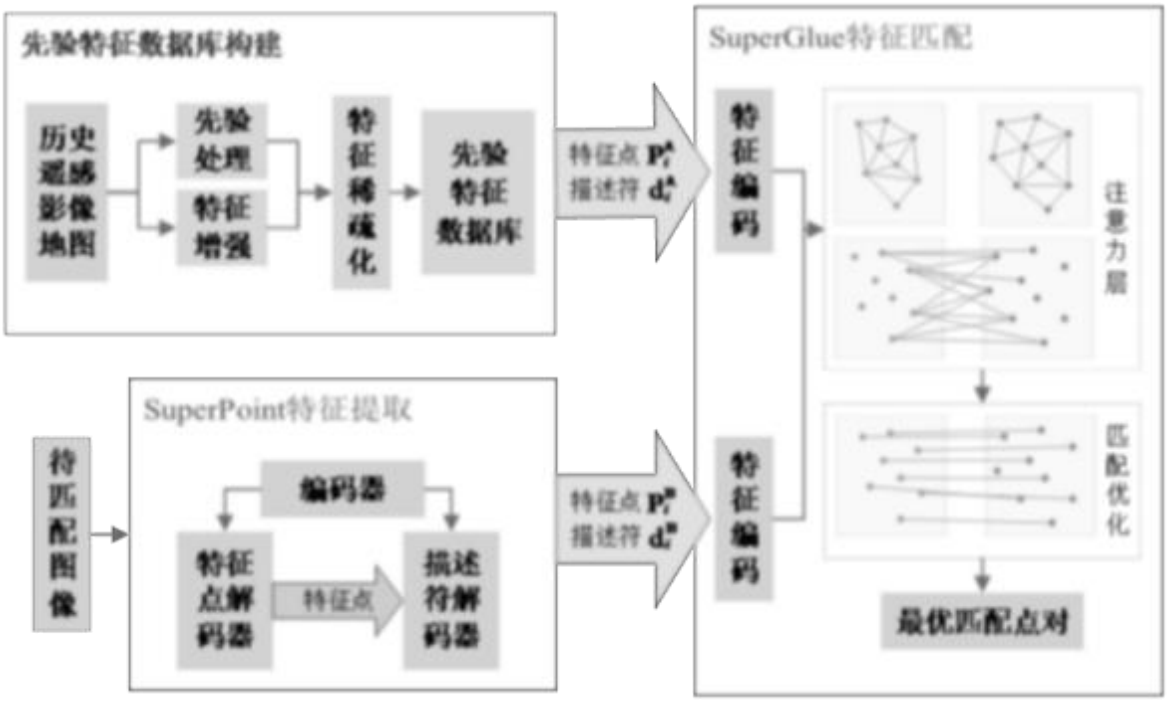


图2.3　基于 SPG 的 I 类异源影像稀疏增强匹配基本框架

（1）引入 SuperPoint 特征，通过自监督的学习机制避免传统特征构造方式限制深度网络的学习能力，从而提升提取到的特征点质量；引入 SuperGlue 匹配网络，通过自注意力和交叉注意力的机制来降低错误匹配的概率；

（2）考虑到无人机定位的应用场景，通过数据增强的方式建立历史遥感影像特征数据库，一方面增加不同时期提取到的特征点以提高匹配成功率，另一方面无人机飞行过程中可以直接加载遥感影像特征点而无需实时提取从而提高算法的实时性；

（3）设计了基于 SuperPoint 特征的特征稀疏化方法，通过建立指标从提取到的大量特征点中筛选出更加优质的特征点用于后续匹配，以降低匹配的计算负载，在保证精度的情况下减少时间消耗。

2.2.2 SuperPoint 特征提取与描述

SuperPoint 是由 DeTone 等人[26]提出的一种能够在具有几何视角变化和较大环境变化的情况下稳定提取特征点和描述符的自监督网络模型，相较于传统对于关键点的模糊定义，自监督学习框架能够避免人工标注关键点对于深度网络学习能力的约束。

如图 2.4 所示显示了 SuperPoint 的模型框架，主要由三个部分组成，网络前半部分是一个共享的 VGG 编码器，用于消除输入图像的冗余信息提取深层特征，网络的后半部分是分别用于提取特征点和生成描述符的解码器。在特征点解码器上，将编码后的特征图输入小型 CNN 网络生成一个特征点响应图，每个像素表示该位置是否有可能存在关键点，特征点解码器还包含了非极大值抑制和特征点筛选部分。描述符解码器部分主要是将提取到的半稠密描述符进行插值细化和归一化最终得到完整的 256 维描述符。

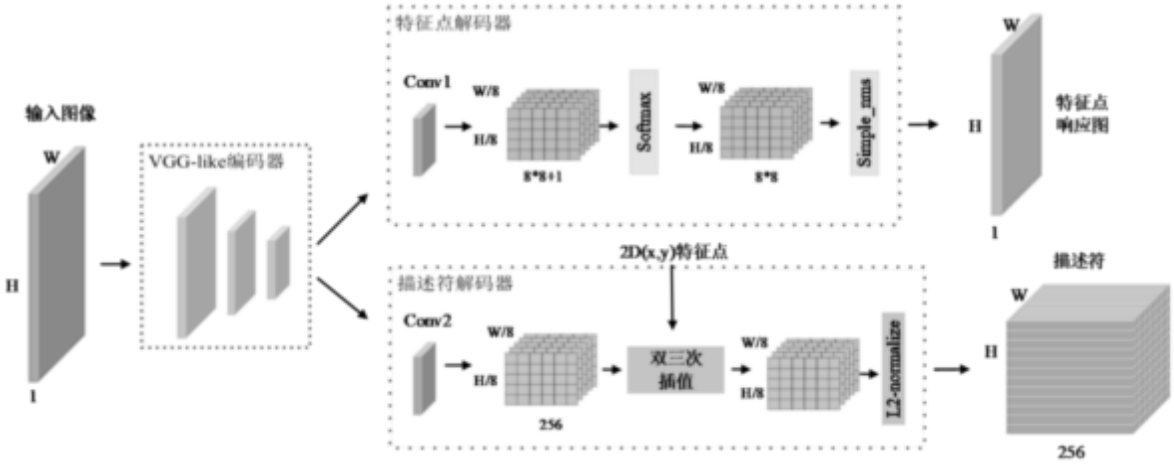


图2.4 SuperPoint 模型框架

图像匹配的特征点并没有一个很清晰的定义，什么是特征点的真值很难界定。在 SuperPoint 算法提出之前，特征提取网络的训练数据主要依赖于 SIFT 算法提取或是

三维重建等方式构建锚点，但这种方式可能会抑制网络模型学习更加深层特征的能力，换句话说，以这种方式训练的网络性能很大程度上会受到锚点构造方法的限制，当附近没有锚点时数据会阻止网络提出新的特征点。SuperPoint 算法创造性地提出了一种合成数据集，实现了特征检测器的自监督训练。其训练主要有三个步骤：

（1）特征点预训练

在图像匹配中特征点并没有一个准确的定义，因此对于一幅任意图像的特征点往往无法确定，但是针对只包含线段、多边形、立方体等简单几何形状的图像来说，线段端点和连接处被视为特征点是毫无争议的。SuperPoint 即采用了这种伪真值来生成数据集并用于训练网络模型中基础的特征点检测部分，如图 2.5 所示展示了人工生成的合成形状数据集部分图案。

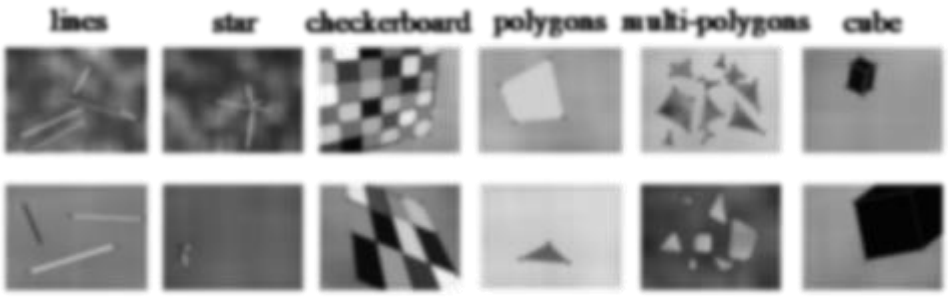


图2.5 合成形状数据集

（2）特征点自标注

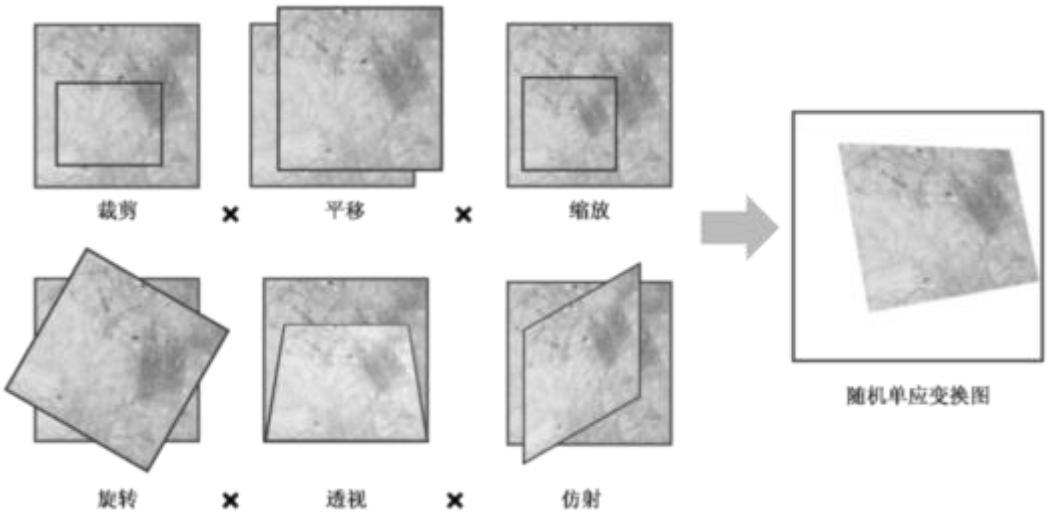


图2.6 随机单应变换

基础特征点检测器的训练集仅仅包含了基本几何形状元素的图像，对于一般的图像（真实图像并不总是有大量的类似角点结构）提取效果并不太好，因此还需要在真

实的大量数据集上进行进一步训练得到更为一般的模型。如图 2.6 所示，为了生成稳定的自标注数据集，还需通过多次随机单应性变换来增强原始图像，从而提高检测器的性能。

在自监督的模式下，首先对无标签图像数据集（如 MS-COCO 数据集）的单幅图像做随机单应变换得到变换后的图像，在这些图像上使用之前训练好的基础特征点检测器分别提取特征点，可以得到多张同一图像的特征点热图，按照单应性变换关系将其变换到原始图像上并叠加可得到最终热图，取阈值截取该热图上的概率值即可作为原始图像特征点的自标注的真值。

（3）联合训练

经过前面两步可以获取一幅图像对应的特征点真值，但使用特征点匹配还需要生成描述符，描述符是对特征点及其邻域信息的描述，通过匹配描述符才能将两幅图像中的特征点对应起来。通常为了使得匹配的特征点对尽量正确，应当使得能够互相匹配的两个特征点的描述符之间的距离尽量“近”，而不匹配的特征点对描述符的距离尽量“远”，这里的“近”和“远”是指在一定度量方法下描述符的相似程度。

同时为了配合特征点解码网络，描述符解码网络也采用自监督的训练方式，首先对于 MS-COCO 真实图像数据集的原始图像做单应性变换，这样两幅图像之间的特征点的对应关系就是变换的单应性矩阵，随后在训练过程中对两张图像的任意点对都求损失函数，去优化损失使得匹配点对间的距离小，非匹配点对距离大，这样就可以得到较优的描述符结构。

2.2.3 SuperGlue 特征匹配与筛选

尽管设计特异性更好的稀疏特征能够帮助图像匹配算法提高匹配成功率和精度，但如果存在大量重复相似特征，这样生成的描述符将也是相似的，例如一栋大楼的外墙窗户，通过为每一个特征搜索其最近邻的匹配可能会造成大量误匹配。针对这个问题，本文引入了 SuperGlue 匹配算法，通过自注意力和交叉注意力的机制来降低错误匹配的概率。

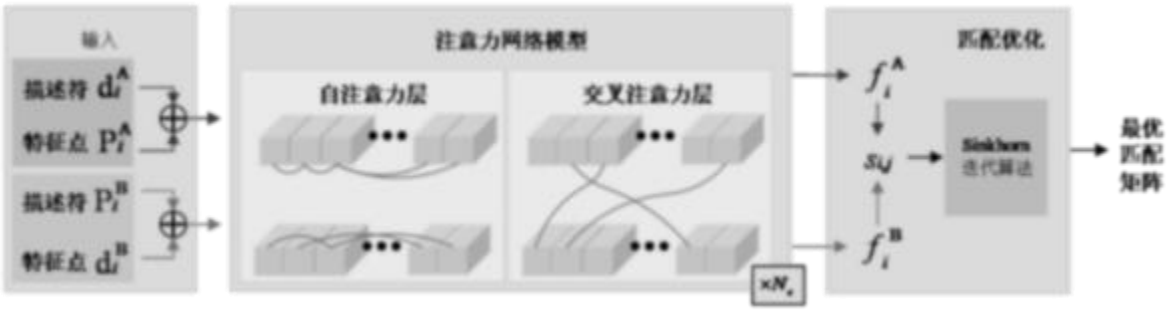


图2.7 SuperGlue 模型框架

SuperGlue 模型框架如图 2.7 所示，输入两幅图像各自的特征点及描述符，输出特征点之间的匹配关系。整个框架由两个主要模块组成：注意力神经网络（AGNN，Attentional Graph Neural Network）以及匹配优化部分。AGNN 将特征点和描述符编码成为一个向量，并利用自注意力和交叉注意力来回增强（重复多次）该向量的匹配性能。在匹配优化阶段，首先计算 AGNN 输出的所有特征匹配向量间的内积，其表示了对应匹配向量的匹配分数，然后其可以等价为一个最优传输模型，进而利用 Sinkhorn 算法迭代计算出最优分配矩阵。

SuperGlue 认为特征点位置+描述可以获得更强的特征匹配特异性，因此首先通过一个特征点编码器将特征点的位置以及描述符合并成为一个向量 $\mathbf{x}_i^{(0)}$：

$$\mathbf{x}_i^{(0)} = \mathbf{d}_i + \mathrm{MLP}_{\mathrm{encode}}(\mathbf{p}_i) \tag{2-1}$$

其中 MLP（Multilayer Perceptrons）表示多层感知器，用于将特征点的位置与其描述进行耦合编码，以使后续的注意力机制能够充分考虑到特征的外观以及位置的相似度。SuperGlue 创造性的将注意力机制用于特征匹配，其核心在于特征聚合，其思想是模拟人类进行特征匹配的过程，即通过来回浏览对比两幅图像，筛选出匹配的特征点，如果不是匹配的特征，还需要观察周围有没有更好的匹配特征，通过主动搜索来增强特征点的特异性。

考虑两幅待匹配图像 I^A 和 I^B，使用其全部特征点来构建一个无向图，这个图包括两种边，分别是连接图像内部所有特征点的 ε_{self}，和连接本图特征点 i 与另外一张图的所有特征点的 ε_{cross}。用 ${}^{(l)}\mathbf{x}_i^A$ 表示图像 I^A 上的第 i 个特征在 AGNN 第 l 层的中间描述，则图像 I^A 中所有特征 i 传递更新的信息可以表示为：

$${}^{(l+1)}\mathbf{x}_i^A = {}^{(l)}\mathbf{x}_i^A + \mathrm{MLP}([{}^{(l)}\mathbf{x}_i^A \mid \mathbf{m}_{\varepsilon \to i}]) \tag{2-2}$$

$[\cdot|\cdot]$表示串联操作，$\mathbf{m}_{\varepsilon \to i}$ 为聚合了所有特征点$\{j:(i,j)\in\varepsilon\}$之后的结果，其描述了特征点 i 与所有其他特征点的相似程度，表示如下：

$$\mathbf{m}_{\varepsilon \to i} = \sum_{j:(i,j)\in\varepsilon} \mathrm{Softmax}(\mathbf{q}_i^T \mathbf{k}_j)\mathbf{v}_j \tag{2-3}$$

该式表示为了查询某一特征描述 $\mathbf{q}_i^T$ 在另一幅图像上对应的值 $\mathbf{k}_i$ 但检索到了元素 $\mathbf{v}_j$，其中 $\mathbf{q}_i, \mathbf{k}_j, \mathbf{v}_j$ 均表示某特征在网络第 l 层的特征中间描述，$\varepsilon = \{\varepsilon_{self}, \varepsilon_{cross}\}$。对于图像 I^B 同样，在网络前向传播过程中自注意力层与交叉注意力层绑定在一起交替进行更新，更新 N_c 次，得到 AGNN 的输出，如对于图像 I^A 有：

$$f_i^A = \mathbf{W} \cdot {}^{(N_c)}\mathbf{x}_i^A + \mathbf{b}, \quad \forall i \in A \tag{2-4}$$

其中 f_i^A 表示了 AGNN 输出的特征最终描述形式，对于图像 I^B 也有类似的形式，于是可以得到 f_i^A 和 f_j^B 得到特征点 i 和 j 的匹配分数：

$$s_{i,j} =< f_i^A, f_i^A >, \forall (i,j) \in A \times B \tag{2-5}$$

接下来构建特征匹配矩阵 $\bar{\mathbf{P}}$，其中矩阵每一个元素 $p_{i,j}$ 描述了特征点 i 和 j 的最终匹配分数，目标是最大化总体得分 $\sum s_{i,j} p_{i,j}$ ，

$$\begin{aligned} \max_{p} \quad & \sum_{i=1}^{N}\sum_{j=1}^{M} s_{i,j} p_{i,j} \\ s.t. \quad & \sum_i p_{i,j} = 1, \forall i = 1,2,...,N \\ & \sum_j p_{i,j} = 1, \forall j = 1,2,...,M \end{aligned} \tag{2-6}$$

该部分是一个经典的最优传输问题，可使用 Sinkhorn 算法进行迭代求解。

网络采用有监督方式进行训练，即根据两幅图像间的单应性变换矩阵得到的匹配点对 $\mathrm{X} = \{(i,j)\} \subset \mathrm{A} \times \mathrm{B}$ 来训练，其中 X 表示匹配点对真值，A 和 B 分别表示两幅图像全部特征点的集合（包含匹配到的和未被匹配到的）其损失函数如下：

$$Loss = -\sum_{(i,j)\in \mathrm{X}} \log \bar{\mathbf{P}}_{i,j} - \sum_{i \in \mathrm{A}} \log \bar{\mathbf{P}}_{i,N+1} - \sum_{j \in \mathrm{B}} \log \bar{\mathbf{P}}_{M+1,j} \tag{2-7}$$

2.2.4 特征稀疏化与先验特征数据库构建

在无人机视觉定位过程中，对于视觉导航系统输出信息的实时性具有较高的要求，这也给图像匹配算法带来了挑战。根据不同的图像匹配算法和应用场景需要考虑合适的能够实现无人机快速定位的策略，针对 SPG 算法，一个有效策略是建立先验特征数据库。

数据库的建立可以分为两个方向：一是静态方法，即提前对预飞行区域的历史数字影像地图进行处理，从而减少无人机实时飞行的资源消耗；二是动态方法，即在无人机飞行过程中实时构建或更新数据库，以应对数字影像地图与无人机实际拍摄影像存在较大变化的问题，如 SLAM 方法。

（1）地图先验处理

基于非端到端图像匹配算法的无人机定位过程分为几个步骤：1）使用特征点检测算法分别提取无人机影像和数字影像地图的特征点和描述符；2）执行匹配算法对两幅图像的特征和描述进行匹配；3）使用外点滤除策略对得到的匹配点对进一步筛选；4）根据图像的点对匹配关系解算无人机位姿。针对以上各步骤在图像匹配过程中使用不同算法的耗时占比情况，使用了 48 对相同大小和尺度的异源图像作为测试求取平均结果进行统计，如图 2.8 所示，可以看出在图像匹配过程中的大部分时间耗时在特征点的检测和匹配上，其中特征点检测（包含描述）的时间占据了大约 70%。

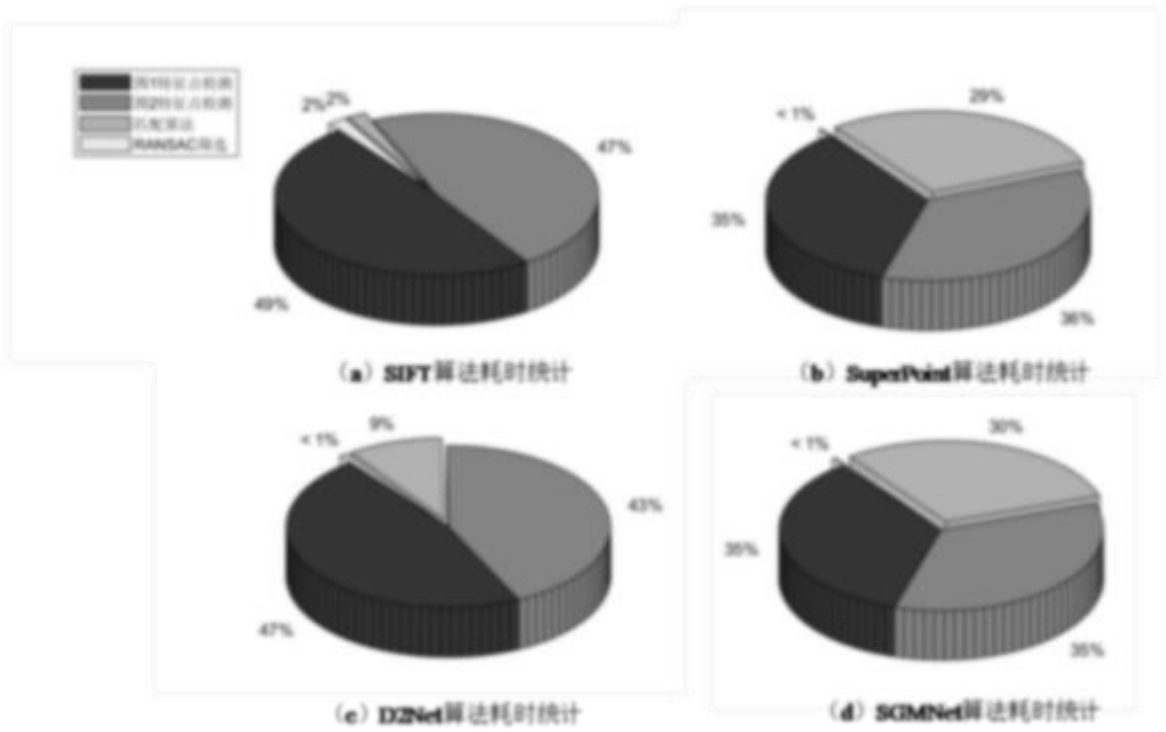


图2.8　不同非端到端算法在匹配阶段耗时占比统计

通常无人机搭载的数字影像地图往往是百万像素级别，若直接使用无人机影像与数字影像地图进行匹配，在特征点检测和匹配上均要面对庞大的数据规模，计算和搜索的时间空间代价都是非常高昂的，因此本文采取预先提取数字影像地图的特征点和描述符，在实际飞行过程中直接在数据库中进行检索的策略，检索时间的代价相比于直接提取特征和描述符是非常小的。

（2）特征增强处理

此外，针对两幅异源图像来说，由于其种种差异原因，算法提取到的特征点并不总是在另一幅图存在相对应的匹配点，或是相匹配的点太多而失去自身的特异性，如图 2.9 所示，对两幅 4000×4000 大小的异源图像进行匹配，匹配后内点的数量可能仅

是提取到的特征点数量的 1/4，即大约提取到的 3/4 的特征点都是外点。这样在执行匹配算法时在特征空间里搜索互相对应的点，大部分的搜索可能都不是必须的，为了提升搜索效率、减少这种无效搜索，对特征空间预先进行特征增强是有必要的。

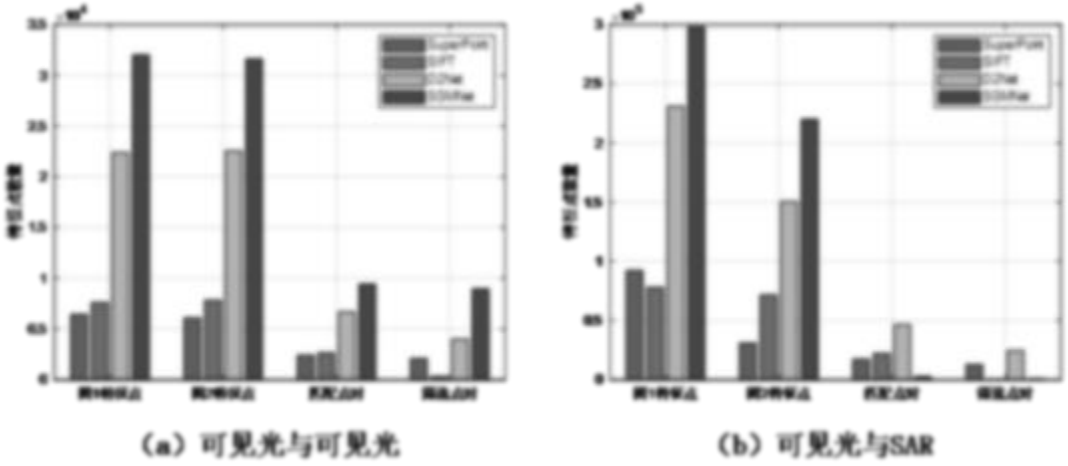


图2.9 异源图像匹配各阶段特征点的数量

特征增强指采取某种方法对特征进行处理以达到使增强后的特征更加“优质”的目的，特征增强技术可以在一定程度上提升无人机在 GNSS 拒止下视觉位姿估计的鲁棒性。此处针对在本任务即匹配导航中定性地给出“优质”的定义，图像匹配定位中使用的优质特征应具有以下特点:

1）突出性。优质特征须具有更突出的特点，如角点、拐点、线段端点等；

2）稳定性。在自然变化（如天气、光照等）、视角变化（如无人机俯仰角变化、飞行高度变化等）、成像设备变化（如红外成像、雷达成像、可见光成像等），特征点应保持稳定提取，例如人工建筑物（道路房屋）或人工作业区（农田湖泊边界）等;

在排除视野变化的遮挡以及地图环境内容变化外，无人机影像中的特征点应当总是能够在数字影像地图上找到与之相匹配的特征点的。为了增强数字影像地图的先验优势，本文作出两条特征点稳定假设：（a）如果一个特征点能够在其影像的增强序列中被重复检测到，则该特征点是稳定的，且重复检测到的概率越高表明其越稳定；（b）如果一个特征点能够在其影像的增强序列中被重复正确匹配，则该特征点是稳定的，且被正确匹配的次数越多表明其越稳定。

其中增强序列定义如下：对于一幅遥感影像 I_0，其存在 M 幅不同时期、不同数据源的历史遥感影像序列 $I_1, I_2, \cdots, I_M$，对每一幅影像 $I_m (m = 0,1,2,\cdots,M)$ 应用 N 次数据增强获得 N 幅变换后的影像序列 $I_{m1}, I_{m2}, \cdots, I_{mN}$，这样生成的 $M \times N$ 幅影像称为遥感图像 I_0 的增强序列。

为了使无人机能够在复杂环境下工作，增强策略应该考虑到异源图像之间的差异

性，因此数据增强的方式包括随机单应变换、添加噪声（如椒盐噪声、白噪声等）、基于 OSG（OpenSceneGraph）三维渲染的雷达/红外/光照等成像模拟等方式，来保证先验特征数据集的合理性和完备性。

这样对于一个特征点 $p \in I_0$，其对应第一条稳定假设的稳定性分数定义为其可以在 I_0 的增强序列中被重复检测到的图像的数量，即：

$$f_a(p) = \frac{1}{M \times N} \sum_{m=1}^{M \times N} F(\min_{q \in I_m} \| H_m(d_p) - d_q \|_2 < h) \tag{2-8}$$

其中，q 是特征点 p 对应于在增强序列中的图像的特征点（q 是位于点 $H_m(d_p)$ 邻域内的所有点），d 表示对应点在图像坐标系下的位置，$F(.)$ 是指标函数，$\|\cdot\|_2$ 表示计算两个点的欧氏距离，$I_m(m=1,2,\cdots,M \times N)$ 是对应的 $M \times N$ 幅增强序列影像，H_m 是图像 I_0 到图像 I_m 的单应性变换关系，h 是匹配成功的判断阈值（邻域大小）。

对于一个特征点 $p \in I_0$，其对应第二条稳定假设的稳定性分数可以定义为其可以在 I_0 的增强序列中被重复匹配到的图像的数量，即：

$$f_b(p) = \frac{1}{M \times N} \sum_{m=1}^{M \times N} G(\| M_m(d_p) - H_m(d_p) \|_2 < h) \tag{2-9}$$

其中，$M_m(d_p)$ 是经 SPG 算法计算出的特征点 p 在增强序列图像 I_m 上对应的匹配点坐标位置，$G(.)$ 是指标函数。需要说明的是，经 SPG 算法计算后，对于特征点 p 来说，在对应的一幅增强序列图像上最多仅有一个对应的匹配点，也可能没有。

如图 2.10 所示，假设数字影像 I_0 中有三个特征点 p_1, p_2, p_3，并生成了三幅增强序列图像，在各个增强序列上检测的特征点表示为 q_1, q_2, q_3, q_4。图（a）与图（b）显示了第二条稳定假设的映射过程，经 SPG 算法得到匹配点对 $(p_3, M(p_3))$，其满足阈值筛选条件，可以作为数据库中的一个候选特征点。图（a）与图（c）显示了第一条稳定假设的映射过程，在特征点检测后，根据增强序列的映射矩阵将图（a）的特征点 p_3 映射到图（c）的坐标 $H(p_3)$，搜索其邻域内所有特征点，显然距离其最近的特征点 q_3 并不在设定阈值内，因此其稳定性分数 $f_a(p_3)$ 就会有一定的下降。此外，值得注意的是特征点 p_4，其在增强序列第 0 幅图像中并未被检测出来，但在其他几幅增强序列图中都被稳定检测到，因此根据第一条稳定假设其也应该被作为先验地图数据库中的一个候选特征点。

通过对增强序列中的所有候选特征综合考虑两条假设的稳定性指标，只有稳定性分数在满足一定阈值条件下才将该特征点存储在先验特征数据库中。特征增强技术是

为了提升无人机位姿估计时图像匹配结果的鲁棒性，但要想达到预期效果还需要结合特征稀疏化技术，下面就稀疏化对视觉位姿估计方面的必要性进行说明：

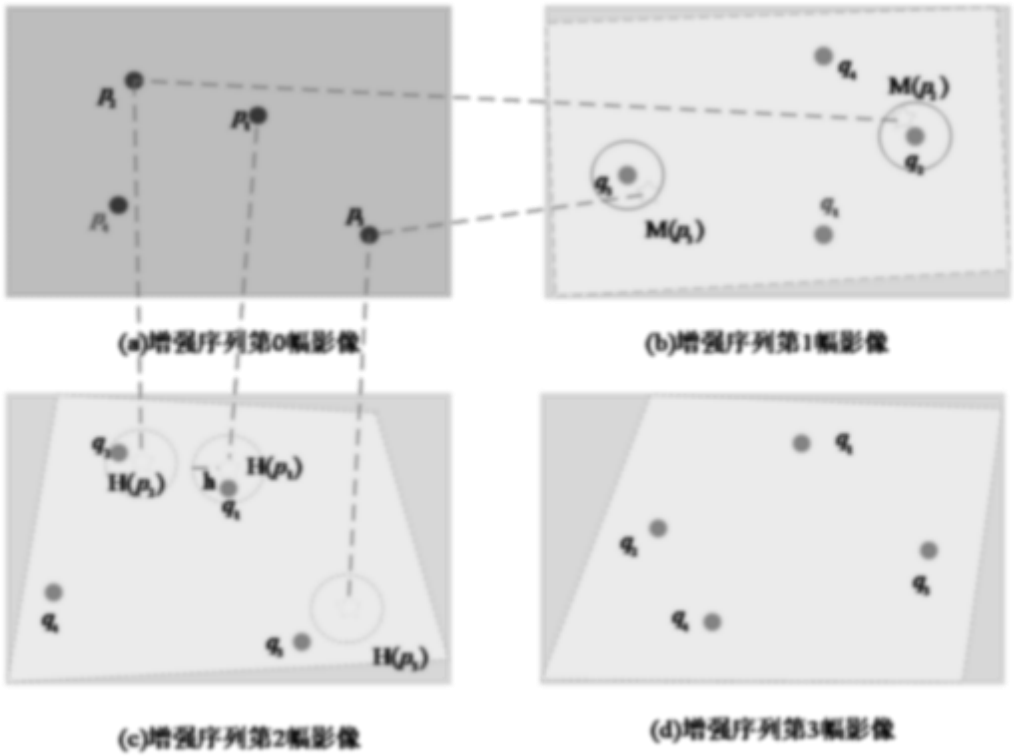


图2.10 增强序列映射示意图

a) 在实时性方面的影响

尽管特征增强技术能够剔除掉特征提取算法提取到的“非优质”特征点，这种点在匹配时可能无法在匹配空间中找到对应的匹配点，但却需要与匹配空间的每个点计算一次或多次距离，增大了搜索空间，若提前筛选掉这些点可以在一定程度上降低匹配时的时间复杂度。但另一方面，特征增强技术也会通过图像增强序列增加特征点，尤其是在特征显著区域，但在正确匹配点对足够的情况下太多匹配点对将会大大增加计算量，因此需要对其进行稀疏化，即在当前邻域内保留少量相对更加优质的特征点。

b) 在鲁棒性方面的影响

由于选取的特征点是全局提取的，不同类型或区域的特征点在匹配效果上存在较大的差异变化，例如图像的某一局部区域特征明显（如人工建筑群系），能够提取到大量特征点，而其他区域（森林草原湖泊等）仅能提取少量，这样特征明显的区域将在匹配结果中产生更大的权重，从而导致视觉位姿估计鲁棒性受到其影响，因此要对其加以区分。

（3）特征稀疏化

为了使图像匹配的特征点对用于解算匹配模型具有良好的精度，一个关键是匹配

点对位置的均匀性，即要求匹配的点对尽量覆盖整个图像，这样解算出的匹配模型才会具有较好的精确性，若匹配点对集中在某一个小范围的区域，这种情况说明匹配模型仅仅是一个局部解而不一定是一个好的全局解，为此需要实现关键点在局部上的稀疏分布和全局上的均匀分布。

一个传统策略是非极大值抑制算法（NMS，Non-Maximum Suppression），即找到局部的“最大值”，并筛选掉（抑制）其邻域内的其他值。NMS 虽然简单高效，但其需要经验阈值来确定邻域窗口大小，针对不同场景、不同分辨率的图像需要不同的窗口，即使是同一幅图像不同的局部也需要不同的阈值来获取最优的分布。

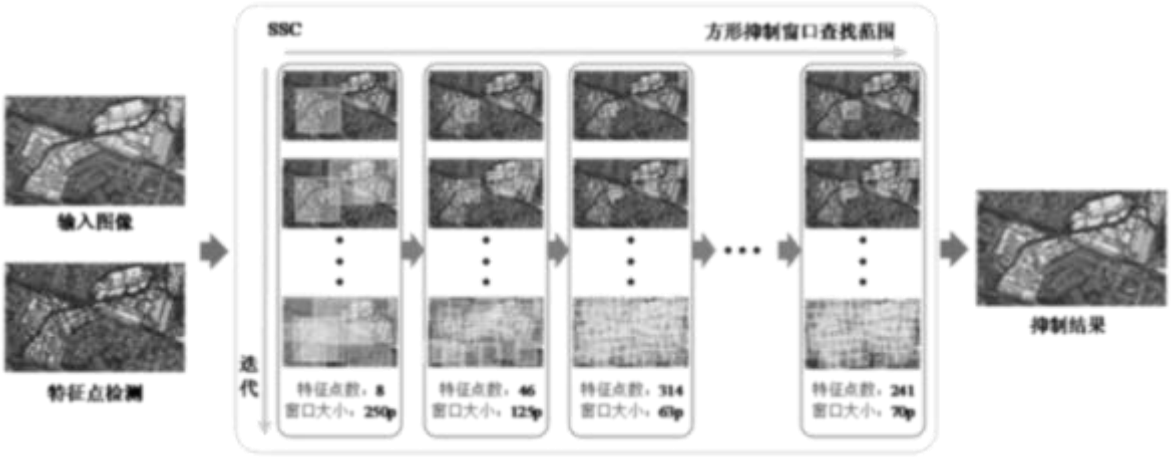


图2.11　SSC 算法流程

本文选择了方形抑制算法（SSC，Suppression via Square Covering）[58]，算法流程如图 2.11 所示，其是一个迭代选取关键点以及试错进而逼近给定数量的过程。SSC 算法的基本思想是先猜测一个响应距离 d，并计算每个点在候选点集合 S 中的响应半径 r，按照响应度从大到小选出响应度最高且彼此相隔至少为 d 的 m 个关键点。对于每个关键点应满足：

$$r_i = \min \| x_i - x_j \|, \text{s.t.response}(x_i) < \text{response}(x_j), x_j \in S \tag{2-10}$$

表示 x_i 的响应能作为区域最大值的半径。算法是一个迭代二分的过程，如果 m<n，则将 d 缩小至 1/2，否则增大 d，然后再执行上面的过程。n 是需要提取的点数量，当 |m-n|<ε 时，停止该流程，ε 是一个预设的阈值。如图 2.12 显示了特征稀疏化后的效果图。

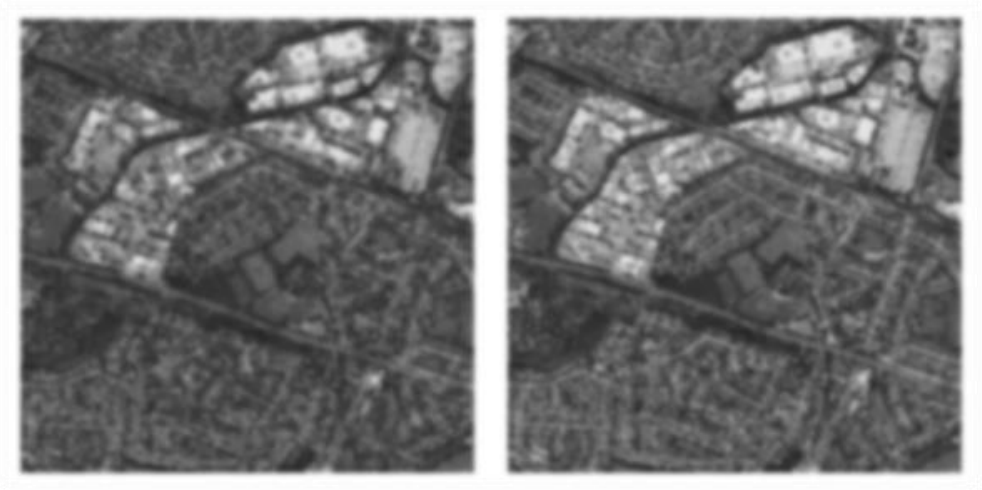


图2.12 原始特征点图（左）和稀疏化效果图（右）

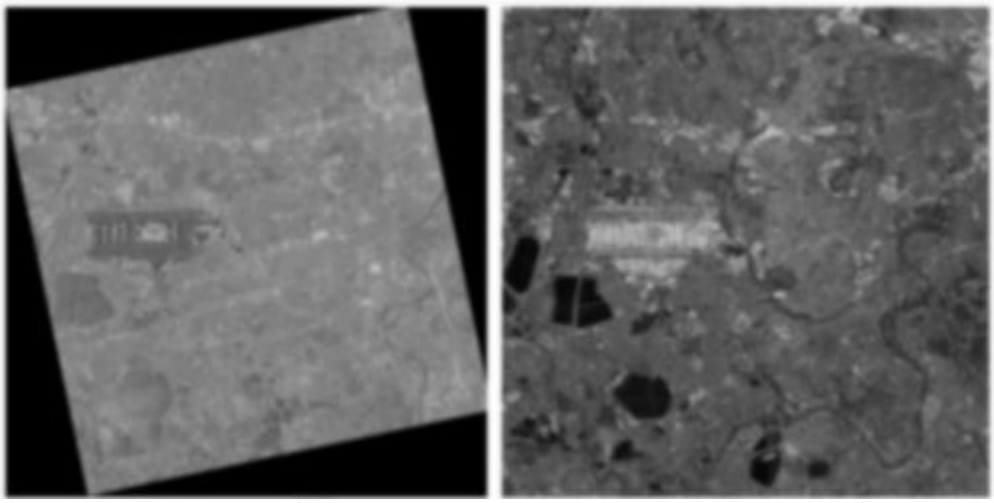
2.2.5 I 类异源影像匹配算法实验对比与分析

为了验证在I类异源场景下，待匹配无人机影像与数字影像地图存在较大差异时，本研究所提出的基于 SPG 的 I 类异源影像稀疏增强（Sparse and Enhance）匹配方法（以下简称“SPG-SE”算法）相较于直接使用 SPG 算法以及其他算法在匹配精度、鲁棒性以及实时性方面的性能差异，本研究选取了 ICEYE 的 SAR 数据集对本研究所提出的算法做了更近一步的分析。

（1）I 类异源影像匹配实验

本实验选取了 ICEYE SAR 数据集中的一幅——英国伦敦的高分辨率 ICEYE 光斑扩展区域（SLEA，Spot Extended Area）成像模式的 SAR 影像，该影像覆盖西思罗国际机场，场景大小为15km×15km，地面分辨率为1.0m×1.0m，入射角 $20^{\circ}-35^{\circ}$。该 SAR 数据集原始 SAR 数据为单视复数（SLC，Single Look Complex）影像，以全分辨率生成，保留幅值和相位信息，但由于存在大量的噪声斑点，难以在局部区域分辨出目标点和噪声点，因此这里直接使用了 ICEYE 提供的全范围检测（GRD，Ground Range Detected）产品影像，即将 SLC 影像处理成振幅影像并投影到地球平面，并以较差的空间分辨率为代价进行多视处理，从而减少噪声斑点，处理后的地面分辨率为 3.0m×3.0m。

该 SAR 数据集场景如图 2.13 所示，场景主体区域为建筑区，包含大量道路、居民房屋以及机场区域等人工建筑，约占到整体比例的 68%，其次是农田、森林、绿化等植被区域，约占到整体比例的 28%，最后是所占比例很少的河流湖泊等水体区域。



（a） SAR 图像　　（b）可见光图像

图2.13　伦敦-西思罗国际机场 SAR 图像及可见光图像

从该 SAR 图像中随机裁剪区域并将其与可见光图像对应区域进行匹配以测试各算法性能，部分测试结果分别如图 2.14-图 2.16 所示。

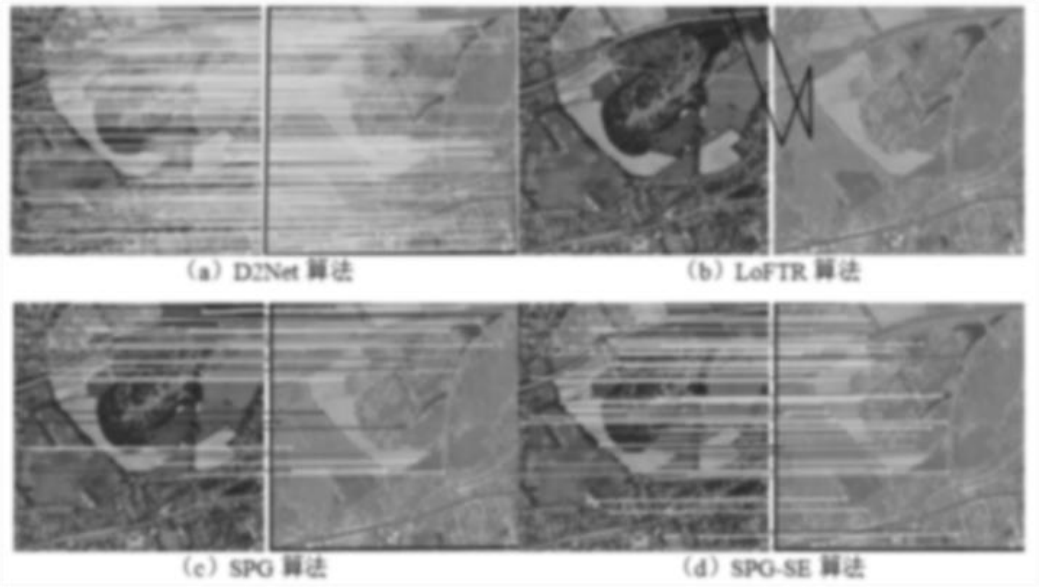


图2.14　机场随机区域一测试结果匹配图

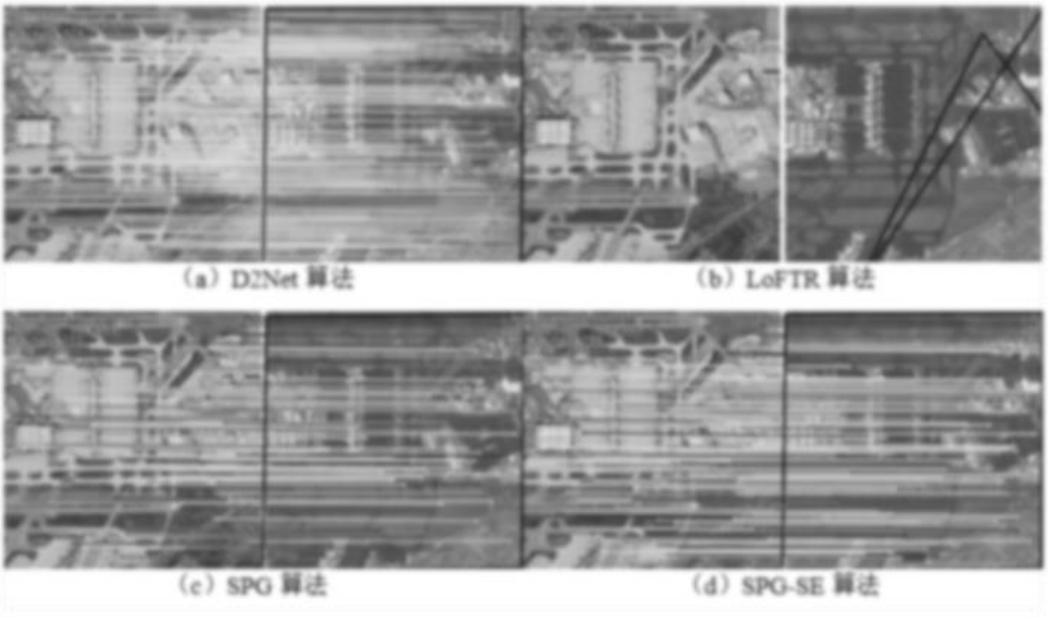


图2.15 机场随机区域二测试结果匹配图

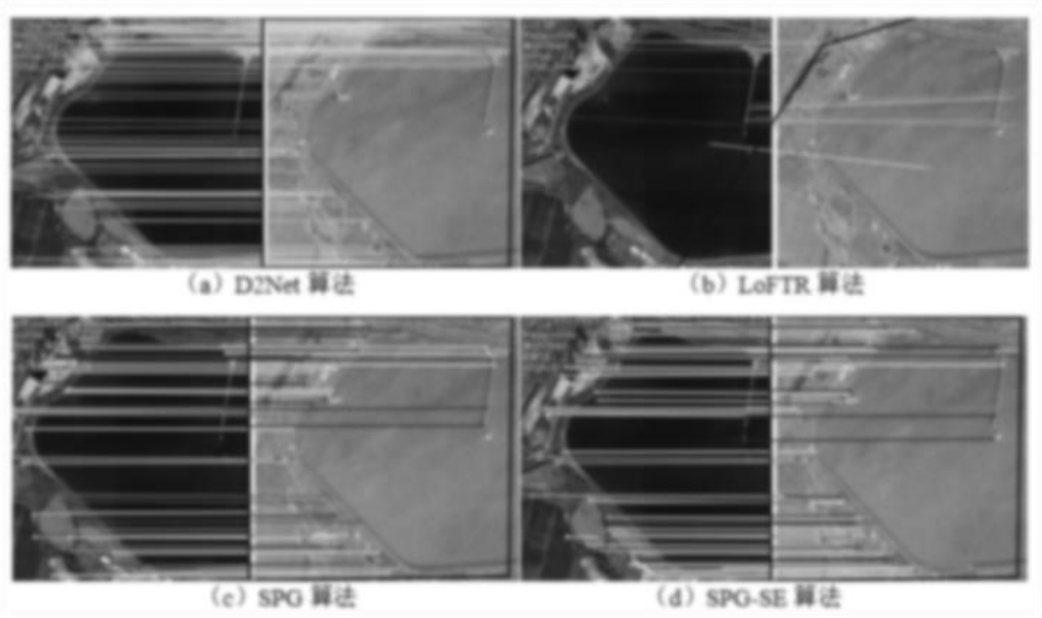


图2.16 机场随机区域三测试结果匹配图

测试结果匹配图中将两幅待匹配图像进行了拼接，其中左侧是可见光图像，右侧是 SAR 图像，连线表示将筛选后的分别位于对应图像的匹配的特征点对连接起来，由于左右两幅图像是同尺度下的对齐图像，因此连线有越多平行线表示匹配结果越准确。而蓝色框线表示将完整左图按照匹配矩阵变换到右图所对应覆盖的区域，理论上其应当是一个覆盖右图的矩形，如图 2.15（a）显示了一个较好的匹配结果。

图 2.14-图 2.16 显示的三幅随机区域是该场景中具有代表性的几个场景，其中图

2.14 随机区域一是城市周边的绿化、空地等区域，但具有较为明显的边界特征；图 2.15 随机区域二是机场部分，完全的人工建筑，具有众多的拐点/边界点、线；而图 2.16 的随机区域三内容的 60%左右部分被一个湖泊占据，但湖泊周围仍具有较明显的特征。按照视觉上的显著程度可以将匹配难度表示为：区域三 > 区域一 > 区域二，根据匹配结果（连线点对的数量和蓝色框的覆盖程度）也可以看出其匹配情况符合难度。

在该 SAR 图像与可见光图像匹配情况下，传统算法 SIFT 和在 SLAM 场景下表现较好的 R2D2 算法在绝大部分场景无法成功匹配，同样使用自注意力和交叉注意力的 LoFTR 算法在多个特征较为显著的区域匹配失败（在大部分区域能进行较好匹配）。而 SPG 算法、D2Net 算法和 NCNet 算法显示出了强大的鲁棒性和精度，均能够得到相当数量的匹配点对数和筛选后的点对数量。

根据匹配结果图也可以直观看出，相比于直接使用 SPG 算法，使用改进后的“SPG-SE”算法得到了更多的匹配点对，匹配结果也相对更加稳定和均匀，在图 2.14 和图 2.16 的（c）和（d）中，直接使用 SPG 算法仅提取到了部分相对更加显著的特征，由于覆盖不均匀且特征点过少，因此解算出的蓝色框线受局部结果影响较大产生了较大误差，在使用特征稀疏增强后提取到的特征点更多且更加均匀，显著改善了直接使用 SPG 算法的情况，但图 2.16 由于水域区域覆盖范围较大，整个中部和右下角部分难以提取到有效点特征，因此各算法解算结果均产生了较大误差。

（2）实验总结与分析

统计各算法在该实验中的测试平均结果如表 2.1 所示，其中 INF 表示均方根误差过大（误差>50 像素）或全部图像匹配失败（筛选后的匹配点对数少于 10）。

表2.1 伦敦-西思罗国际机场数据集匹配综合测试结果（平均）

算法		特征点总数目		匹配点对数目	RANSAC 筛选点对数目	匹配时间/s	均方根误差/像素
		可见光	SAR				
非端到端算法	SIFT	4247	2834	1007	7	0.13	INF
	D2Net	3664	2389	737	387	0.45	3.43
	SPG	1285	491	212	156	0.17	4.24
	SPG-SE	1463	491	271	202	0.14	3.36
端到端算法	R2D2		109		15	0.35	INF
	LoFTR		540		294	0.18	7.12
	NCNet		788		340	0.22	4.64

根据表中数据可以看出，非端到端算法提取出的特征点数目，相比于直接从可见光图像提取，SAR 图像的特征点数目显著减少，这也体现了异源图像的差异性。传统

算法中最具代表性的 SIFT 算法无法处理异源图像的匹配，其虽然具有 1007 组匹配点对，但匹配点对的正确率仅有 0.7%左右；D2Net 算法是针对异源图像匹配设计的，同样采用了密集提取特征的方式，其匹配效果极佳，RANSAC 筛选后的匹配点对数是几个算法中最高的，且均方根误差平均仅有 3.34 像素，但代价是较高的时间和内存消耗；端到端算法 R2D2 针对同源图像设计，在异源图像匹配效果上表现不佳；LoFTR 算法具备一定的异源图像匹配能力，但在 SAR 图像上也表现不佳；NCNet 算法将在下一小节进行介绍。

SPG 算法是针对角点特征特化训练的算法，在异源图像上也更多关注角点特征，因此受光照、角度、异源图像构成方式之间的差异影响较小，表现出了较好的数据结果。而本文对其改进的“SPG-SE”算法相比于直接使用 SPG 算法，首先是在对可见光图像的处理上增加了约 10%左右的特征点数量（对 SAR 图像未进行稀疏增强），但却分别增加了 27.8%和 29.5%的匹配点对和筛选后的匹配点对，可见优质特征的选取策略是有效的。同时由于先验处理的原因，在算法耗时上减少了 35.3%，精度增加了 1 个像素点左右。

总体来说，本研究提出的基于 SPG 的 I 类异源影像稀疏增强匹配算法达到预期效，在针对 I 类异源影像场景中，在不大幅增加特征提取数量的情况下，本文提出的特征增强和稀疏化策略保证了提取特征的优质和均匀。本算法相比于直接使用 SPG 算法，在保证精度的情况下，匹配速度和鲁棒性均有一定的提升。

2.3 基于邻域共识的 II 类异源影像匹配算法

前文提到，SuperPoint 特征提取的主要是角点、拐点、边界点等点线属性显著的特征，但这种方法存在一个根本的局限性：假设无人机飞行在一个弱纹理的区域，如森林、草原等，SuperPoint 将很难提取到有效的特征，提取到的少量描述符也很难进行区分或匹配。

为了解决这个问题，一方面考虑到近年来在图像匹配领域端到端算法展现出来的优势，即在算法中根据参数的几何模型（例如仿射变换、透视变换等）生成图像之间的对应关系，而不是仅仅通过计算描述符之间的相似性来确定匹配点对；另一方面，在这种具有挑战性的场景下，通过提取稀疏特征显然很难解决这个问题，因此考虑使用密集特征。在此基础上，提出了基于邻域共识（NCNet，Neighbourhood Consensus Networks）[34]的 II 类异源图像匹配算法，该算法是一种建立在邻域共识（也被称为领域一致性）的模式匹配思想基础上的算法，能够分析一对图像之间的密集匹配，并从图像数据中学习邻域一致对应的模式。

2.3.1 II 类异源图像匹配算法模型

类似于 SuperGlue 算法中模拟人类视觉匹配图像过程中的自注意力和交叉注意力机制，为了消除重复特征模式上的歧义匹配，我们需要分析包含唯一非重复匹配特征的场景的更大的上下文，这样，来自该非重复特征的匹配信息将能够传播到相邻的不确定匹配上，从而确定这些不确定匹配，这就是邻域共识的思想。将邻域共识的鲁棒性和神经网络的可学习性相结合，通过识别邻域中支持匹配的模式来识别可靠的匹配，如图 2.17 所示，网络主要由三个组成部分：（i）密集特征提取；（ii）NCNet 邻域共识网络；（iii）互近邻匹配。

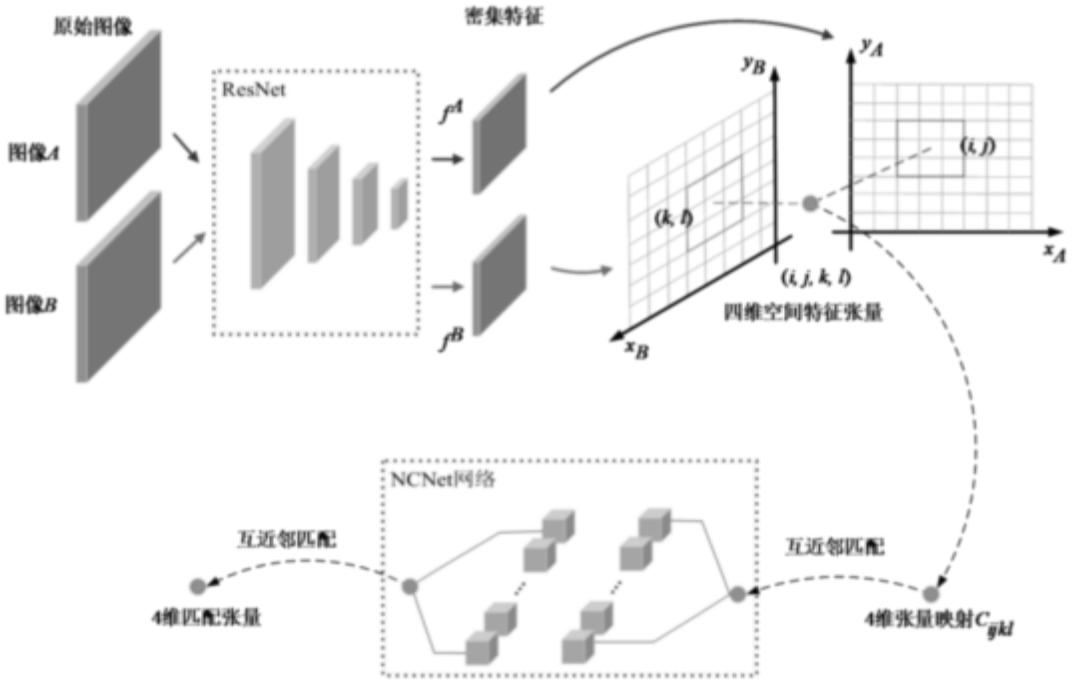


图2.17 基于 NCNet 网络的 II 类异源图像匹配算法模型

（1）密集特征提取

使用卷积深度网络 CNN 来提取密集特征，不同于稀疏特征仅选取某些关键点作为特征点位置提取相应描述，密集特征需要为图像中每一个像素生成对应的特征描述。本文选择了残差卷积神经网络 ResNet-32 作为密集特征提取器，给定一幅大小为 (h, w) 的图像 I，ResNet 将为其生成一组密集的描述符 $\{f_{ij}^I\} \in \mathbb{R}^d, (i=1,...,h, j=1,...,w)$，$d$ 表示特征描述符的维度。

传统算法在提取特征和描述符后选择了最近邻算法 kNN 进行匹配，但在存在重复特征或描述符不够独特的情况下，将一个特征分配给第一个最近邻特征可能会导致一个错误匹配。为了避免丢失掉其他近邻特征的有价值信息，在给定两组待匹配图像的密集描述符 $\{f_{ij}^A\}$ 和 $\{f_{ij}^B\}$ 的情况下，计算两幅图像之间的全部特征之间的余弦相似

性，将其存储在一个四维空间张量 $\mathbf{C} \in \mathbb{R}^{h \times w \times h \times w}$ 中，其中每一个张量元素 c_{ijkl} 分别对应于图像 I^A 中位置 (i, j) 的描述符 $\{f_{ij}^A\}$ 与图像 I^B 中位置 (k,l) 的描述符 $\{f_{kl}^B\}$ 的相关性映射关系，计算如下式所示：

$$c_{ijkl} = \frac{< f_{ij}^A, f_{kl}^B >}{\left\| f_{ij}^A \right\|_2 \left\| f_{kl}^B \right\|_2} \tag{2-11}$$

（2）NCNet 邻域共识网络

存储相关性映射关系的张量 $\mathbf{C}$ 中包含了所有匹配对的分数，大小为 $(h,w)^2$，然而正确匹配的理想数量是 (h,w)，即需要为图 I^A 的每一个像素在图 I^B 中找到一个唯一对应的匹配点，这意味着该张量中绝大部分信息都是错误的匹配。为了进一步对张量空间 $\mathbf{C}$ 中的匹配进行处理和过滤，这里选择四维卷积神经网络来完成邻域共识任务，如图 2.18 所示。

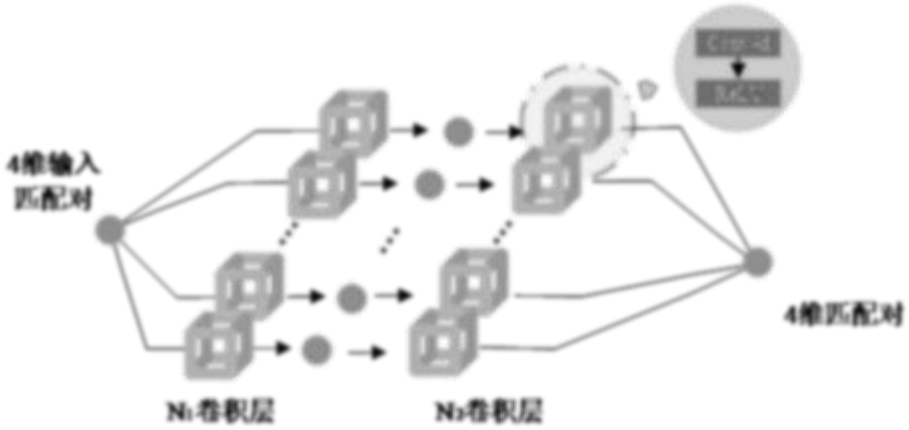


图2.18 邻域共识网络 NCNet

在邻域共识思想的支持下，一个正确的匹配点对在四维空间中应当有一组正确且连贯的匹配围绕着它，这些匹配模式与输入图像的变换模型相同，即如果原图像之间发生了平移，无论四维空间中的位置如何，这个匹配模式也在四维空间中被等量平移。

该网络的第一层卷积滤波器跨越了匹配空间的局部 4 维区域，该区域分别对应于两幅图像中局部邻域的余弦相似性，这样第一层的 4D 滤波器可以处理和检测这两个邻域的所有成对匹配中的模式，每个滤波器均用于学习不同的几何形变，并产生对应于相关张量的每个 4 维点对的局部形变。这些通道由后续的 4 维卷积层进一步处理，其目的是通过前一层捕获更加复杂的特征。该网络生成一个单通道的输出，其尺寸与输入的 4 维匹配对相同。

为了使网络输出结果对于图像的输入顺序保持不变，即无论图像对是作为 (I^A, I^B) 输入还是 (I^B, I^A) 输入都会产生相同的匹配，将输入图像应用到该网络两次，如下：

$$\tilde{c} = N(c) + (N(c^T))^T \tag{2-12}$$

其中，c^T 是指交换对应第一幅和第二幅图像的维度对，即 $(c^T)_{ijkl} = c_{klij}$，若两幅图像具有不一致的匹配模式，将会被降权或删除。

（3）互近邻匹配

尽管该邻域共识网络可以基于匹配点邻域的支持来抑制或者放大匹配，但其不能在全局上施加一个约束，传统方法中典型的互近邻匹配要求匹配点对之间是相互的最近邻关系，这种互近邻条件可以用于消除大部分的候选匹配，但其不可微，因此称之为硬互近邻匹配。

为此这里选择了一种软版本的互近邻匹配，从决策和可微性的角度来看，其可以应用于密集的四维匹配分数。用 r^A_{ijkl} 和 r^B_{ijkl} 分别表示图像 I^A 和 I^B 每个维度上特定匹配 c_{ijkl} 的最佳得分的比率：

$$r^A_{ijkl} = \frac{c_{ijkl}}{\max(c_{abkl})}, r^B_{ijkl} = \frac{c_{ijkl}}{\max(c_{ijab})}, 0 < a < w, 0 < b < h \tag{2-13}$$

软互近邻匹配作为输入的选通机制，以降低非互近邻匹配的分数。对于某一个匹配点在另一幅图像上的所有匹配点对，如果是互近邻匹配点对，该分数将为 1，如果不是，该分数将会降低到 $[0, c_{ijkl})$ 中的某一值。相比于硬互近邻分数会直接置为 0 的方式来说，这种软互近邻策略能够抑制但不会直接“杀死”非互近邻匹配点对，有助于对匹配执行全局约束。

输入的两幅图像 (I^A, I^B) 经过该网络模型之后，将产生一个四维的过滤后的相关性映射张量 $\mathbf{C}$，其中包含了所有成对匹配的分数，为了获取两幅图像之间的对应关系，需要从该张量中获取对应与图像 I^A 中每个特征点的唯一匹配点对。

类似于软互近邻匹配那样，首先使用一个 soft-max 函数将两幅图像的匹配点对分数进行归一化：

$$s^A_{ijkl} = \frac{\exp(c_{ijkl})}{\sum_{a,b}\exp(c_{abkl})}, s^B_{ijkl} = \frac{\exp(c_{ijkl})}{\sum_{a,b}\exp(c_{ijab})}, 0 < a < w, 0 < b < h \tag{2-14}$$

这样对于每一个点都有与之相匹配的所有点的概率分数，可以选择一个最有可能的匹配来完成一个模式的确切分配。

2.3.2 II 类异源影像匹配算法实验对比与分析

为了验证在 II 类异源场景下本研究所提出的基于 NCNet 的 II 类异源影像匹配算法（以下简称“NCNet 算法”）相较于其他算法在匹配精度、鲁棒性以及实时性方面的性能差异，本研究选取了 ICEYE SAR 数据集中的另一幅更具挑战性的场景：美国堪萨斯州的条带（Strip）成像模式的 SAR 影像，覆盖科普兰德的一部分农田区域，场景大小为$30\text{km}\times50\text{km}$，地面分辨率为$3.0\text{m}\times3.0\text{m}$，入射角$15^{\circ}-30^{\circ}$。

（1）II 类异源影像匹配实验

伦敦-西思罗国际机场存在大量建筑物，整幅图像的特征均较为显著，然而无人机执行任务过程中不可避免的需要在山区或是特征较弱的地区飞行，为了测试弱特征区域的算法性能，堪萨斯州-科普兰德农田场景如图 2.19 所示。

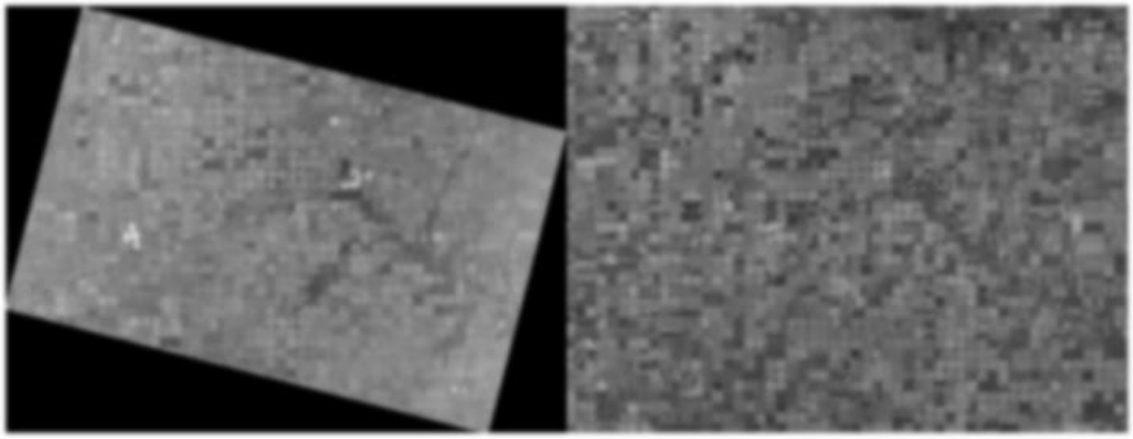


图2.19 科普兰德农田 SAR 图像（左）和可见光图像（右）

科普兰德地区较为干旱缺水，因此将田地修建成了圆形，将灌溉用水管道沿半径铺设，绕圆心旋转，从而方便灌溉。该场景包含一条东北-西南方向贯穿的主干道，主干道上下部分分布着大量的圆形或半圆形的农田，也有部分方形农田，同时存在极少量的建筑区域，约占整体比例的 3%左右。总体上来看该场景存在大量的纹理相似区域，且显著的角点或拐点特征很少，由于地面性质属性相近，SAR 图像的成像方式决定了其本身数据同质性严重，局部区域可能包含大量冗余信息。

从该 SAR 图像中随机裁剪区域并将其与可见光图像对应区域进行匹配，部分测试结果分别如图 2.20-图 2.22 所示。

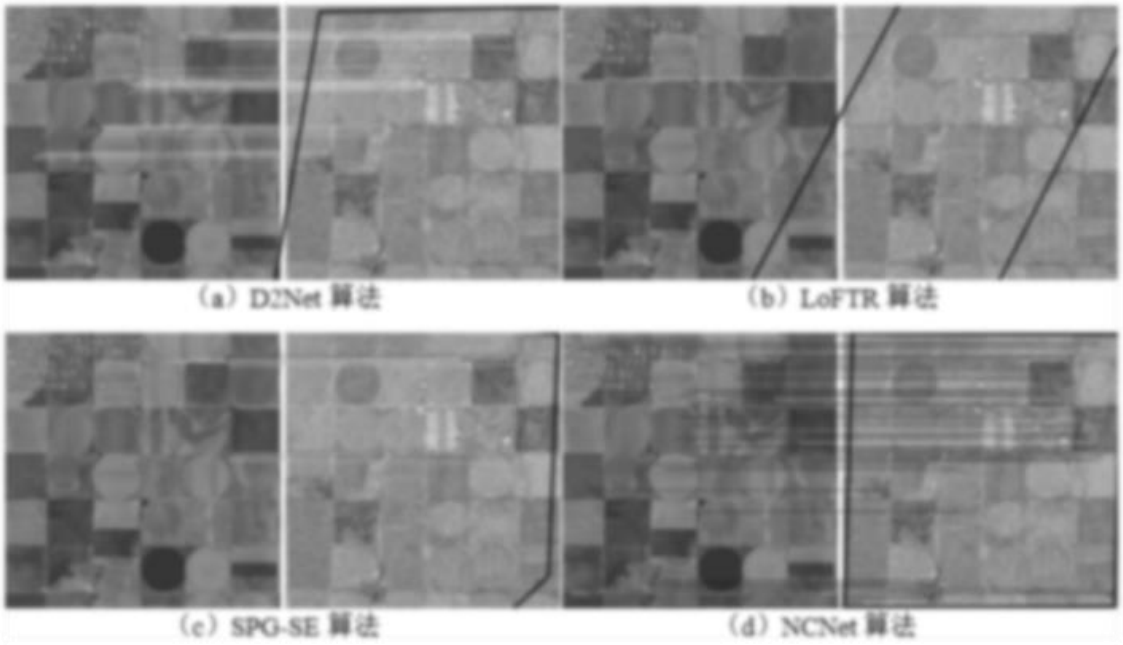


图2.20　农田随机区域一测试结果匹配图

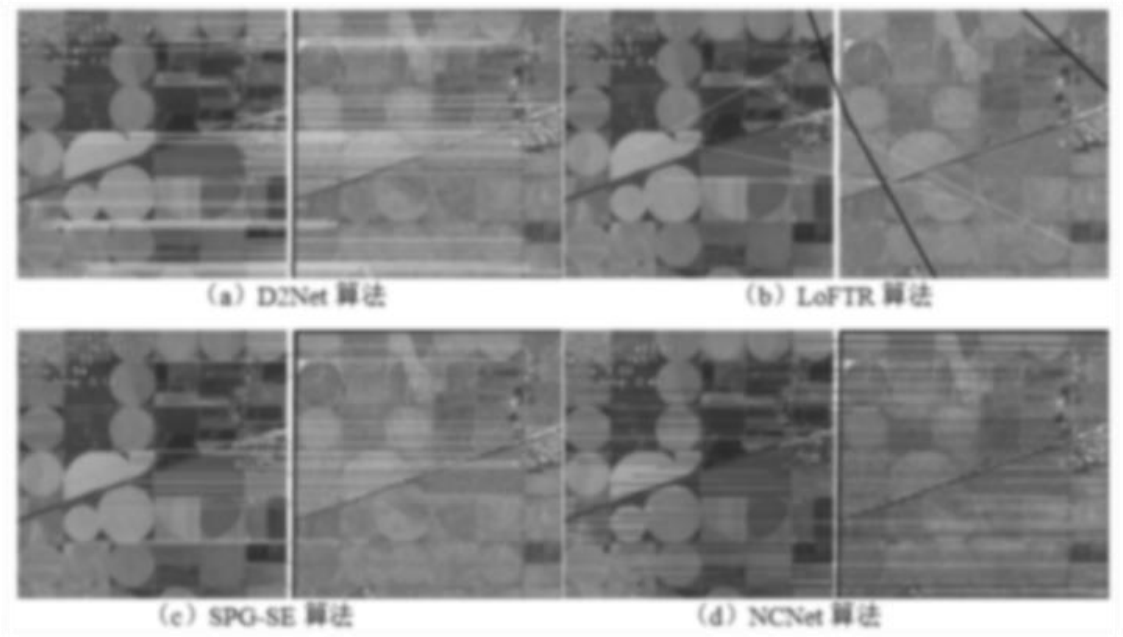


图2.21　农田随机区域二测试结果匹配图

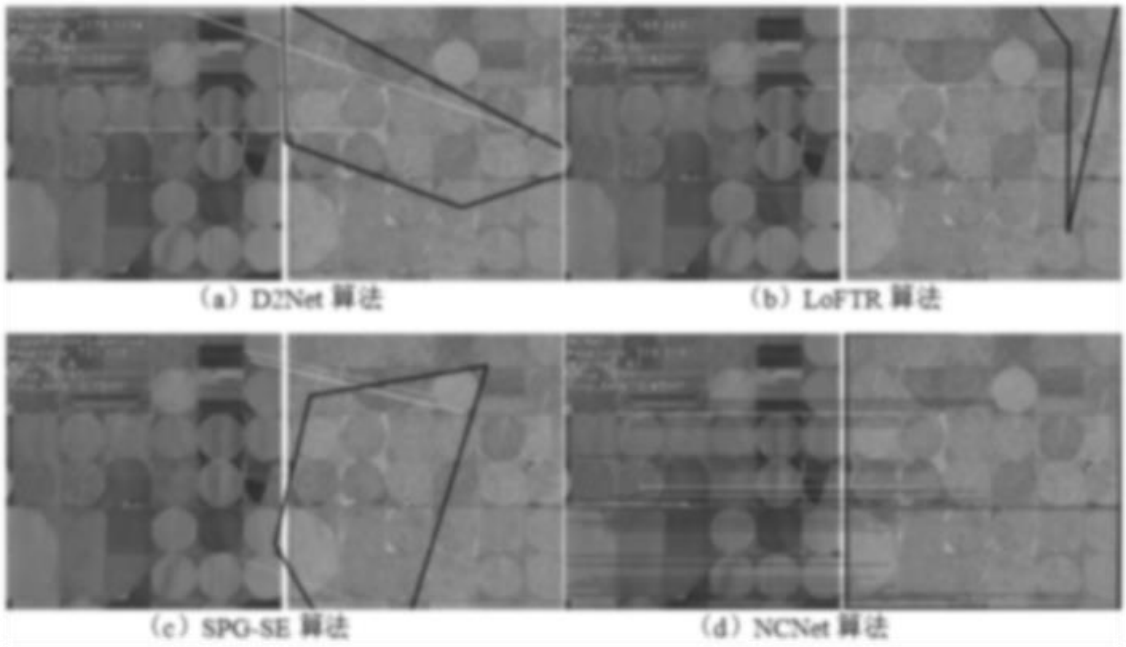


图2.22 农田随机区域三测试结果匹配图

在该 SAR 图像与可见光图像匹配情况下，SIFT、R2D2 和 LoFTR 算法基本无法成功匹配。而 SPG-SE 算法只能在存在较多显著点特征（如人工建筑物）的区域匹配成功，如图 2.21 区域二；D2Net 算法相比于 SPG 算法鲁棒性更优，但在只存在少量显著特征的区域也只能有一定的概率匹配成功，如图 2.20 区域一，尽管其找到了部分正确的匹配点对，但由于匹配点对聚集导致在解算单应性矩阵过程中产生了较大误差。

（2）实验总结与分析

统计各算法在该实验中的测试平均结果如表 2.2 所示：

表2.2 伦敦-西思罗国际机场数据集匹配综合测试结果（平均）

算法		特征点数目		匹配点对数目	RANSAC 筛选点对数目	匹配时间/s	均方根误差/像素
		可见光	SAR				
非端到端算法	SIFT	764	973	198	7	0.13	INF
	SPG-SE	1176	137	27	19	0.08	29.19
	D2Net	3769	1193	313	61	0.59	25.88
端到端算法	R2D2		130		11	1.08	INF
	LoFTR		172		8	0.21	INF
	NCNet		465		123	0.17	3.81

可以看出，相比于可见光图像，本区域的 SAR 图像无论是 SPG-SE 还是 D2Net 都难以提取出足够量的特征点，而通常这类依靠提取稀疏特征匹配的算法都需要足够量的特征点，以通过大量正确匹配特征点对去抑制错误匹配特征点对对匹配结果的影响。然而针对当前数据图难以提取足够的特征点，在特征点对不足的情况下将无法筛除错误的匹配点对，这导致了两种特征点算法在处理弱显著性区域性能显著下降。

而 NCNet 算法在该极具挑战性的区域依然显示出了其强健的稳定性，其经过筛选后的特征点对数平均保持有 123 对，且均方根误差仅有 3.81 像素，相比于其他算法保持了可观的精度。

2.4 基于点特征的异源影像匹配算法性能测试与分析

为了验证基于点特征的异源影像匹配算法针对异源图像的实际泛化性能，本小节选择了不同类型的异源图像对各算法性能进行测试，下面首先介绍相关性能指标。

（1）匹配精度指标

为了更加精确地对各算法的性能进行评估，本文选择如下指标进行量化：

a) 匹配时间（MT，Matching Time）

为了统一端到端算法与非端到端算法的差异，本文的匹配时间包含了特征提取和描述时间、特征匹配时间、外点滤除时间。

b) 均方根误差（RMSE，Root Mean Square Error）

均方根误差衡量了误差结果的平均离散程度，通常其也等价于图像匹配结果的精度。对于单应性变换矩阵 H 已知的两幅图像 A 和 B，假设匹配后共有 N 个匹配点对，即位于 A 图的特征点 p_i^A 对应与 B 图中唯一的特征点 p_i^B，则对于 RMSE 有：

$$RMSE(p_i^A, p_i^B) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\| H(p_i^A) - p_i^B \right\|_2} \tag{2-15}$$

c) 匹配准确率（MMA，Mean Matching Accuracy）

其中匹配准确率衡量了在阈值 t 下图像匹配算法匹配的准确程度，其定义如下：

$$MMA(p_i^A, p_i^B; t) = \frac{\sum_{i=1}^{N} \alpha(t - \| H(p_i^A) - P_i^B \|_2)}{N} \tag{2-16}$$

其中 t 是两点间二维距离的阈值，$\alpha(\cdot)$ 是一个二进制指示器函数，正值为 1 负值为 0。

（2）图像匹配实验

本实验选择了武汉大学 skyearth 团队的异源影像数据集 MRSI[59]以及研究者搜集的部分异源图像作为测试数据集，该数据集成像数据多种多样，包括卫星和飞机实际拍摄、仿真软件仿真、经二次处理的图像等。数据集共 96 个场景的不同异源图像对，图像分辨率大多在 500x500 像素左右，包含“可见光-可见光”、“红外-可见光”、“白天-夜晚”、“SAR-可见光”、“抽象地图-可见光”、“深度图-可见光”共 6 种类型，选取部分如图 2.23-图 2.28 所示。

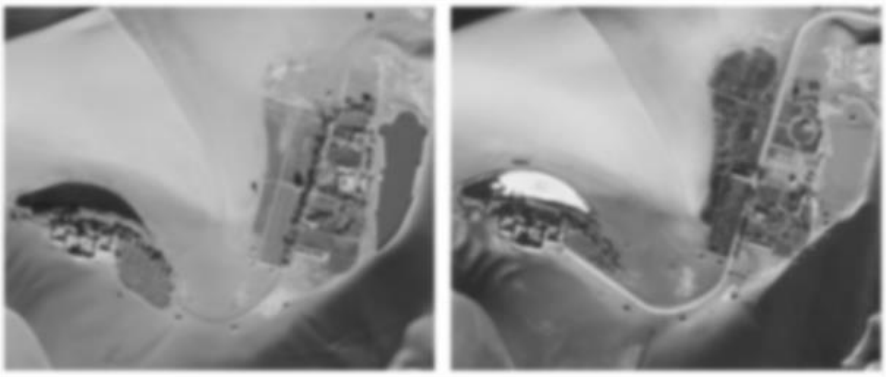


图2.23 MRSI 数据集 I：“可见光-可见光”影像（ID：2）

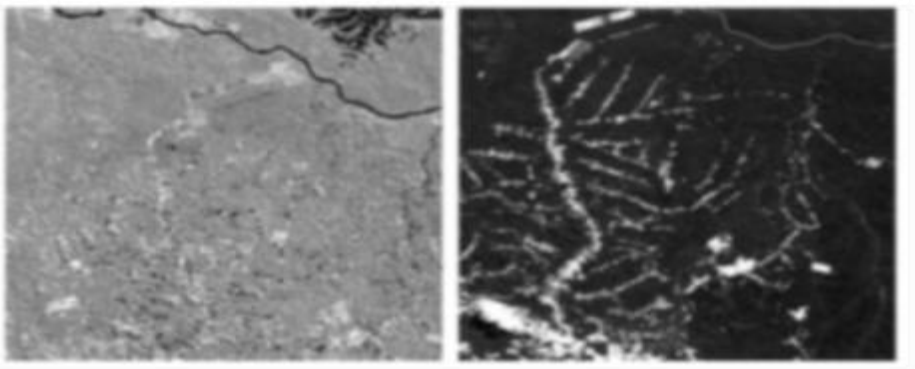


图2.24 MRSI 数据集 II：“SAR-红外”影像（ID：5）

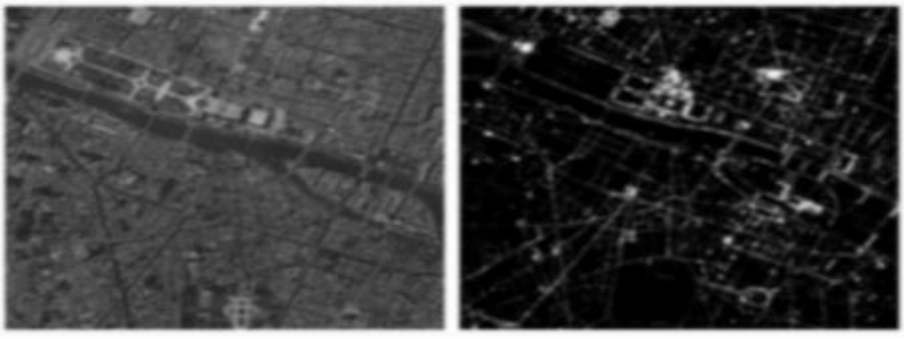


图2.25 MRSI 数据集 III：“白天-夜晚”影像（ID：1）

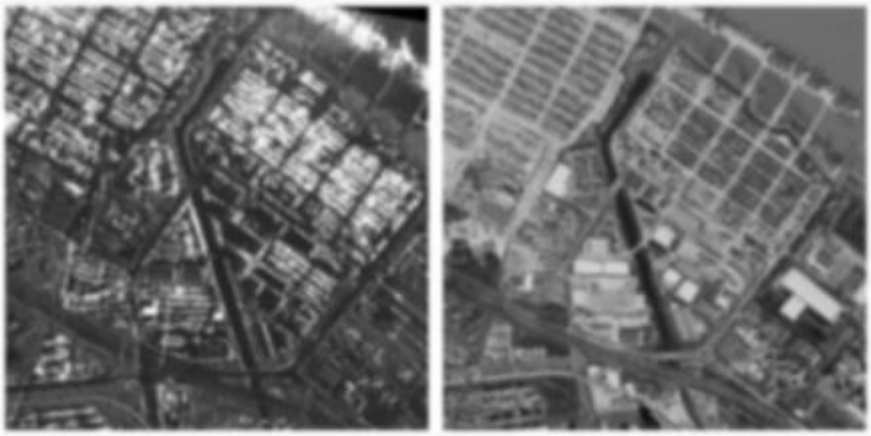


图2.26 MRSI 数据集 IV：“SAR-可见光”图像（ID：6）

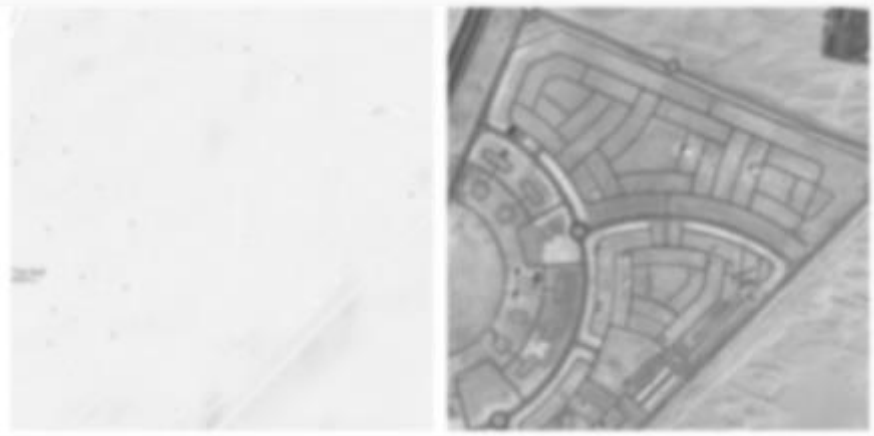


图2.27 MRSI 数据集 V：“抽象地图-可见光”图像（ID：7）

图2.28 MRSI 数据集 VI：“深度图-可见光”图像（ID：3）

可以看出，不同成像设备（如红外、SAR、可见光）、不同数据源（如仿真图像、可见光图像）、不同光照（如白天、夜晚）、不同角度、不同时间等等原因呈现出的影

像存在巨大差异。

整个数据集测试平均结果如表 2.3 所示，其中匹配点对数、筛选后点对数、MT 和 RMSE 仅统计成功匹配的图像对。表中算法的匹配点对数目与 RANSAC 筛选后的点对数目存在较大差异，原因主要有两方面：其一是 RANSAC 筛选为了从所有匹配点对中迭代搜寻误差最小的最多符合两幅图像之间的映射关系的点对，因此不符合映射关系的匹配点对将会被剔除；其二是因为 RANSAC 使用了平面一致性假设，即假设图像中的所有点都在同一平面上，这样就会筛选掉图像中由于视角不同而带来的建筑物高度差异生成的匹配点对，这一步的目的是为了在后续能够实现根据匹配点对解算无人机的位姿。

表2.3 基于点特征的异源影像匹配算法性能测试实验（平均）

算法		匹配点对数	筛选后点对数	MT/s	RMSE/像素	MMA/10
非端到端算法	SIFT	695	22	0.073	INF	0.093
	D2Net	687	311	0.426	2.23	0.937
	SPG	242	201	0.124	2.04	0.887
	SPG-SE	273	235	0.089	1.75	0.898
端到端算法	R2D2	159	86	0.547	INF	0.479
	LoFTR	692	610	0.204	4.12	0.792
	NCNet	538	336	0.171	2.64	0.979

（3）实验总结与分析

可以看出，传统算法 SIFT 虽然速度快但并不具备匹配异源图像的能力，因此仅将其作为时间上的参考标准；而深度方法除 R2D2 外，其他算法均在该数据集上有着较为不错的性能，这是因为 R2D2 算法更加着重于解决提取到的特征点的可重复性和可靠性的问题，并不具备针对具有显著差异的异源图像的匹配能力，泛化能力较差；LoFTR 算法匹配点对数目以及 RANSAC 筛选后的点对数目最多，但其误差情况并没有因此表现较好，且在部分场景也会存在匹配失败的问题。

SPG 算法和 SPG-SE 算法在 RMSE 和 MMA 上结果相近，这是因为稀疏增强策略并没有改变提取的特征点的性质，因此算法在角点缺失的图像中仍无法工作，值得说明的是，两种算法的 MMA 不高，但稀疏增强的策略能够让 SPG 算法提取出更多性质更优的角点特征，从而提升正确匹配点对在全部匹配点对中的比例，相对的降低错误匹配点对在计算单应性矩阵时的权重。此外由于先验提取的关系，相比于 SPG 算法，SPG-SE 能够在速度上提升约 30%。而 NCNet 算法相比于其他算法匹配准确率很高，且在精度上也保持了可观的性能。

2.5 本章小结

为了解决视觉辅助的无人机在实际导航定位过程中可能遇到的异源差异性场景，本章根据场景的优质特征比例将其分为了两类，分别提出了两种基于点特征的异源影像匹配技术，以应对异源差异性场景下进行图像匹配的挑战。

首先，对于优质特征较多（即显著角点数量多、范围广）的 I 类异源场景，引入了 SuperPoint 特征提取，目的是能够在光照、视角、成像手段等不同条件的影响下稳定的提取这些优质特征点。随后引入了 SuperGlue 匹配算法，以解决在纹理相似或重复区域下的特征点精确匹配。此外，为了保证无人机定位导航的实时性需求，针对图像匹配算法采取了建立先验特征数据库的策略，提出了基于 SPG 的 I 类异源影像稀疏增强匹配算法，通过特征增强+特征稀疏化的手段预先处理数字影像地图并构建数据库，相比于直接使用 SPG 算法，在保证匹配精度的同时，该算法显著提升了图像匹配的速度和鲁棒性。

最后，对于优质特征不显著的 II 类异源场景，考虑了密集特征提取的方案并引入了邻域共识网络 NCNet，提出了基于邻域共识的 II 类异源影像匹配算法，通过在两幅图像所形成的 4D 空间中查找对应的一致模式来进行图像匹配。相比于其他异源匹配算法，该算法在具有挑战性的场景下能够稳健匹配，虽然相比 SPG 算法精度有一定下降，但其具有更强的鲁棒性。

第三章　基于异源影像匹配的无人机位姿估计技术

无人机预先存储任务区域的遥感数字影像地图，在飞行过程中，底部搭载的视觉传感器以一定的频率向下采集地面影像数据，随后处理器对采集的地面影像与遥感地图进行视觉上的“对齐”，即可根据两幅图像的像素对应关系计算出无人机在采集影像时相对于遥感地图的位置和姿态，从而实现无人机视觉上的定位。如图 3.1 所示显示了基于图像匹配的无人机视觉定位过程。

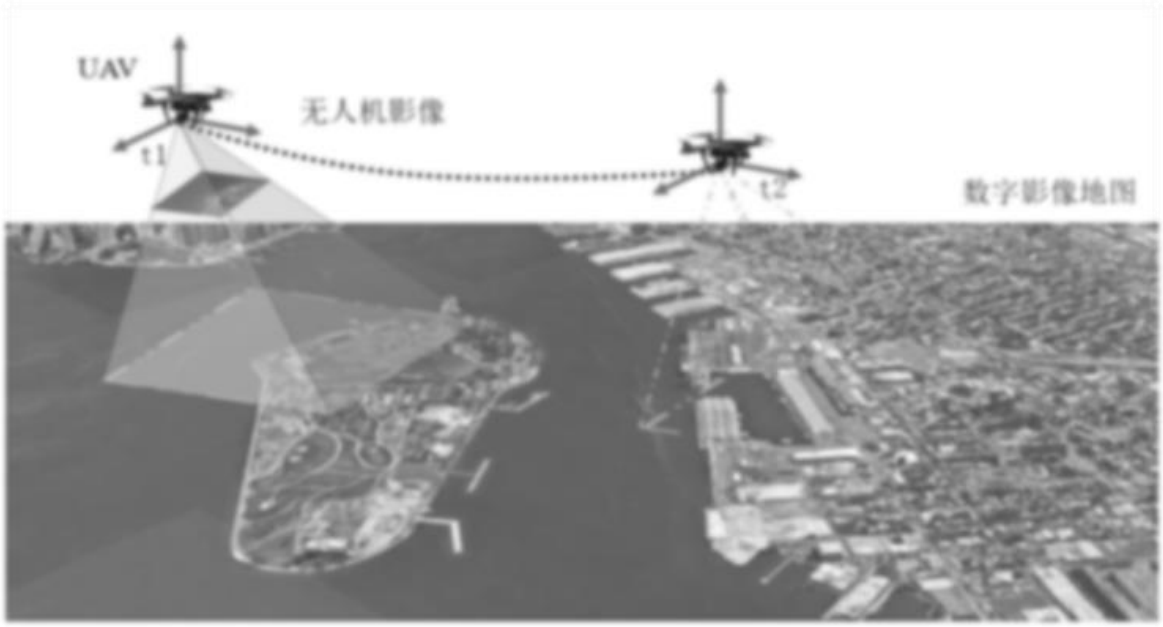


图3.1　基于图像匹配的无人机视觉导航过程

无人机在三维空间中的位置和姿态可以用一个六维向量来描述，其包含了三维的位置信息和三维的旋转信息，这个六维向量被称为无人机的位姿。无论是位置信息还是旋转信息都是相对于坐标原点而言的，因此本章将首先介绍坐标系有关概念，包括其建立和转换，在此基础上提出基于异源影像匹配的视觉位姿估计技术并建立导航方案。

3.1 无人机导航坐标系的构建和转换

图像匹配的最终目的是根据两幅图像之间的匹配关系解算出无人机获取图像时的位姿，无人机的位姿描述包括位置和姿态两部分，共六个自由度：

（1）无人机的位置信息

位置通常是指无人机在地理坐标系下的绝对位置，这种位置是三维的，包含了无

人机在三维空间中的平移运动信息。基于图像匹配获取无人机的绝对位置信息，其中两维可以从先验地图中获得（先验地图的信息中包含了地图中每一个像素在地理坐标系下的位置），第三维信息即高度将在接下来的位姿估计中进行介绍。因此位置信息使用相对于先验地图坐标原点的偏移量来进行描述，可根据实际需求将其转换为不同的地理坐标系下，例如 WGS-84（GPS 使用的坐标系）等。

（2）无人机的姿态信息

无人机的姿态通常是指无人机在世界坐标系下的姿态，描述了无人机在三维空间中的旋转运动信息，通常使用偏航角 ψ 、俯仰角 φ 、滚转角 θ 来进行表示，如图 3.2 所示，下面首先介绍导航所需坐标系的构建，随后介绍各坐标系间的转换关系。

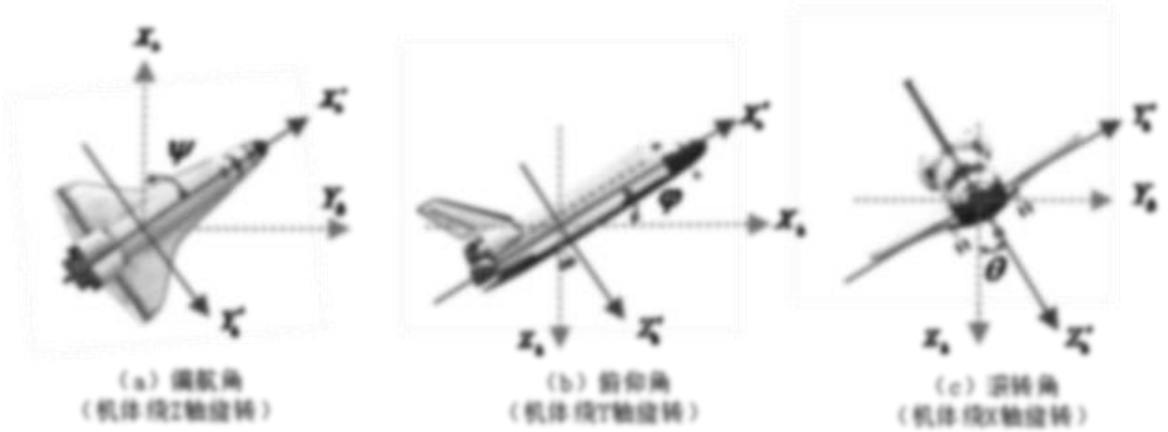


图3.2 偏航角、俯仰角与滚转角

3.1.1 导航坐标系的构建

（1）世界坐标系 $(O_w - X_w Y_w Z_w)$

世界坐标系也是导航定位的全局坐标系，通常其是一个描述物体在地球上具体位置坐标的量，由于先验地图存储了这种位置信息，实际上只需要计算出无人机在先验地图中的位置就可确定其在世界中的位置，因此为了描述方便，本文以先验地图的左上角为坐标系原点，Z_w 轴垂直与地图平面向上，图像的水平方向为 X_w 轴，竖直方向为 Y_w 轴，接下来的无人机位姿结果都是相对于该坐标系来说的。世界坐标系的单位为米。

（2）机体坐标系 $(O_b - X_b Y_b Z_b)$

由于相机通常是固定在无人机上的某一位置，通常相机安装在无人机机体的重心位置，因此相机坐标系 $(O_c - X_c Y_c Z_c)$ 和机体坐标系 $(O_b - X_b Y_b Z_b)$ 重合，这样下文关于相机坐标系和无人机机体坐标系的描述是等价的。机体坐标系原点 $O_b(O_c)$ 位于无人机的质心，$X_b(X_c)$ 轴方向从原点开始向着飞机头部方向，$Y_b(Y_c)$ 轴方向为从原点开始向着无人机右侧方向。按照右手螺旋定则，无人机 $Z_b(Z_c)$ 轴垂直于机体所在平面竖直

向下。机体坐标系单位为米。

（3）成像平面坐标系(O_1-XY)

成像平面坐标系是相机中距光心距离f（焦距）的物理成像平面，其原点O_1位于图像的正中心，也对应着相机视野的中心。该系的X轴沿图像的行方向，Y轴沿图像的列方向。成像平面坐标系的单位为米。

（4）图像坐标系(O_2-UV)

图像坐标系的建立主要是为了描述图像中的像素位置。该坐标系的原点O_2位于图像的左上角，横纵轴分别表示图像的行方向和列方向，用U和V来表示。其单位为像素。

3.1.2 导航坐标系间的转换

（1）世界坐标系与机体坐标系的转换

无人机的运动可视为刚体运动，机体共有三个轴，其任何复杂的角度变换均可以通过矩阵分解为仅依赖三个正交轴的有限次基本旋转，为了将机体坐标系与世界坐标系进行转换，定义了三个正交轴的基本旋转矩阵，下面对其进行分析。假设无人机初始位于世界坐标系原点，其机体坐标系与世界坐标系完全重合。

考虑无人机的航向发生转动，称为偏航角$\psi\in[-90^\circ,90^\circ]$，则对应新的机体坐标系是原机体坐标系$O_b-X_bY_bZ_b$绕Z轴顺时针转$\psi$得到的，如图3.3（a）所示，新的坐标系记为$O_\psi-X_\psi Y_\psi Z_\psi$，旋转矩阵为：

$$C_\psi^b=\begin{bmatrix}\cos\psi & \sin\psi & 0\\ -\sin\psi & \cos\psi & 0\\ 0 & 0 & 1\end{bmatrix} \tag{3-1}$$

若无人机的俯仰变化角度为$\varphi\in[-90^\circ,90^\circ]$，即新的无人机机体坐标系是绕无人机的两幅机翼连接线的平行轴 Y 轴产生了转动，如图3.3（b）所示，新的坐标系记为$O_\varphi-X_\varphi Y_\varphi Z_\varphi$，旋转矩阵为：

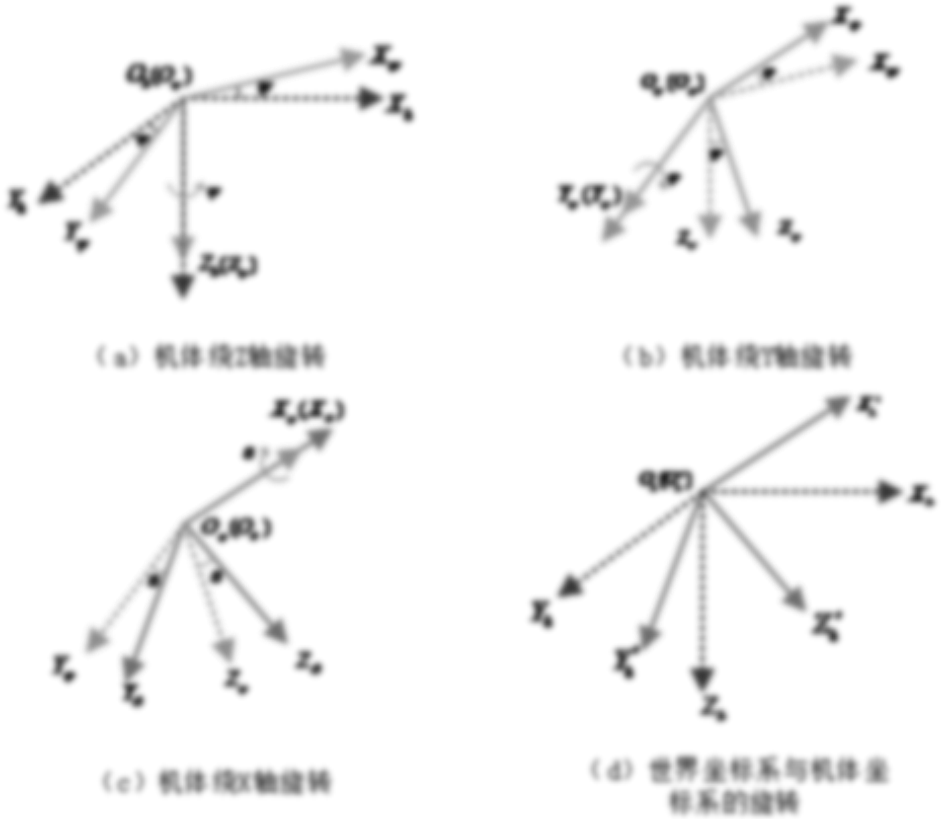
$$C_\varphi^b=\begin{bmatrix}\cos\varphi & 0 & -\sin\varphi\\ 0 & 1 & 0\\ \sin\varphi & 0 & \cos\varphi\end{bmatrix} \tag{3-2}$$

若无人机绕飞行方向的中心轴发生了转动，即有一个滚转角$\theta\in(-180^\circ,180^\circ]$，如图3.3（c）所示，新的坐标系记为$O_\theta-X_\theta Y_\theta Z_\theta$，旋转矩阵为：

$$C_\theta^b=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & \sin\theta\\ 0 & -\sin\theta & \cos\theta\end{bmatrix} \tag{3-3}$$

以上三个矩阵描述了无人机绕三轴的旋转变换，接下来考虑一种任意旋转情况，如图 3.3 所示，无人机初始机体坐标系为$O_b-X_bY_bZ_b$，首先将其绕Z_b轴旋转ψ，得到新的机体坐标系$O_\psi-X_\psi Y_\psi Z_\psi$，在新的坐标系基础上绕$Y_\psi$轴旋转$\varphi$，得到新的机体坐标系$O_\varphi-X_\varphi Y_\varphi Z_\varphi$，同理再绕$X_\varphi$轴旋转$\theta$，这样得到新的坐标系$O_b'-X_b'Y_b'Z_b'$，其复合旋转矩阵表示如下：

$$\begin{aligned}\mathbf{R}&=C_\theta^\varphi C_\varphi^\psi C_\psi^W\\ &=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & \sin\theta\\ 0 & -\sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}\cos\varphi & 0 & -\sin\varphi\\ 0 & 1 & 0\\ \sin\varphi & 0 & \cos\varphi\end{bmatrix}\begin{bmatrix}\cos\psi & \sin\psi & 0\\ -\sin\psi & \cos\psi & 0\\ 0 & 0 & 1\end{bmatrix}\\ &=\begin{bmatrix}\cos\varphi\cos\psi & \cos\varphi\sin\psi & -\sin\varphi\\ -\cos\theta\sin\psi+\sin\theta\sin\varphi\cos\psi & \cos\theta\cos\psi+\sin\theta\sin\varphi\sin\psi & \sin\theta\cos\varphi\\ \sin\theta\sin\psi+\cos\theta\sin\varphi\cos\psi & -\sin\theta\cos\psi-\cos\theta\sin\varphi\sin\psi & \cos\theta\cos\varphi\end{bmatrix}\end{aligned} \tag{3-4}$$



（a）机体绕Z轴旋转　（b）机体绕Y轴旋转

（c）机体绕X轴旋转　（d）世界坐标系与机体坐标系的旋转

图3.3　世界坐标系与机体坐标系间的旋转变换

该旋转空间 **R** 可以覆盖旋转的任意情况，实际中无人机位姿除了旋转变化外，还会产生移动，因此还需要一个位移描述量，使用三维平移向量 $\mathbf{T}=\begin{bmatrix}x_{oc} & y_{oc} & z_{oc}\end{bmatrix}^T$ 来进行描述，即对于任意的无人机机体坐标系 $O_b-X_bY_bZ_b$，其在世界坐标系下可以表示为由初始坐标系（与世界坐标系重合）原点平移位置量 **T** 然后再做一个旋转操作 **R** 得到。即对于世界坐标系下的任意一点 $P_w(x_w,y_w,z_w)$，将其转换到无人机机体坐标系下的点 $P_b(x_b,y_b,z_b)$（注意该点并不随坐标系的变化而发生运动），满足关系式：

$$\begin{bmatrix}x_b\\y_b\\z_b\\1\end{bmatrix}=\begin{bmatrix}\mathbf{R} & \mathbf{T}\\0 & 1\end{bmatrix}\begin{bmatrix}x_w\\y_w\\z_w\\1\end{bmatrix} \tag{3-5}$$

其中 $[\mathbf{R},\mathbf{T};0,1]$ 是外参矩阵，描述了相机坐标系从与世界坐标系重合到当前位置的运动，换句话说，其也表示了无人机在世界坐标系下的位姿，图像匹配的目的是根据匹配点对的关系求解这两个坐标系间的旋转矩阵 **R** 和平移矩阵 **T**，并进一步解算无人机的六维状态 $P=\begin{bmatrix}x_v,y_v,z_v,\psi_v,\varphi_v,\theta_v\end{bmatrix}^T$，对于计算得到的一个外参矩阵有：

$$\begin{bmatrix}\mathbf{R} & \mathbf{T}\\0 & 1\end{bmatrix}=\begin{bmatrix}r_0 & r_1 & r_2 & t_1\\r_3 & r_4 & r_5 & t_2\\r_6 & r_7 & r_8 & t_3\\0 & 0 & 0 & 1\end{bmatrix} \tag{3-6}$$

根据式(3-4)、式(3-5)和式(3-6)联立可解得，

$$\begin{cases}x_v=t_1\\y_v=t_2\\z_v=t_3\\\psi_v=\arctan(r_1/r_0)\\\varphi_v=-\arcsin(r_2)\\\theta_v=\arctan(r_5/r_8)\end{cases} \tag{3-7}$$

（2）机体坐标系与成像平面坐标系的转换

本文中机体坐标系与成像平面坐标系间的关系符合针孔相机模型（机体坐标系等价于相机坐标系），如图 3.4 所示，

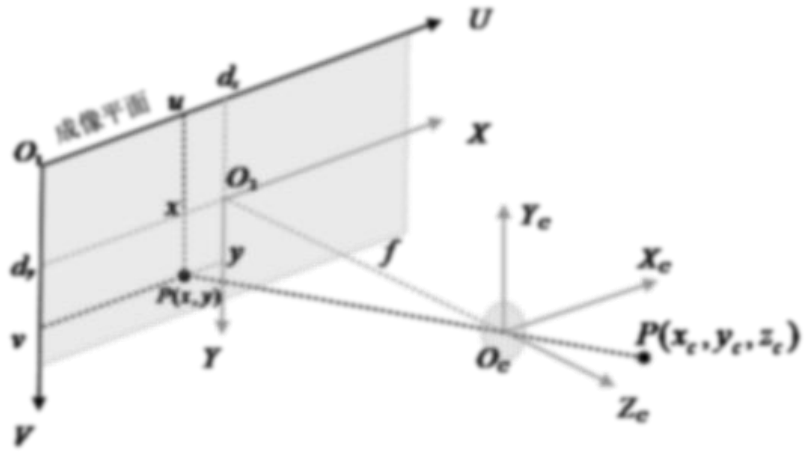


图3.4 针孔相机模型

对于相机坐标系$O_c - X_cY_cZ_c$下任意一点$P(x_c, y_c, z_c)$，经过相机光心O_c投影到成像平面坐标系$O-XY$下的点$P'(x, y)$，焦距为f（单位为米），则根据三角形相似关系有：

$$\frac{z_c}{f} = -\frac{x_c}{x} = -\frac{y_c}{y} \tag{3-8}$$

由于小孔成像是倒立成像的，然而在实际中，相机通常会在后期将成的倒立图像自动进行翻转，因此在这里不考虑负号，整理可得，

$$\begin{cases} x = f\dfrac{x_c}{z_c} \\ y = f\dfrac{y_c}{z_c} \end{cases} \tag{3-9}$$

这样相机坐标系中的三维点就会被转换为成像平面坐标系的二维点。

（3）成像平面坐标系与图像坐标系的转换

对于成像平面坐标系$O-XY$中的点$P'(x, y)$，式(3-9)描述了点 P 和它的像之间的空间关系，但通常相机采集的图像是由像素组成的，因此还需要对空间进行采样和量化，为此建立图像坐标系$O-UV$，将其转换为像素坐标$P''(u, v)$。

首先需要将原图像中心的坐标系原点平移到图像的左上角，x方向和y方向的平移量分别为c_x、c_y，通常这两个坐标系的转换还包含了缩放系数，假设在U轴方向上缩放了α倍，在V轴方向上缩放了β倍，则有

$$\begin{cases} u = \alpha x_c + c_x \\ v = \beta y_c + c_y \end{cases} \tag{3-10}$$

根据式(3-9)和式(3-10)整理可得从相机坐标系到图像坐标系的完整转换关系，

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z_c} \begin{bmatrix} f_x & 0 & d_x \\ 0 & f_y & d_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \frac{1}{z_c} KP \tag{3-11}$$

其中，$f_x = \alpha f$，$f_y = \beta f$，α 和 β 的单位为像素/米，f_x 和 f_y 的单位均为像素。中间的量组成的矩阵 K 即是相机的内参数矩阵，每个相机的内参矩阵均是固定的，可通过对相机进行标定得到。

3.2 基于异源影像匹配的视觉位姿估计技术

通过图像匹配得到的匹配点对之间的映射关系仅仅表示了所采集的图像位于先验地图中的具体位置，以及两幅图像像素间的映射关系，而不是无人机的姿态。获取映射关系仅仅是第一步，据此解算出无人机的姿态才是图像匹配的主要目的，本小节将就如何基于图像的匹配点对实现无人机的位姿解算做出介绍。

无人机的位姿解算本质上是求解无人机的运动，即在世界坐标系下，无人机的机体坐标系从状态 A 运动到状态 B 的过程。其可以通过一个旋转矩阵 **R** 和一个平移矩阵 **T** 来进行描述。本文采用了 PnP 模型建立二维图像到三维地图坐标点的关系，进而实现对无人机位姿状态的估计。

PnP 模型如图 3.5 所示，其是一种求解 3 维到 2 维点对运动的方法，它描述了当知道 n 个 3 维空间点及其投影位置时如何估计相机的位姿，通常特征点的 3 维位置可以通过以下方式确定：（1）无人机搭载的先验数字影像地图包含高度信息，例如数字高程模型（DEM）或三维点云等；（2）使用无人机搭载的传感器采集图像的同时采集深度信息，例如采用双目或者 RGB-D 深度相机等；（3）三角测量法，即通过在两个位置观察同一个点的夹角从而确定与该点的距离。（4）同一平面假设，假设无人机在一定高度时，忽略相机视野范围内的地形起伏，认为其所有像素点在同一平面上。

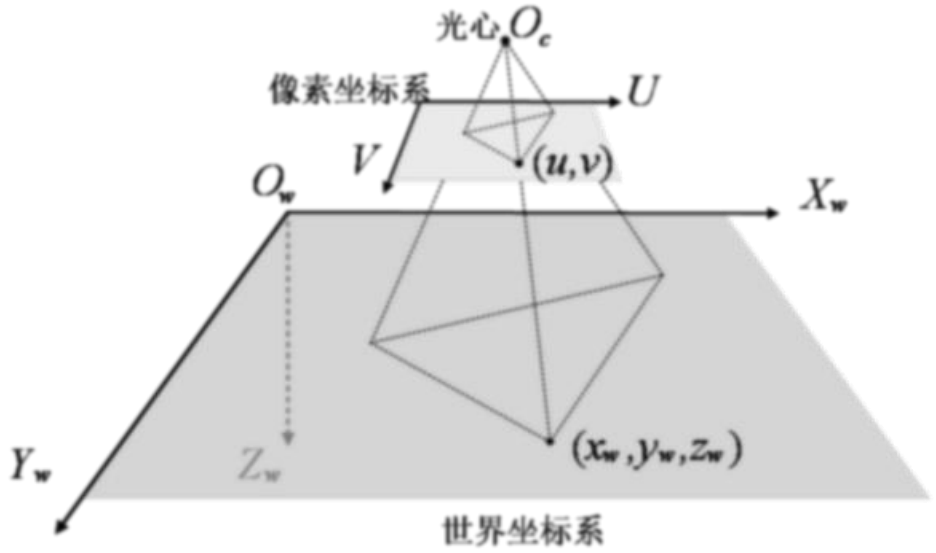


图3.5 PnP 模型示意图

基于本研究所使用的异源图像地图数据源的考虑，本文选择同一平面假设来建立先验数字影像地图的三维点信息，在提取特征点之后使用平面约束筛选掉不在主平面上的特征点，从而实现位姿解算算法的精度保证。

PnP 问题可以使用 P3P[60]、EPnP[61]、UPnP[62]，或非线性优化如光束法平差（BA，Bundle Adjustment）等方式进行求解，为了能够充分的利用匹配点对的信息，本文采用直接线性变换（DLT，Direct Linear Transform）来进行求解，并联合其他算法对其进行迭代优化。

考虑空间中的某点 $P(x_w,y_w,z_w)$，为了计算方便，将其转为齐次坐标表示 $\mathbf{P}=[x_w,y_w,z_w,1]^T$，对应与图像坐标系下的像素点齐次坐标 $(u_1,v_1,1)$，定义增广矩阵 $[\mathbf{R},\mathbf{T}]$，其为一个 3×4 的矩阵，将其代入式(3-6)和(3-11)可得：

$$z_c\begin{bmatrix}u\\v\\1\end{bmatrix}=\mathbf{K}\begin{bmatrix}\mathbf{R} & \mathbf{T}\end{bmatrix}\begin{bmatrix}x_w\\y_w\\z_w\\1\end{bmatrix}=\mathbf{K}\begin{bmatrix}r_1 & r_2 & r_3 & t_1\\r_4 & r_5 & r_6 & t_2\\r_7 & r_8 & r_9 & t_3\end{bmatrix}\begin{bmatrix}x_w\\y_w\\z_w\\1\end{bmatrix} \tag{3-12}$$

根据最后一行消去 z_c 可得到两个约束：

$$\begin{cases}u_1=\dfrac{r_1x_w+r_2y_w+r_3z_w+t_1}{r_7x_w+r_8y_w+r_9z_w+t_3}\\[2ex]v_1=\dfrac{r_4x_w+r_5y_w+r_6z_w+t_2}{r_7x_w+r_8y_w+r_9z_w+t_3}\end{cases} \tag{3-13}$$

定义$[\mathbf{R},\mathbf{T}]$的行向量$\mathbf{t}_1=[r_1,r_2,r_3,t_1]^T$，$\mathbf{t}_2=[r_4,r_5,r_6,t_2]^T$，$\mathbf{t}_3=[r_7,r_8,r_9,t_3]^T$，于是有：

$$\begin{cases}\mathbf{t}_1^T\mathbf{P}-\mathbf{t}_3^T\mathbf{P}u_1=0\\ \mathbf{t}_2^T\mathbf{P}-\mathbf{t}_3^T\mathbf{P}v_1=0\end{cases} \tag{3-14}$$

其中 $\mathbf{t}_1$、$\mathbf{t}_2$、$\mathbf{t}_3$ 是待求的量，可以看到，每个特征点提供了两个关于 $\mathbf{t}$ 的线性约束方程，假设一共有 N 个特征点，则可以得到如下线性方程组，

$$\begin{bmatrix}\mathbf{P}_1^T & \mathbf{0} & -u_1\mathbf{P}_1^T\\ \mathbf{0} & \mathbf{P}_1^T & -v_1\mathbf{P}_1^T\\ \vdots & \vdots & \vdots\\ \mathbf{P}_N^T & \mathbf{0} & -u_N\mathbf{P}_N^T\\ \mathbf{0} & \mathbf{P}_N^T & -v_N\mathbf{P}_N^T\end{bmatrix}\begin{bmatrix}\mathbf{t}_1\\ \mathbf{t}_2\\ \mathbf{t}_3\end{bmatrix}=\mathbf{0} \tag{3-15}$$

由于 t1、t2、t3 一共有 12 维，因此至少需要 6 对匹配点才能实现矩阵$[\mathbf{R},\mathbf{T}]$的线性求解，当匹配点对多于 6 对时，本文采用 MAGSAC 算法对其进行迭代优化。

在消除误匹配点对中，最常用的算法是 RANSAC 算法，通过迭代求解找到一个符合尽可能多的匹配点对的模型，但其需要人为的设定内点的阈值，以求取一个内点点集，这个阈值很大程度上影响了模型的评估。为了解决这个问题 Daniel Barath 等人提出了 MAGSAC 算法[63]，该算法将内点的阈值看做是一个随机变量，通过边缘化内点阈值，计算每个特征点是内点的概率，将其作为一个权重进行加权最小二乘法优化，从而迭代计算得到一个最优模型。

通过这种方式得到的矩阵$[\mathbf{R},\mathbf{T}]$包含的是无人机从起点到当前位置的运动模型，即旋转量和平移量，由于无人机起点坐标被初始化为与世界坐标系重合，因此无人机当前位置的运动模型也可以视为无人机在世界坐标系下的状态。

3.3 基于异源影像匹配的无人机定位技术流程

通常对于无人机先后采集的两幅连续图像帧来说，其重叠区域可能达到 80%甚至 90%（这取决于无人机的飞行速度和传感器设备的采集速率设置），对于每一对相邻帧都采取图像匹配的策略是不必要的。考虑这两幅图像属于同源图像，无论是光照、角度等条件变化均很小，为了降低图像匹配算法对资源的消耗，满足位姿估计的实时性需求，可通过光流跟踪的方式来处理相邻帧。

无人机定位技术的算法流程如图 3.6 所示，所需信息包括高度计、惯性导航系统 INS、无人机影像和根据卫星数字影像地图构建的先验特征数据库，输出视觉估计的

无人机的位置和姿态。此外，为了保证视觉定位的速度、精度和鲁棒性，本研究还采用了如下策略：

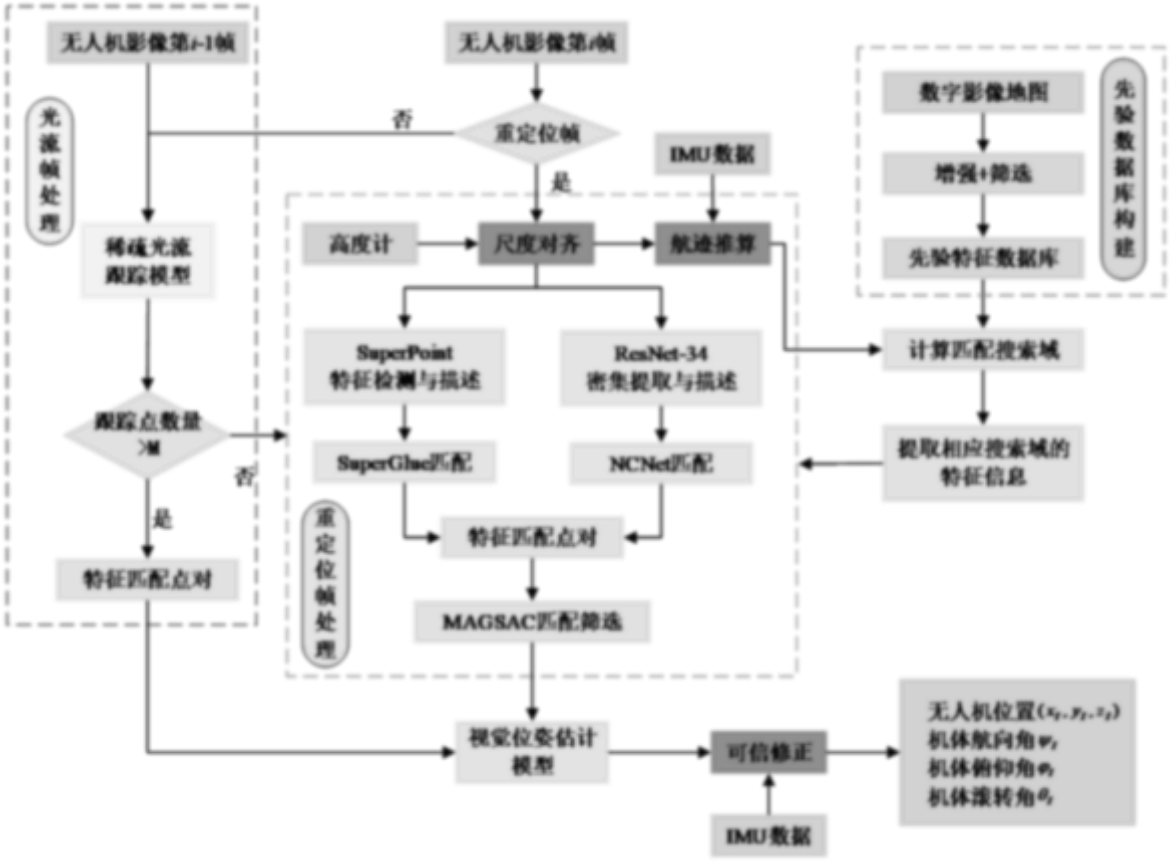


图3.6 基于异源影像匹配的无人机定位技术基本流程

（1）航迹推算

在无人机飞行过程中，由于无人机的运动过程是平滑连续的，这反映在图像上就是无人机当前时刻获取的影像相对于上一时刻的影像必然存在较大的重叠范围，因此在先验特征数据库中检索时只需要在上一时刻位置附近进行检索即可，这样能够大大降低匹配的时间消耗。

本文采取的策略是进行航迹推算，即假定上一时刻 t_1 的位姿（位置和姿态）是准确的，根据 INS 的短时间高精度的特点，用 INS 数据与上一时刻位姿进行姿态解算，就可以得到当前时刻的位姿估计值，然后根据估计结果去数据库相应位置的邻域进行搜索，从而快速得到匹配点对关系。然而实际运动过程中 t_1 时刻的位姿并不总是可靠的，因此在数据库中搜索时可能需要扩大搜索范围。

（2）尺度对齐

在对无人机影像和基准地图配准处理中，先验地图的尺度往往是确定的，而无人机的飞行高度可能会时刻改变，因此还需要注意到地图和无人机影像间的尺度问题，

即需要根据先验地图的尺度和无人机实时飞行高度等信息，将无人机影像的尺度变换到和数字影像地图大致同一尺度下，尽管图像匹配算法能够处理图像间的缩放关系的问题，但先将其变换到与先验地图同一尺度下能够更好的界定搜索空间的范围，通常能够使图像匹配算法表现更好的效果。

（3）可信修正

在无人机实际飞行过程中，由于地图的复杂性，图像匹配算法并不总是能将无人机图像与先验数字影像地图全部准确匹配，因此需要对当前的匹配结果进行验证，确定其是否是一个有效可信的位姿状态。

当相邻的两个时刻解算到的位置之间所需速度远大于无人机当前的飞行速度，表现在数据上即图像匹配得到的位姿结果与上一时刻相比是跳跃的且阶跃值大于阈值（该阈值与无人机帧采集速度和飞行速度呈正相关），应当将该位姿结果视为一个无效数据。为了填补该数据的缺失，利用 INS 的短时高精度特性，使用上一时刻无人机的位姿（融合值）和 INS 采集的加速度数据计算当前帧的位姿来作为视觉解算数据，并将下一帧设置为重定位帧，重新进行视觉定位。

本研究设计的视觉导航图像匹配方法实现步骤如下：

I. 先验处理阶段

对数字影像地图进行具有重叠区的分块处理；由于特征提取算法关注的是一个像素点的周围邻域区域，因此在分块后据边界的 Δx 部分内的像素特征描述会丢失信息，为了充分提取和描述这部分图像，需要在分块时各分块的边界存在一定的重叠，对于 SPG-SE 算法来说，$\Delta x = 8$ 像素。

使用数据增强的方式对每一个图像块进行数据增强，包括 1/2 降采样、添加噪声、随机单应性变换等方式组合获取增强图像，并使用 SuperPoint 算法提取特征与描述符，选取能够稳定性达到 50%（经验值）以上的特征作为备选特征，随后使用稀疏化方法再次筛选特征，将其存储在先验特征数据库中。

II. 实际飞行阶段

（1）根据当前无人机高度，将当前无人机影像帧尺度缩放至与先验地图尺度一致，记录尺度缩放因子 σ；

（2）借助当前时刻 INS 速度和加速度 $[v_t, a_t]$ 与无人机上一时刻的位置参数 P_{t-1} 进行航迹推算，预测无人机相机视角的中心位置 P_t。

（3）若当前无人机影像帧序号能被 N_h 整除则确定为重定位帧，执行基于图像匹配的视觉位姿估计算法，否则确定为光流帧，执行步骤（6）。

（4）先验数字影像地图数据库提取描述符。SPG 算法：从先验地图数据库中使用 SuperPoint 网络提取 P_t 周围 1k×1k 像素区域的特征描述信息 $\mathbf{KP}^{map} = [kp_1^{map}, ..., kp_n^{map}]$，表示该区域一共提取到了 n 个特征点，对于每个特征点 $kp_i^{map} = (x_i^{map}, y_i^{map})$ 有对应的

$des_i^{map}=[d_i^1,...,d_i^{128}]$，表示在地图中位置$(x_i,y_i)$的特征点的128维描述符；

（5）无人机影像提取描述符。a. SPG算法：对无人机拍摄的影像使用SuperPoint算法提取特征和描述符$\mathbf{KP}^{uav}=[kp_1^{uav},...,kp_t^{uav}]$，使用SuperGlue算法对两组描述符$\mathbf{KP}^{map}$和$\mathbf{KP}^{uav}$进行匹配，获得匹配结果点对$T_{coarse}=[kp_1^{map},kp_1^{uav};...;kp_{mc}^{map},kp_{mc}^{uav}]$；b.NCNet算法：使用ResNet-34网络提取P_t周围邻域的密集描述f^{map}，将无人机拍摄的影像输入到ResNet-34网络中得到密集描述f^{uav}，随后计算f^{map}与f^{uav}之间的全部特征的余弦相似性，将其输入邻域共识网络NCNet，得到匹配点对$T_{coarse}=[kp_1^{map},kp_1^{uav};...;kp_{mc}^{map},kp_{mc}^{uav}]$；

（6）根据上一帧无人机影像提取到的特征点$T_{coarse}[:,2]=[kp_1^{uav};kp_2^{uav};...;kp_{mc}^{uav}]$，使用稀疏光流法跟踪得到这些特征点在当前帧的位置，并执行一次反向稀疏光流筛选掉错误跟踪点得到$T_{flow}=[kp_1^{uav2};kp_2^{uav2};...;kp_{mf}^{uav2}]$，据匹配关系使用跟踪点$T_{flow}$替换匹配$T_{coarse}[:,2]$，得到新的$T_{coarse}$。

（7）使用边缘样本共识MAGSAC算法从全部的匹配点对T_{coarse}中筛选符合平面一致性的匹配点对T_{fine}，并根据缩放因子σ，将计算的匹配点对坐标映射回原始图像中。

（8）对筛选后的精匹配点对T_{fine}使用PnP位姿解算模型对无人机的位姿进行估计，计算该无人机影像拍摄时无人机的绝对位姿$P_v=[x,y,z,\psi,\varphi,\theta]$。

（9）获取INS数据，根据上一帧的无人机位姿状态与加速度数据估计当前帧的无人机位姿$P_{INS}=[x,y,z,\psi,\varphi,\theta]$，根据无人机速度$v$设置阈值$\delta(v)$，当视觉绝对位姿与INS估计位姿即$\|P_v-P_{INS}\|\le\delta(v)$时认为视觉位姿可信，将下一帧设置为光流帧；当$\|P_v-P_{INS}\|>\delta(v)$时视觉位姿不可信，用INS航迹推算值替换视觉位姿，并将下一帧设置为重定位帧，重新执行视觉匹配位姿估计算法。

3.4 基于异源影像匹配的高低空无人机协同位姿估计系统

尽管基于邻域共识的图像匹配方法显示出了很好的鲁棒性，但视觉定位方法有一个根本缺陷：完全不一致的图像无法匹配。这往往发生在飞行高度较低的无人机上，由于低空无人机视野范围狭小，异源性差异大时将很难找到一致对应关系。而高空无人机由于具有更广阔的视野，能够更加关注整体地形地貌，因此视觉定位失败的概率要显著低于低空无人机。此外，低空无人机图像会包含更多地面细节，因此定位精度往往较高。考虑到以上方面，本文提出了一种基于异源影像匹配的高低空无人机协同导航方案，以在具有挑战性的区域实现无人机长时间的稳定导航定位。

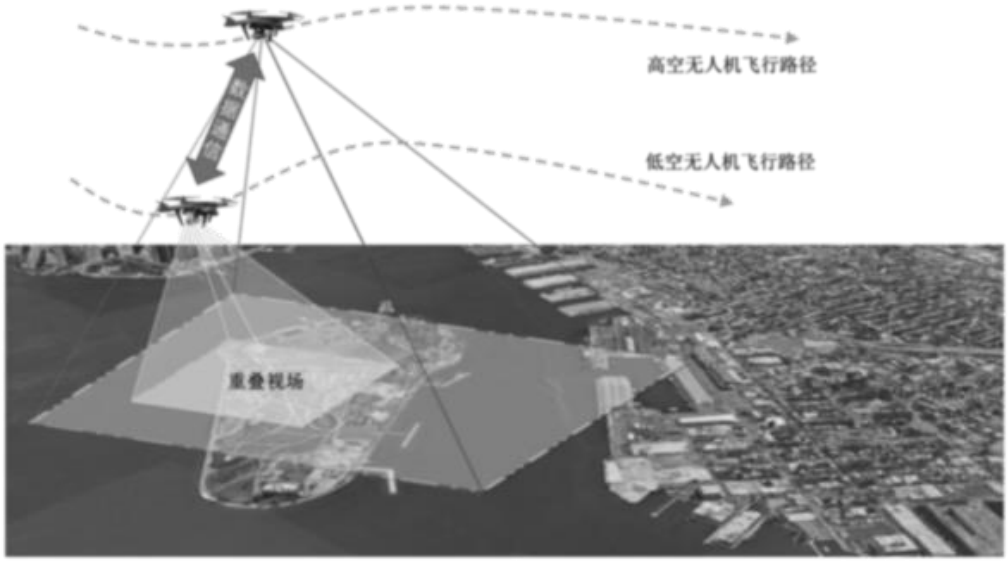
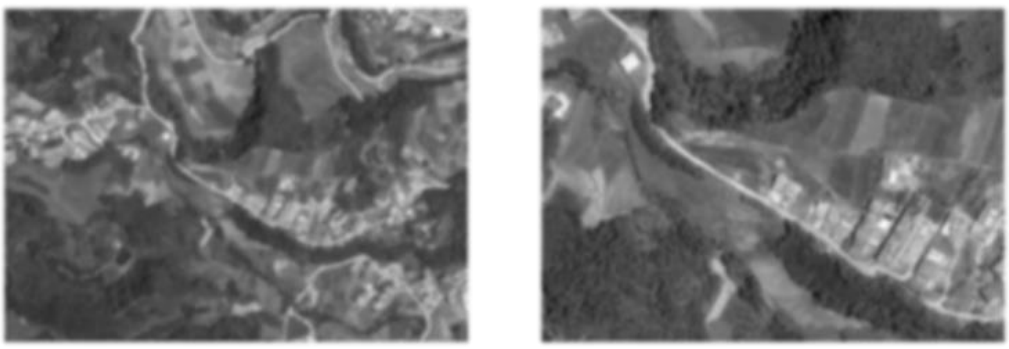


图3.7 无人机视觉协同定位示意图

如图 3.7 所示显示了无人机协同视觉定位导航的示意图，高低空无人机首先独立进行位姿解算和定位，当高空或低空一方定位失准时，基于重叠视场获取相对位姿关系用于对失准方的位姿结果进行校正。低空是相对于高空来说的，高空具有更广的视野也能够探测到更大的地面范围，其处理对象通常是如建筑群系或地形地貌的边缘、轮廓等，相应的对于具体的目标如一栋单独的房屋其可能在高空无人机成像影像中为孤立的像素点。

如图 3.8 所示分别显示了同一区域的高低空无人机的视野示意图，相比于高空影像，低空影像的一个显著特点是地面目标物特征细节更加丰富，例如建筑、道路、树木、河流等的形状、角点、轮廓、纹理、颜色等特征，这样在低空场景下，能够提取更多的特征信息用于其它任务。当然从另一个方面来说，由于低空无人机视野范围较小，在飞行经过诸如林地、湖泊等区域上空时，无人机狭小的视野范围将难以提取到有效特征从而无法完成图像匹配。



（a）高空无人机视野　　（b）低空无人机视野

图3.8 高低空无人机视野示意图

在基于图像匹配的视觉位姿估计过程中，当连续多帧的重定位失败（即图像匹配失败），此时每一帧均不得不使用 IMU 数据进行航迹推算来作为视觉临时值，这样长时间下 IMU 的累计误差将导致融合得到的位姿估计值产生越来越大的偏差，最终定位失准。

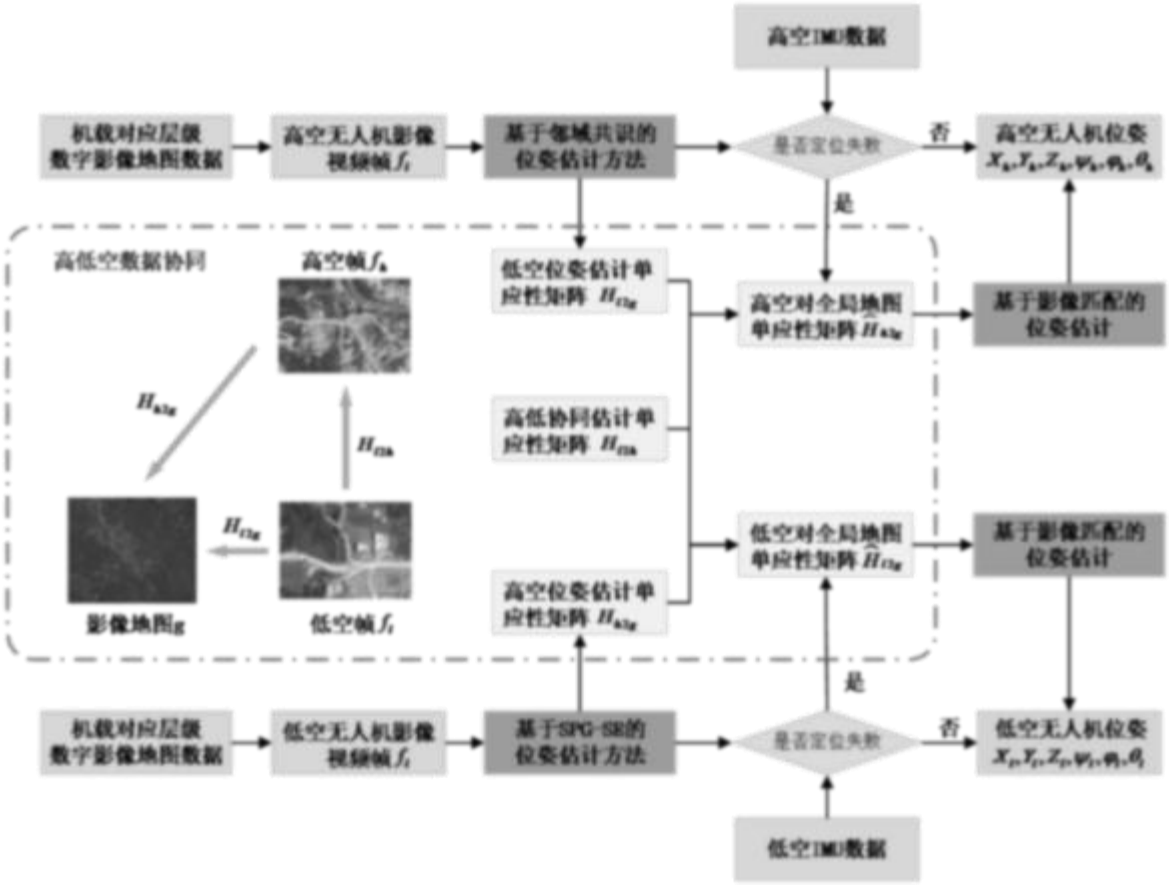


图3.9　高低空无人机视觉协同位姿估计流程图

如图 3.9 所示，高低空无人机视觉协同定位框架由高空无人机位姿估计、高低空视觉数据协同和低空无人机位姿估计三个部分构成，其中高低空无人机之间需要进行数据通信，在实际工作中根据数据时间戳进行同步，本文假设通信处于理想状态下。

高低空无人机的视觉位姿估计选择了不同的算法，主要基于以下方面考虑：低空图像相对于高空来说一方面内容上具有更多的细节信息，能够更关注物体的视觉角点特征，从而具有更高的精度；另一方面低空无人机视觉场景变化迅速，对图像匹配的速度提出了更高要求，因此使用 SPG-SE 算法来进行图像匹配。相似的，高空图像相对于低空图像来说具有更多的内容，且视觉场景变化不那么迅速，较为适合 NCNet 这种基于邻域搜索的密集匹配模型。此外，高空无人机在整个协同导航方案中应当是作为核心枢纽的存在，需要能够提供持续稳定的视觉位姿信息，因此选择能够在更复杂场景下实现稳定匹配的 NCNet 的算法。假设在低空无人机视觉定位失败或视觉位姿

不可信时，协同位姿更新流程如下（高空同理）：

a) 获取高空无人机影像帧 f_h、低空无人机影像帧 f_l 以及数字影像全局地图 I_g。

b) 对高空影像 f_h 与全局地图 I_g 使用 SPG-SE 图像匹配算法计算得到图像间的单应性变换矩阵 H_{g2h}。对于高空影像帧任意坐标点 $p^h=(x_p^h,y_p^h)$，其在全局地图匹配的对应点假设为 $p^g=(x_p^g,y_p^g)$，有

$$\begin{bmatrix} x_p^g \\ y_p^g \\ 1 \end{bmatrix} = H_{g2h} \begin{bmatrix} x_p^h \\ y_p^h \\ 1 \end{bmatrix} \tag{3-16}$$

c) 对低空影像 f_l 和高空影像 f_h 使用基于邻域共识的图像匹配算法解算对应单应性变换矩阵 H_{h2l}，则对于低空影像任意坐标点 $p^l=(x_p^l,y_p^l)$，其在全局地图的对应匹配点为 $p^g=(x_p^g,y_p^g)$，则有

$$\begin{bmatrix} x_p^g \\ y_p^g \\ 1 \end{bmatrix} = H_{g2h}\cdot H_{h2l}\cdot \begin{bmatrix} x_p^l \\ y_p^l \\ 1 \end{bmatrix} = H_{g2l}\cdot \begin{bmatrix} x_p^l \\ y_p^l \\ 1 \end{bmatrix} \tag{3-17}$$

d) 通过高低空图像匹配关系使用基于图像匹配的无人机位姿估计技术重新估计当前低空无人机位姿，若仍然失败，则使用 IMU 数据进行航迹推算。

3.5 本章小结

为了实现视觉辅助无人机导航的需求，本章节主要研究了基于异源影像匹配的无人机位姿估计技术，即如何基于图像匹配结果估计无人机拍摄图像时的六轴姿态。

首先，为了实现对无人机位姿的描述，对本研究中坐标系的构建及各坐标系间的数据转换进行了介绍。随后为了从匹配点对的映射关系中解算出对应的相机姿态，介绍了 PnP 模型并阐述了 PnP-MAGSAC 位姿估计技术的基本原理和公式推导。

考虑到图像匹配的速度问题，提出重定位帧的策略，通过帧间光流跟踪获取相对位姿关系而不要求每一帧都进行匹配。最后将整个视觉方案进行整合提出基于异源影像匹配的无人机位姿估计技术，仅依靠先验数字影像地图和无人机视频影像即可解算出无人机拍摄视频帧时的六轴姿态。

此外，为了提高视觉位姿估计的速度和精度，通过读取 IMU 数据进行航迹推算获取当前帧的无人机可能位置，一方面可以减少图像匹配的搜索空间从而提升匹配速度，另一方面可以限制图像匹配得到的位姿结果，从而进行可信修正。

最后，为了保障在复杂环境下的视觉定位，借助于高低空无人机的特点，本文给出了一种高低空视觉协同导航框架，以保证视觉辅助导航方案在视觉长时间失准情况下的稳定性。

第四章　基于卡尔曼的多源信息融合导航处理技术

视觉定位系统的定位精度不受之前时刻的影响，而完全取决于当前时刻图像匹配和位姿解算的精度，具有长时间稳定性和准确性的特点。但由于图像匹配过程需要一定的时间，无法满足无人机导航定位的实时性需求，此外，视觉匹配受限于工作区域的视觉特征，对于部分地区仍有可能出现匹配失败或定位误差过大的情况，因此基于图像匹配的视觉位姿估计方案通常是作为一种辅助手段使用。

在无人机导航中，惯性导航系统 INS 往往是一个必备的选择，其不需要采集外界信息、不受外界环境条件干扰、更新频率高、短期内精度高，能够满足无人机导航定位的全部需求。然而单独使用 INS 时，由于其完全依赖于初始状态，因此飞行过程中误差会不断累积，无法适应长时间的导航需求。

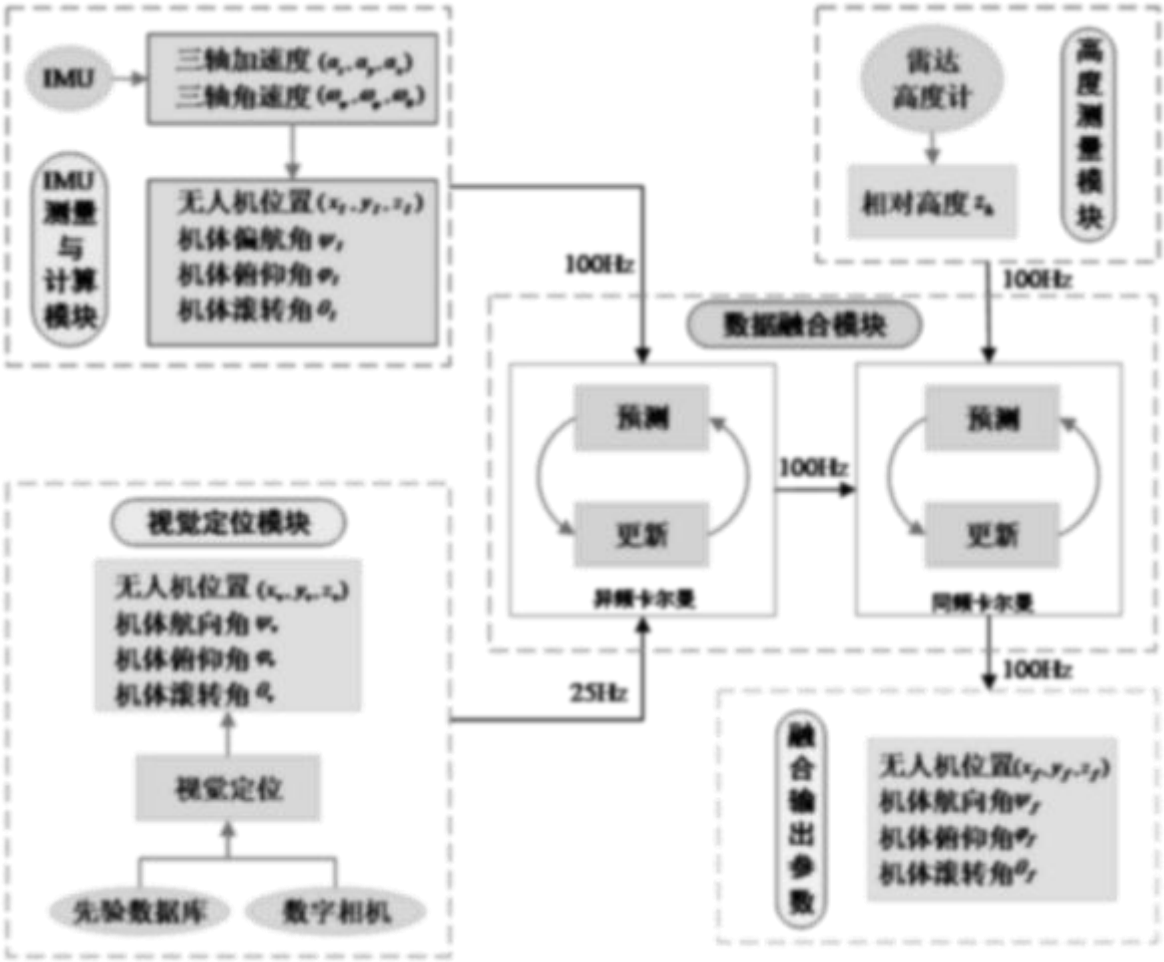


图4.1　多源传感器融合导航流程

因此，利用视觉定位长期准确性和 INS 短期高精度的特点，视觉定位系统和惯性导航系统能够借助彼此的优势进行互补，通过对 INS 数据和视觉定位数据进行融合，

可以有效降低 INS 累积误差的影响，并且弥补视觉定位更新频率低的不足。本文提出的基于视觉定位和 INS、雷达高度计的多源传感器融合导航处理流程如图 4.1 所示。

4.1 惯性导航系统基本原理介绍

惯性导航系统（INS）是一种不依赖于外界信息，仅通过测量载体在惯性导航系下的加速度来计算载体速度、位置、偏航角等信息的自主式导航部件。其有着隐蔽性好、全天候、工作范围广、更新率高、短期精度好、稳定性高等优点，广泛应用于一些战略、战术武器中，而随着微电子机械系统等器件的发展，目前商业级、消费品级的 INS 也已经成为了很多飞机、车辆等的必备系统部件[65]。

INS 使用加速度计和陀螺仪等惯性器件，利用基准方向和初始的位置信息来确定载体的方位、位置和速度，该系统根据陀螺仪的输出建立导航坐标系，根据加速度计的输出来解算载体的速度和相对位移。其中加速度计和陀螺仪被统称为惯性测量单元（IMU，Inertial Measurement Unit），其解算位姿过程如图 4.2 所示。

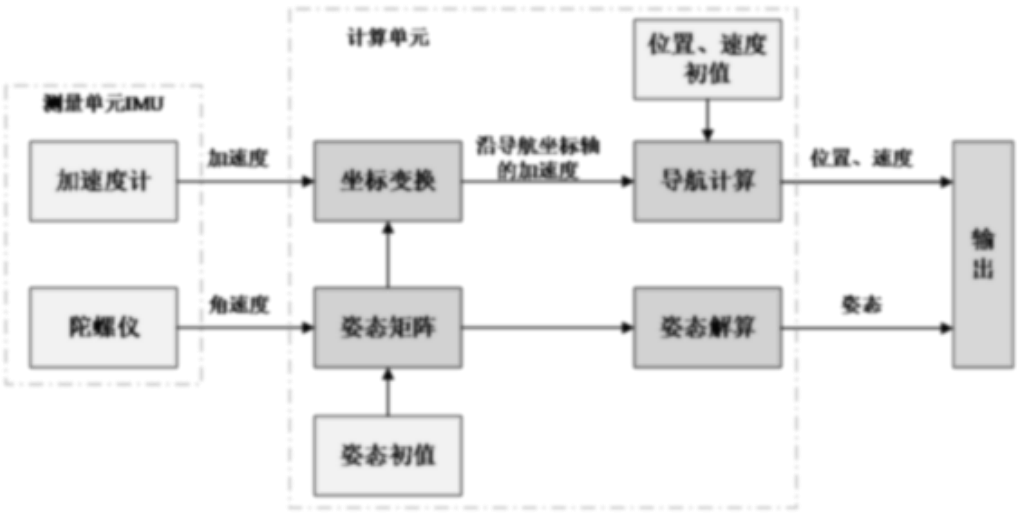


图4.2 惯性导航系统位姿解算流程

在实际系统中，为了降低陀螺仪和加速度计的输出噪声对系统解算精度的影响，同时能够完全利用输出信息，陀螺仪和加速度计的输出全部采用增量形式[64]，即输出量分别为速度增量和角增量。尽管这在一定程度上降低了陀螺仪和加速度计的零偏误差的影响，但其它类型的系统误差，例如陀螺仪和加速度计的随机游走误差、比例因子误差、温度误差等，仍是不可避免的。

由于 INS 完全不依赖于外界信息，其位姿解算的结果都是相对于初始化之后的位置、速度、姿态的初值来说的，因此随着运动时间的增加，误差会不断地在积分过程中累积，造成导航的精度逐渐降低，甚至呈指数级下降。因此为了保证无人机能够

完成长时间的导航任务需求，需要定期对其进行校正，下面就基于卡尔曼的多源信息融合技术进行介绍。

4.2 基于卡尔曼的多源信息融合技术

由于单一传感器的不足，使用多源信息融合来提升无人机导航系统的稳定性和精度是有必要的。卡尔曼滤波是在测量方差已知的情况下，从一系列测量噪声中估计系统优化状态的方法[66]，其具有轻量级、快速简单的优势。本小节将介绍本系统所使用的九维状态卡尔曼滤波器，并对其中的部分细节进行介绍。

4.2.1 九维状态卡尔曼滤波器

假设一个离散线性动态系统模型如下：

$$\begin{cases} x_k = Fx_{k-1} + Bu_k + \xi_k \\ z_k = Hx_k + \delta_k \end{cases} \tag{4-1}$$

其中状态向量 x_k 描述了 k 时刻的系统状态，u_k 表示 k 时刻的系统输入，z_k 表示 k 时刻观测到的系统状态，系统的状态转移矩阵、控制增益矩阵和状态观测矩阵分别由 F 、 B 、 H 表示，ξ_k 和 δ_k 分别表示系统的过程噪声和测量噪声，两种噪声均为高斯白噪声，其对应的协方差矩阵分别为 Q 和 R 。

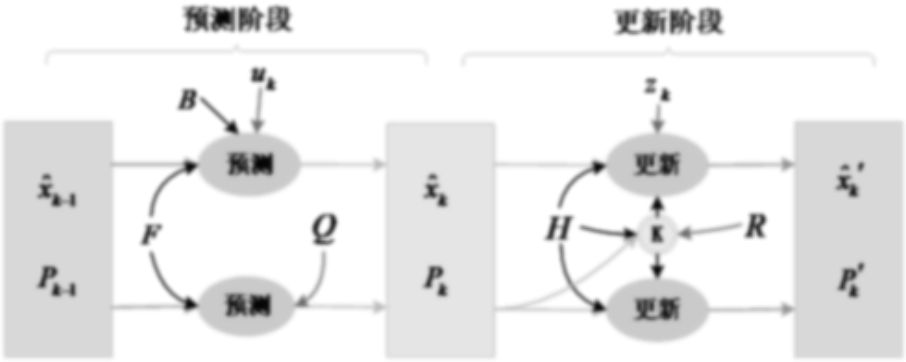


图4.3 卡尔曼滤波过程图

卡尔曼滤波过程分为预测和更新两个阶段，其预测过程如下：

$$\hat{x}_k' = F\hat{x}_{k-1} + Bu_k \tag{4-2}$$

$$\hat{P}_k' = F\hat{P}_{k-1}F^T + Q \tag{4-3}$$

即首先根据前一时刻系统估计状态 $\hat{x}_{k-1}$ 预测当前时刻状态 $\hat{x}'_k$，随后更新当前时刻状态向量 $\hat{x}'_k$ 对应的协方差矩阵 $\hat{P}'_k$。

接下来是校正部分，这里需要通过卡尔曼增益 K_k 来校正当前时刻系统估计状态 $\hat{x}_k$，并更新状态向量的协方差矩阵 $\hat{P}_k$：

$$\begin{cases} K_k = \hat{P}'_k H^T (H\hat{P}'_k H^T + R)^{-1} \\ \hat{x}_k = \hat{x}'_k + K_k(z_k - H\hat{x}'_k) \\ \hat{P}_k = (I - K_k H)\hat{P}'_k \end{cases} \tag{4-4}$$

接下来对本研究的无人机导航系统建立起卡尔曼滤波相应的系统动态模型。在 IMU 与视觉信息融合的过程中，由于 IMU 的频率更高更稳定，因此使用 IMU 数据作为轨迹增量的预测。对于无人机的位姿估计，选择位置向量 $\mathbf{P}=[x,y,z]$，速度向量 $\mathbf{v}=[v_x,v_y,v_z]$，姿态角向量 $\mathbf{\Theta}=[\varphi,\theta,\psi]$ 作为状态量 $[\mathbf{P},\mathbf{v},\mathbf{\Theta}]^T$，对于从时刻 *k*-1 到时刻 *k*，时间变化量为 Δt，则速度和位置的预测值有

$$\begin{cases} \hat{\mathbf{v}}'_k = \hat{\mathbf{v}}'_{k-1} + \hat{\mathbf{a}}_k \Delta t + \xi_k^v \\ \hat{\mathbf{P}}'_k = \hat{\mathbf{P}}_{k-1} + \hat{\mathbf{v}}_{k-1}\Delta t + \frac{1}{2}\mathbf{a}_k \Delta t^2 + \xi_k^{\mathbf{P}} \end{cases} \tag{4-5}$$

其中 $\hat{\mathbf{a}}_k$ 表示 *k* 时刻的 INS 三轴加速度输入值，ξ_k^v 和 ξ_k^P 分别表示时刻 *k* 的速度噪声和位置噪声，$\hat{\mathbf{v}}'_k$ 和 $\hat{\mathbf{P}}'_k$ 分别表示 *k* 时刻的速度预测值和位置预测值，$\hat{\mathbf{v}}_{k-1}$ 和 $\hat{\mathbf{P}}_{k-1}$ 分别表示 *k*-1 时刻的速度和位置的估计值。同理，对于姿态角有

$$\hat{\mathbf{\Theta}}'_k = \hat{\mathbf{\Theta}}_{k-1} + \boldsymbol{\omega}_k \Delta t + \xi_k^{\Theta} \tag{4-6}$$

其中 $\boldsymbol{\omega}_k$ 表示 *k* 时刻 INS 的三轴角速度输入值，ξ_k^{Θ} 表示 *k* 时刻的三轴角速度噪声，$\hat{\mathbf{\Theta}}_{k-1}$ 表示 *k*-1 时刻的姿态角估计值，$\hat{\mathbf{\Theta}}'_k$ 表示 *k* 时刻的姿态角预测值。

将式(4-5)和式(4-6)合并可以得到状态转移方程：

$$\begin{bmatrix} \mathbf{P}'_k \\ \mathbf{v}'_k \\ \mathbf{\Theta}'_k \end{bmatrix} = \begin{bmatrix} I_3 & \Delta t \cdot I_3 & O_3 \\ O_3 & I_3 & O_3 \\ O_3 & O_3 & I_3 \end{bmatrix} \begin{bmatrix} \mathbf{P}_{k-1} \\ \mathbf{v}_{k-1} \\ \mathbf{\Theta}_{k-1} \end{bmatrix} + \begin{bmatrix} 0.5\cdot\Delta t^2 \cdot I_3 & O_3 \\ \Delta t \cdot I_3 & O_3 \\ O_3 & \Delta t \cdot I_3 \end{bmatrix} \begin{bmatrix} \mathbf{a}_k \\ \boldsymbol{\omega}_k \end{bmatrix} + \begin{bmatrix} \xi_k^{\mathbf{P}} \\ \xi_k^{\mathbf{v}} \\ \xi_k^{\mathbf{\Theta}} \end{bmatrix} \tag{4-7}$$

观测向量则由基于异源影像匹配的位姿估计模块得到，如下：

$$\mathbf{z}_k = \begin{bmatrix} x_k & y_k & \tilde{z}_k & \phi_k & \theta_k & \psi_k \end{bmatrix}^{\mathrm{T}} \tag{4-8}$$

4.2.2 卡尔曼滤波器的数据同步

在进行多传感器融合的过程中，传感器的工作频率可能各不相同。在本文中所使用的 IMU 的工作频率为 100Hz，视觉模块的频率为 25Hz，但由于环境等因素影响，其也不是恒定不变的。尤其是视觉模块，图像匹配算法所需的时间与采集到的图像内容有较大关系，例如部分特征点密集的区域在匹配时需要在更大的搜索空间匹配对应点对，相比于稀疏特征区域需要消耗更多的时间。

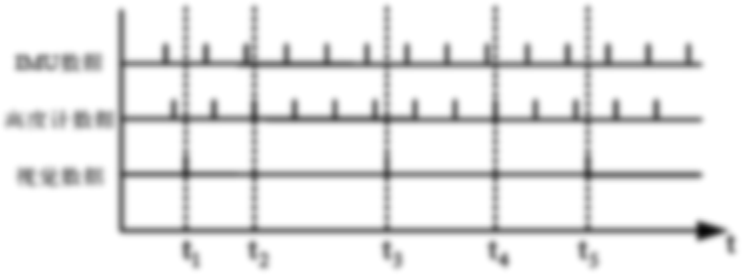


图4.4 数据同步示意图

如图 4.4 所示，在实际工作过程中，各传感器并不会总是保持在相同间隔时间进行数据输出，一方面其输出频率与预期频率可能有波动，另一方面也可能出现诸如数据丢失的情况，因此需要根据各传感器的时间戳来确定数据的采集时间，为了保证各传感器进行数据融合时的信息同步，需要在相应进行数据融合的位置对 IMU 和高度计的数据进行插值处理。

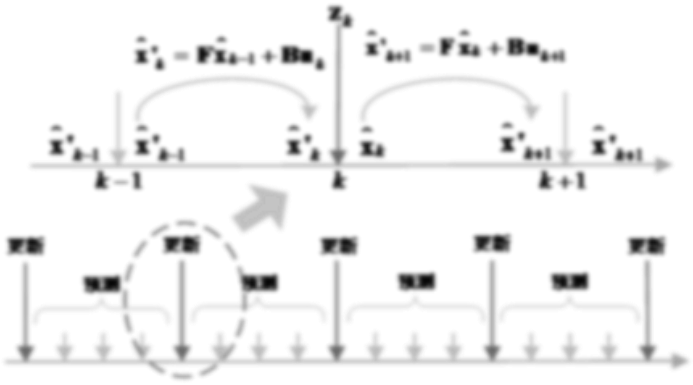


图4.5 卡尔曼预测和更新过程示意图

此外，由于 IMU 数据的更新频率约是视觉模块的更新频率四倍以上，为了满足

无人机导航的需求，也为了充分利用 IMU 的数据信息，将整个导航系统的位姿估计信息输出频率保持与 IMU 的频率一致，在使用卡尔曼滤波时，每做三次预测然后使用视觉信息做一次更新，如图 4.5 所示。

4.2.3 视觉位姿估计的航迹推算

在融合导航方案中，数据融合模块需要对视觉模块解算到的无人机位姿数据和 INS 输出的无人机位姿数据进行融合，然而视觉位姿解算模块中的图像匹配算法需要一定的时间（通常在数百毫秒级），从而无法实时性计算出无人机的视觉位姿信息。也就是说当数据融合模块读取到一个视觉位姿数据时，这可能是无人机数个时刻前的位姿信息，其具有一定的滞后性，为此需要对视觉定位获取的位姿数据进行一定的处理。

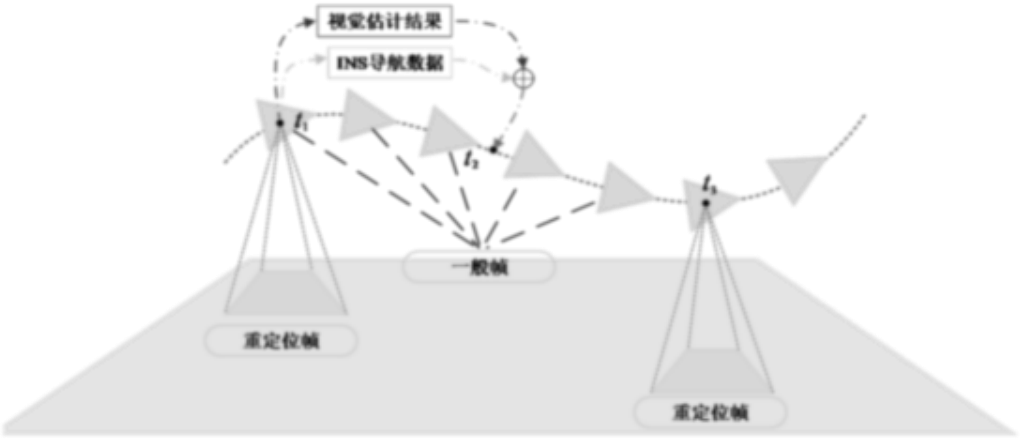


图4.6 视觉滞后补偿（航迹推算）示意图

如图 4.6 所示，视觉位姿解算模块在t_1时刻采集地面影像并启动位姿解算流程，在t_2时刻得到位姿解算的结果，显然这个位姿是t_1时刻的无人机位姿，为此利用 INS 短时间内具有高精度的特性，读取 IMU 单元在$t_1 \sim t_2$时刻的加速度和角速度数据从而计算系统两个时刻的位姿增量，据此推算在t_2时刻的视觉位姿信息，完成对视觉算法滞后性的补偿，本文将根据 IMU 数据来推算下一时刻的位姿的方法称之为航迹推算。

4.3 多源信息融合导航实验及数据分析

本小节将对本章提出的多源信息融合导航方法的效果进行验证，以比较相对于仅使用视觉位姿估计方法在导航精度、稳定性上的提升。本小节分别针对基于 SPG 的稀疏增强匹配算法和基于邻域共识的匹配算法设置了不同的区域进行实验。如图 4.7-图 4.10 所示展示了所选区域的不同影像。

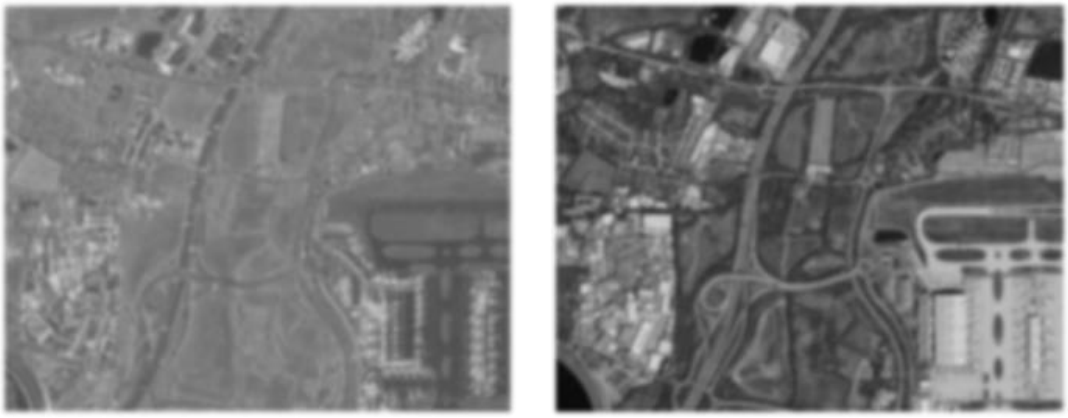


图4.7 西思罗国际机场局部（I 类异源）

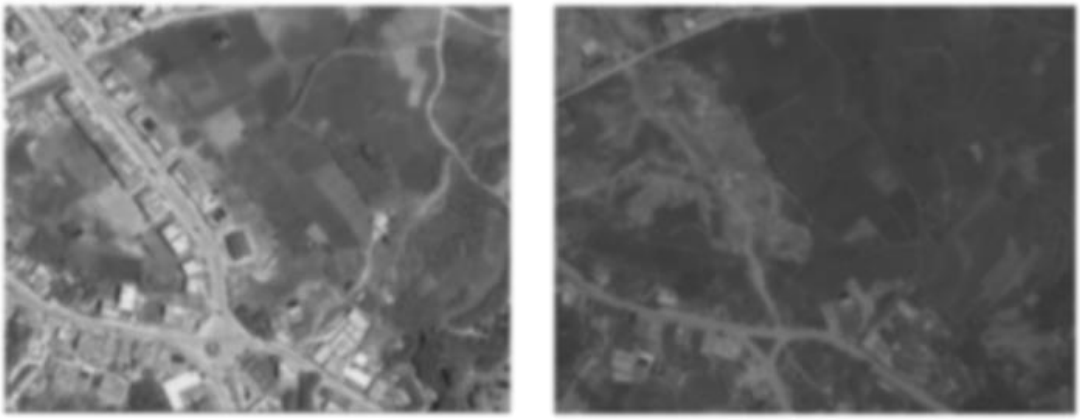


图4.8 某乡村小镇区域局部（I 类异源）

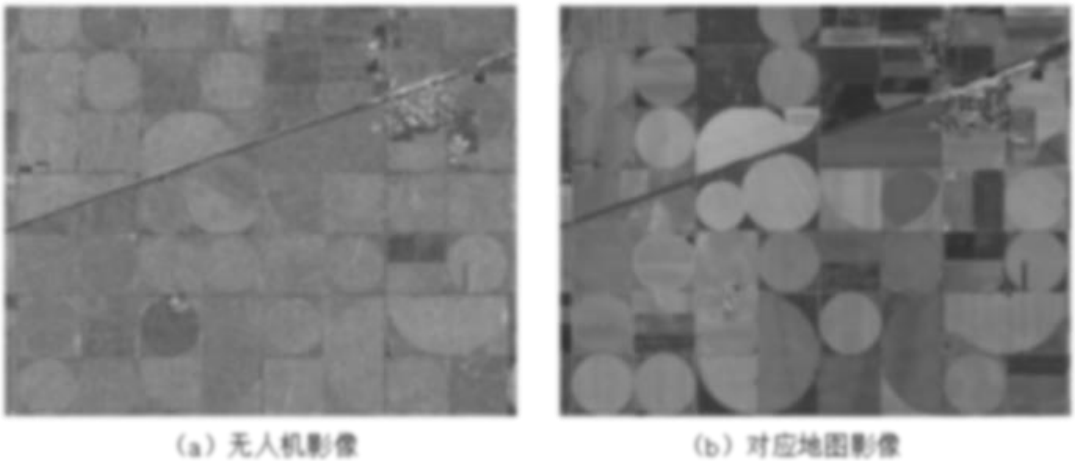


图4.9 科普兰德农田局部（II 类异源）

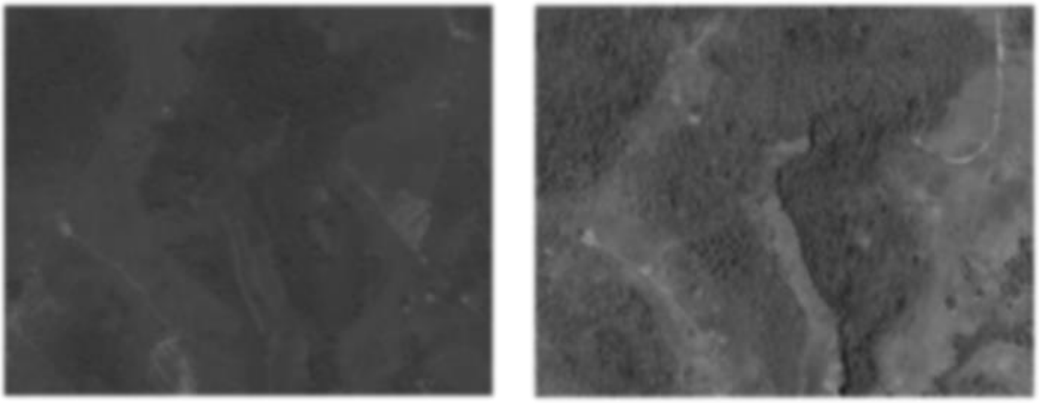


图4.10 某乡村外围区域局部（II 类异源）

图 4.7 是英国伦敦西思罗国际机场的地理影像和模拟无人机拍摄的 SAR 影像，二者主要是由于成像方式不同形成的异源影像；图 4.8 是某山区小镇部分在不同时期下的影像，由于人类活动以及光照的变化产生了较大差异。图 4.9 是科普兰德农田区域的地理影像和 SAR 影像；图 4.10 是某山区的村镇以外区域部分不同时期的影像，其主要是受时间或光照变化影响较大。其中图 4.7 和图 4.8 包含了大量的建筑物、道路或是点特征显著的优质区域，因此属于 I 类异源影像，而图 4.9 和图 4.10 由于优质点特征占比极低，属于 II 类异源影像。

4.3.1 I 类异源影像融合导航实验

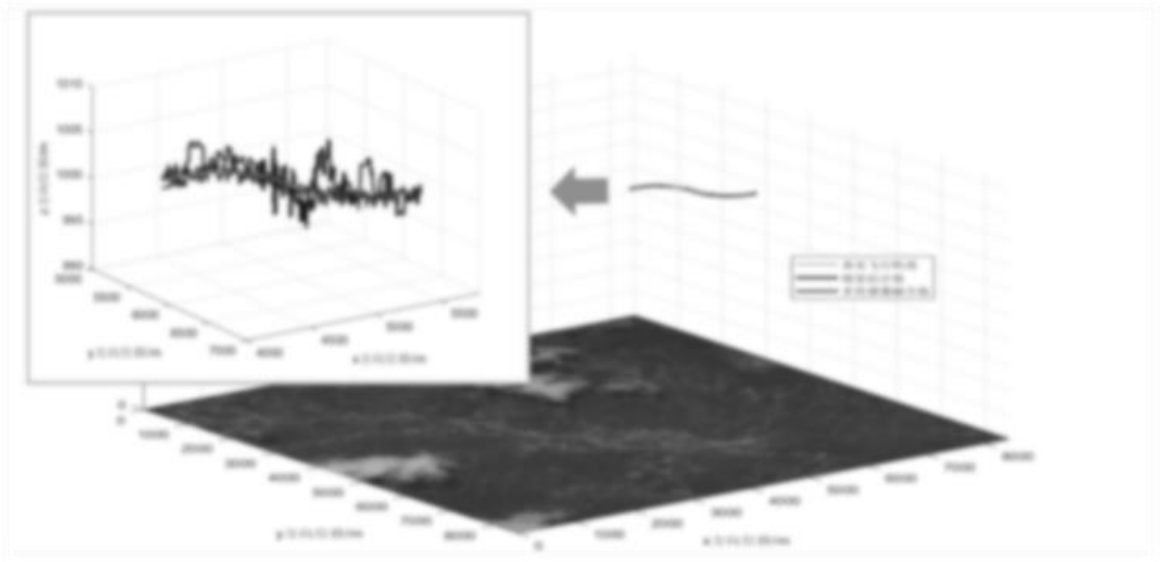


图4.11 I 类异源影像多源信息融合导航飞行轨迹示意图（路径一）

如图 4.11 所示展示了 I 类异源影像下多传感器融合导航的无人机飞行轨迹示意

图，本实验选取了 8558×8419 像素范围，地面分辨率大约为 1.0m×1.0m，整个路径中无人机飞行时长为 100s，无人机飞行高度为 1000m，无人机飞行起点坐标为（4360，5288），初始速度为 x 方向 25m/s，重定位帧间隔 N_h 设置为 25。

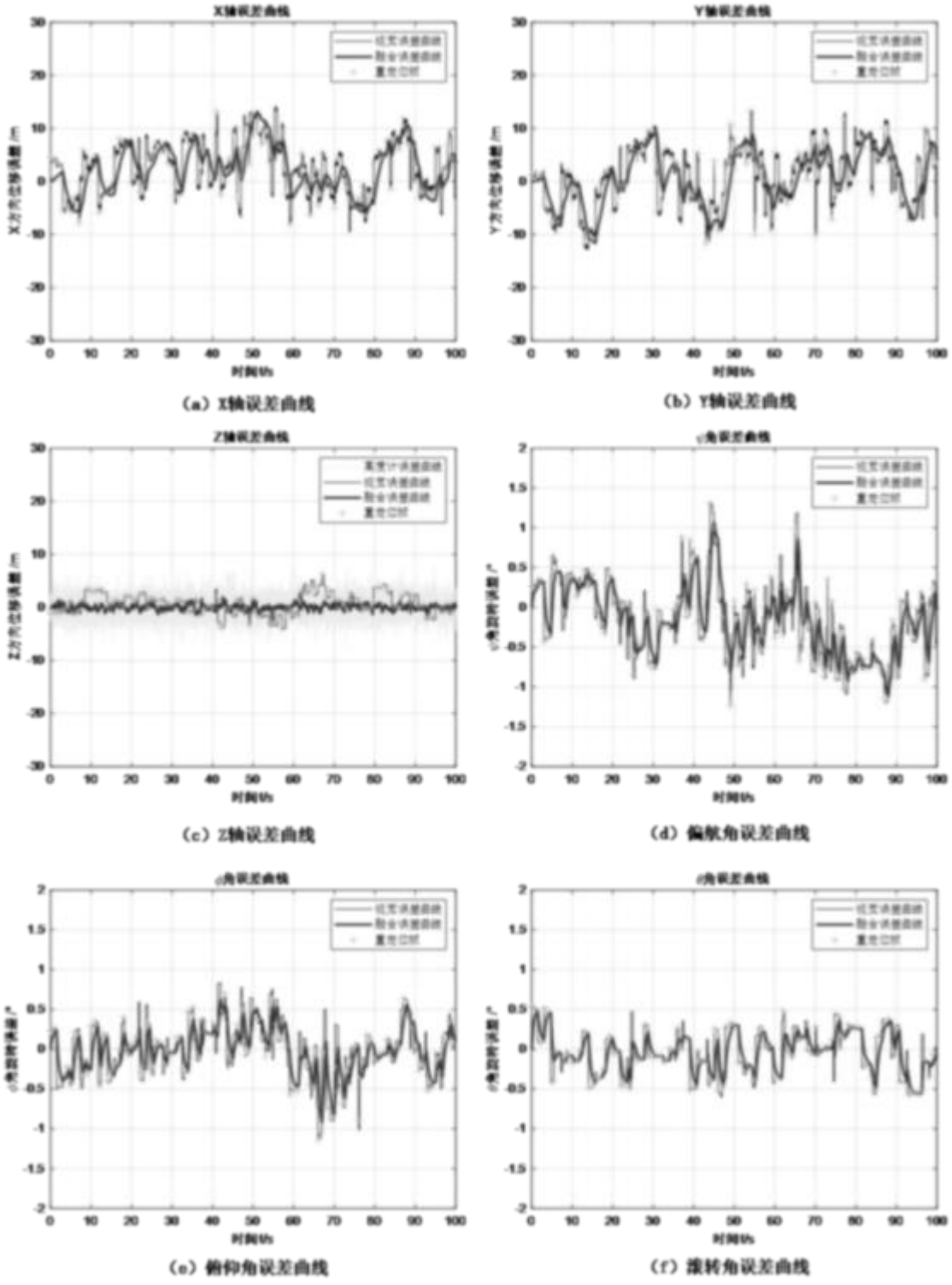


图4.12 I 类异源影像多源信息融合导航六轴误差曲线（路径一）

该区域在视觉定位上是基于 SPG 的 I 类异源影像稀疏增强匹配算法进行的，视觉位姿估计和融合结果误差曲线如图 4.12 所示，其中蓝色曲线表示使用视觉方法得到的位姿估计结果误差曲线，包含图像匹配和光流跟踪以及匹配失败下的航迹推算部分；红色曲线表示多源信息融合估计的位姿结果，包含视觉、IMU 和高度计（黄色曲线）；绿色点表示重定位帧得到的图像匹配结果，重定位发生在上一帧图像匹配失败、位姿解算结果误差过大、帧数达到重定位帧设定值等情况。

根据图 4.12 可以看出，在该异源影像区域的飞行路径下，基于图像匹配解算的位姿误差（绿色点）基本分布在 0 轴上下，其中 X 轴和 Y 轴曲线上部分误差最大达到了±15m，但大部分 90%以上的结果均在±10m 之间，而 Z 轴误差曲线视觉基本在±5m 之间波动，偏航角和俯仰角的误差大多在±1.5°内，而滚转角误差在±1.0°。基于光流跟踪的误差大致在±1m 范围，且围绕在图像匹配后的误差附近小幅波动，因此从总体来看，基于图像匹配解算的无人机位姿误差决定了整个视觉误差曲线的误差范围。

从多源信息融合结果来看，融合的 X 轴、Y 轴，以及俯仰角、偏航角和滚转角的误差最大最小值基本与视觉相当，但波动性较视觉结果有了一定的改善，相较于视觉结果的骤然变化，融合结果更加平滑。此外值得说明的是，由于高度计的精度通常要优于其他传感器得到的结果，且其误差符合典型的高斯分布，因此在融合过程中其置信度较高，尽管视觉有一定波动，但高度的融合结果误差要小的多。实验的精度统计结果将在后文表格进行展示。

针对不同数据源的数字影像地图进行了另一组实验，如图 4.13 所示。

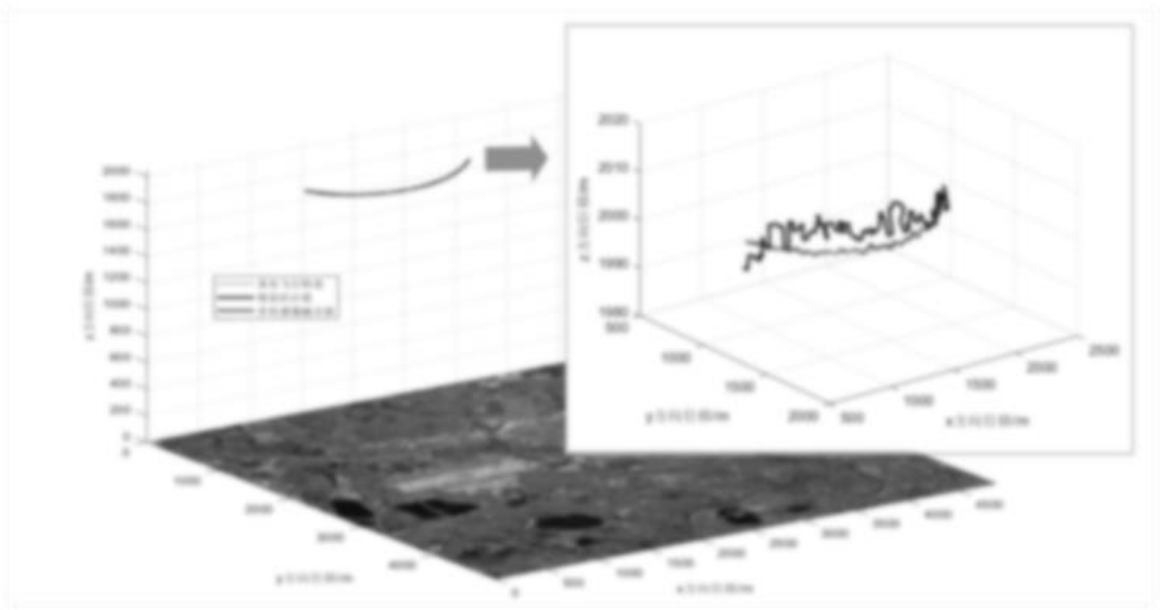


图4.13 I 类异源影像多源信息融合导航飞行轨迹示意图（路径二）

如图 4.13 所示展示了 I 类异源影像下另一幅多传感器融合导航的无人机飞行轨迹示意图，本实验选取了 4858×4859 像素范围，地面分辨率大约为 3.0m×3.0m，整个路径中无人机飞行时长为 60s，无人机飞行高度为 2000m，无人机飞行起点坐标为（800，1050），初始速度为 x 轴方向 10m/s，y 轴方向 30m/s，重定位帧间隔 $N_h = 25$ 。

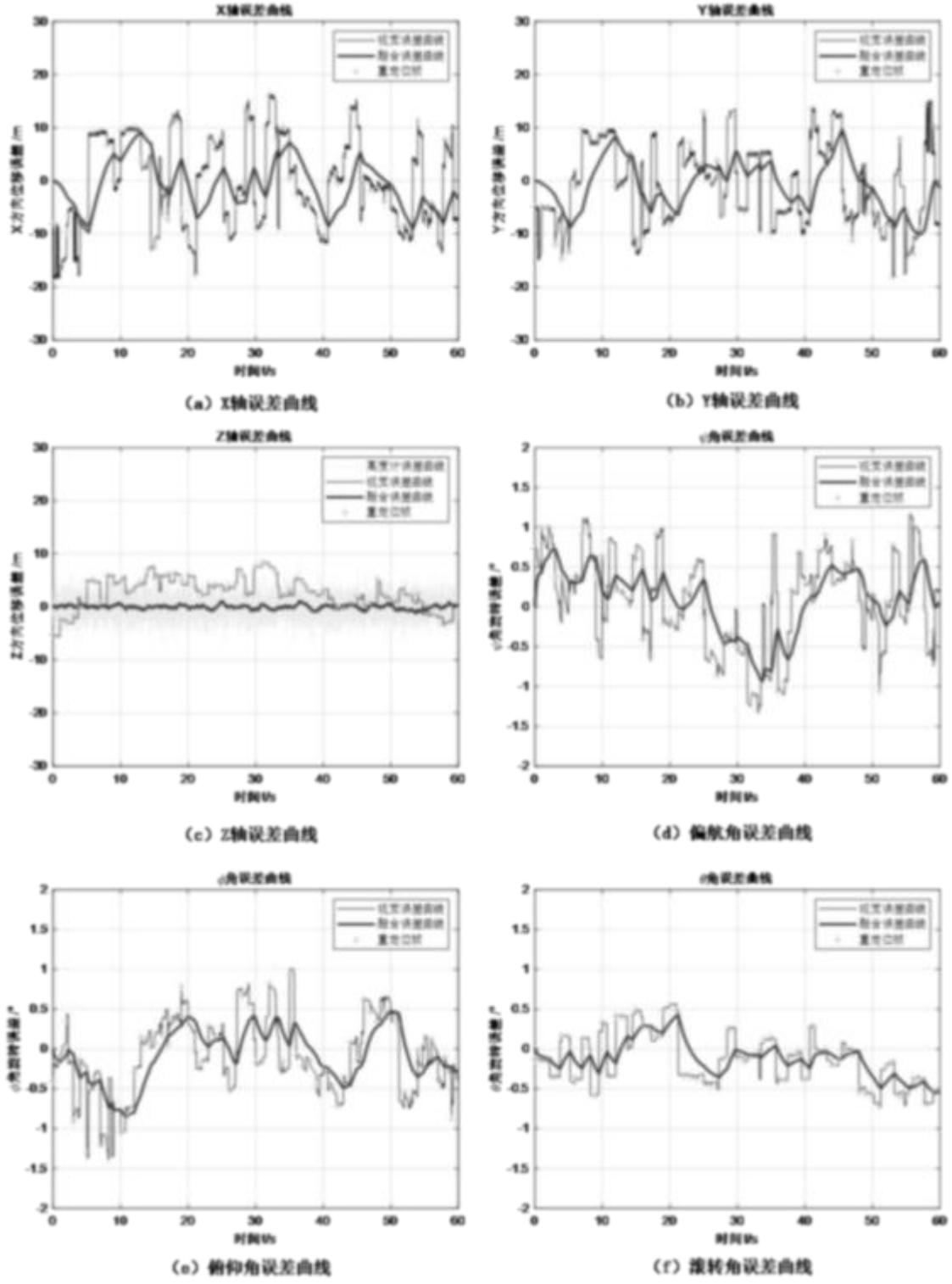


图4.14 I 类异源影像多源信息融合导航六轴误差曲线（路径二）

该区域在视觉定位上同样是基于 SPG 的 I 类异源影像稀疏增强匹配算法进行的，可以看出相比于路径一，在该区域视觉定位误差更小且更加稳定（相对于 2000m 高空来说），这主要是因为相比于路径一的乡镇区域，机场区域附近的建筑物分布更广更均匀，有效特征能够在连续帧里被稳定提取和处理。在该路径下，视觉位姿误差 X 轴、Y 轴和 Z 轴均在±15m 范围内，偏航角、俯仰角和滚转角均在±1.5°范围内，角度误差相比于路径一并未有太多变化。

相似的，多源信息融合结果相比于单纯使用视觉位姿估计结果有了较大改善，X、Y 和 Z 三轴融合后误差均在±10m 范围内波动，偏航角、俯仰角和滚转角三轴融合后误差大致在±0.5°范围。可以看出融合得到位姿解算结果相比于视觉结果更加平滑且误差更小。这是因为 IMU 数据测量的是无人机加速度，因此在融合过程中能够对视觉定位结果形成约束，使用卡尔曼融合时，尽管视觉误差允许情况下会给予视觉定位更高的可信度，但由于加速度的约束使融合结果无法产生突变。

值得说明的是，在路径一的导航实验中，部分区域由于提取到的特征点过少因此位姿解算产生了较大误差，但在可信修正的过程中丢弃了这些异常点。在上述实验基础上又模拟飞行了多组路径，表 4.1 和表 4.2 分别统计了各组路径下的视觉估计位姿和多源信息融合位姿的精度。

表4.1 I 类异源影像视觉估计位姿精度

路径组	X 轴方向/m		Y 轴方向/m		Z 轴方向/m	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	6.759	7.692	6.848	7.697	2.618	2.033
2	7.352	6.394	5.098	6.082	2.260	2.972
3	8.344	8.547	6.560	7.604	3.313	3.955
4	9.811	12.254	8.724	9.680	2.966	3.440
均值	8.067	8.722	6.808	7.766	2.789	3.100
路径组	**偏航角 ψ /°**		**俯仰角 φ /°**		**滚转角 θ /°**	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	0.416	0.511	0.279	0.354	0.237	0.288
2	0.382	0.464	0.405	0.491	0.257	0.313
3	0.496	0.592	0.415	0.473	0.296	0.351
4	0.564	0.682	0.441	0.516	0.323	0.373
均值	0.465	0.562	0.385	0.459	0.278	0.331

表4.2 I 类异源影像多源信息融合位姿精度

路径组	X 轴方向/m		Y 轴方向/m		Z 轴方向/m	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	3.926	4.944	4.163	5.043	0.400	0.503
2	4.388	5.249	4.150	5.099	0.399	0.502
3	5.914	5.655	4.870	4.648	0.297	0.384
4	7.779	9.000	5.414	6.659	0.415	0.537
均值	5.502	6.212	4.649	5.362	0.378	0.482
路径组	**偏航角 ψ /°**		**俯仰角 φ /°**		**滚转角 θ /°**	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	0.354	0.435	0.213	0.270	0.186	0.229
2	0.311	0.385	0.330	0.395	0.202	0.244
3	0.338	0.398	0.280	0.341	0.201	0.246
4	0.531	0.625	0.389	0.463	0.254	0.304
均值	0.384	0.461	0.303	0.367	0.211	0.256

在经过多组实验取平均值后，根据表中结果，多源信息融合结果相比于仅使用视觉位姿估计方案平均绝对误差 MAE 在 X 轴、Y 轴和 Z 轴方向上分别减少了 31.8%、31.7%和 86.5%，在偏航角、俯仰角和滚转角上分别减少了 17.4%、21.3%和 24.3%。而均方根误差在 X 轴、Y 轴和 Z 轴方向上分别减少了 28.8%、31.0%和 84.5%，在偏航角、俯仰角和滚转角上分别减少了 18.1%、19.9%和 22.8%。

表4.3 I 类异源影像视觉定位时间结果

	重定位帧耗时/s	平均定位耗时/s
4 组路径均值	0.103	0.007

对于 25 帧/s 的无人机视频来说，要达到实时性要求的理论帧定位耗时为 0.04s，尽管重定位帧耗时接近 0.1s，但光流帧定位速度远远高于实时性要求，对于 25 帧仅需定位一次（重定位帧间隔 $N_h = 25$ ）的视觉位姿估计系统能够达到实时性要求，且最低允许的重定位帧数量为每秒 9 帧。

关于多源信息融合对于视觉辅助导航鲁棒性和速度的提升主要表现在以下方面，一方面图像匹配过程中需要知道当前无人机拍摄影像在先验数字影像地图中的位置，通过上一时刻的无人机位姿和 IMU 数据进行航迹推算得到当前时刻无人机位姿，再

进行反变换即可得到位于数字影像地图的中心位置，这样能够显著减少图像匹配过程中搜索数据库的大小，也能减少非映射区域产生错误匹配的数量；另一方面由于 IMU 短时高精度的特性，也可以根据当前无人机推算位姿对视觉实际解算位姿进行约束，由于图像匹配并不一定能够每一帧都匹配成功，因此这种约束也能够减少视觉异常值出现的概率。

可见对于基于多源信息融合的导航方法能够显著提升仅依赖基于 SPG 稀疏增强匹配的视觉定位方案的精度和稳定性，满足实验预期：既达到所期望的长时间鲁棒性导航定位的目的，也在整个过程中优化了视觉位姿估计结果，提升了导航定位的精度。

4.3.2 II 类异源影像融合导航实验

如图 4.15 所示展示了 II 类异源影像下多传感器融合导航的无人机飞行轨迹示意图，本实验像素范围 8558×8419，地面分辨率约 1.0m×1.0m，整个路径中无人机飞行时长为 100s，无人机飞行高度 1000m，无人机飞行起点坐标为（920，776），初始速度 x 方向 30m/s，重定位帧间隔 N_h 设置为 25。

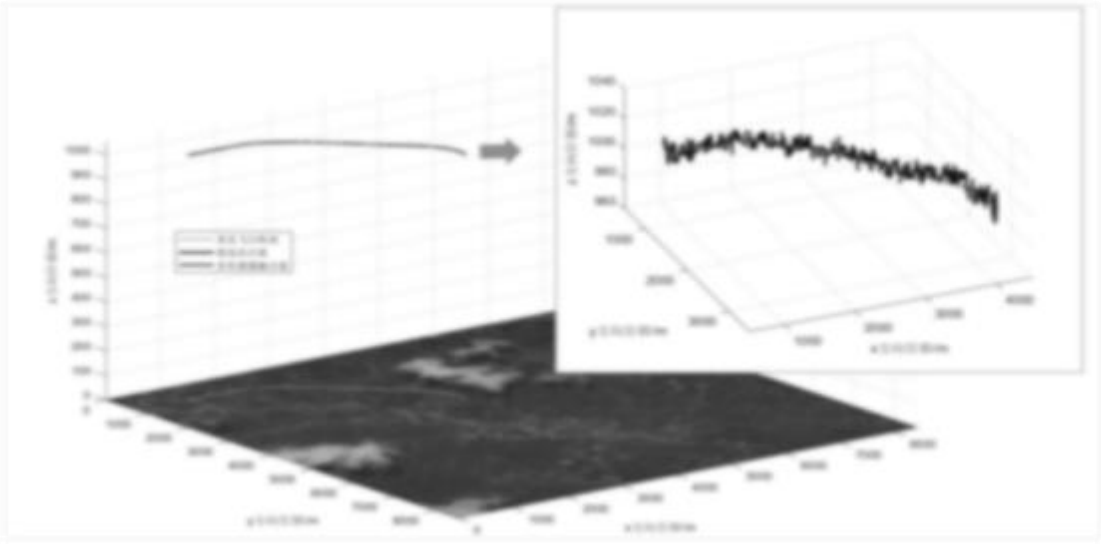


图4.15 II 类异源影像多源信息融合导航飞行轨迹示意图（路径一）

该区域在视觉定位上是基于邻域共识的 II 类异源影像匹配算法进行的，视觉位姿估计和融合结果误差曲线如图 4.16 所示，可以看出，多源信息融合结果要显著优于视觉结果。该实验中，视觉误差曲线在 X 轴、Y 轴方向上的位姿估计误差在 20m 左右，相似的，仅稀疏光流跟踪的视觉误差保持在图像匹配解算的位姿结果约 5m 范围内，可见图像匹配算法的精度和鲁棒性决定了视觉位姿估计的精度和鲁棒性。在 Z 轴上视觉估计误差保持在 10m 以内，而偏航角、俯仰角和滚转角更是维持在了±1°以内，保持了很好的结果。

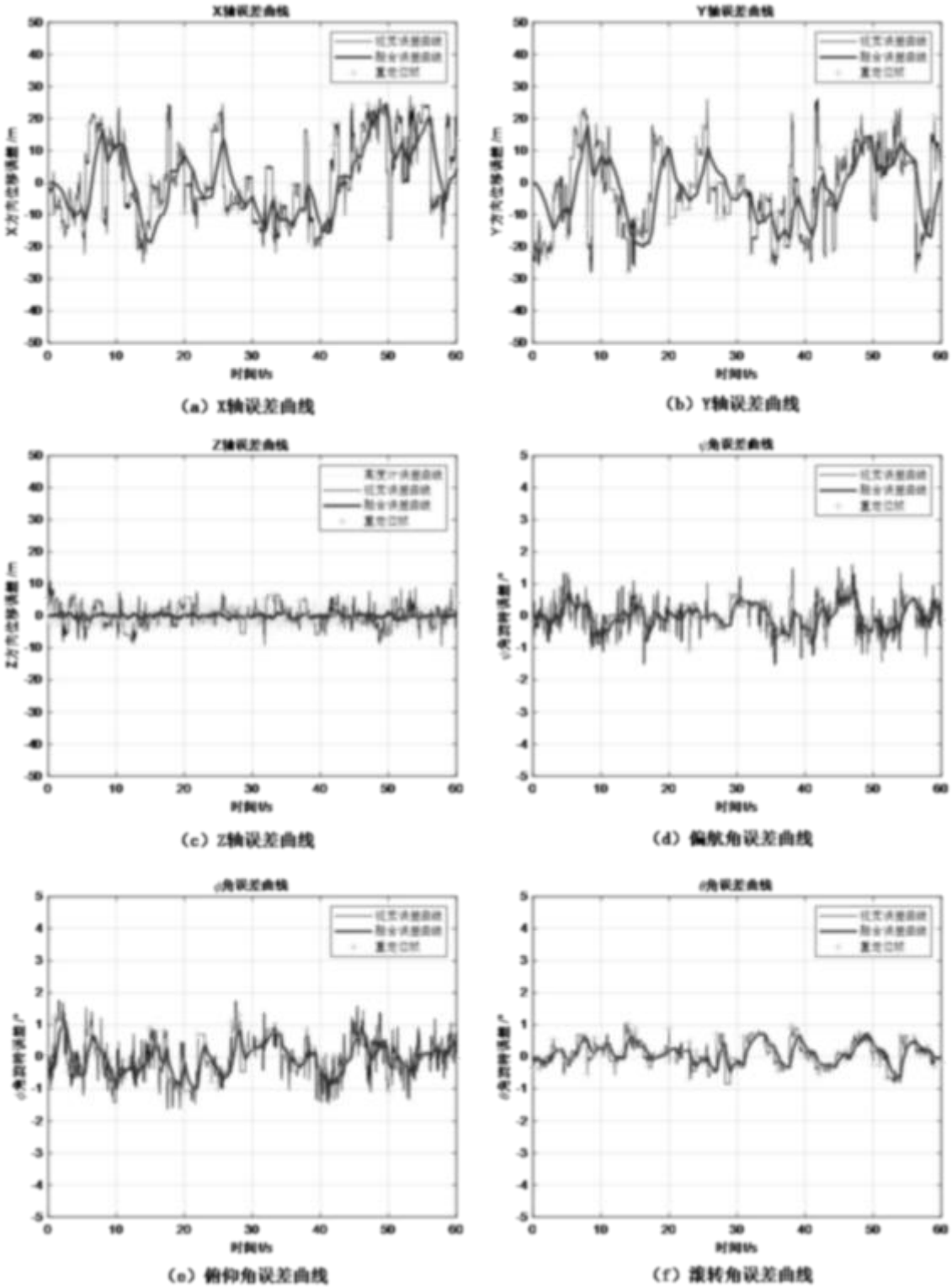


图4.16 II 类异源影像多源信息融合导航六轴误差曲线（路径一）

同样的，在上述实验基础上又模拟飞行了多组路径，表 4.4 和表 4.5 分别统计了各组路径下的视觉估计位姿和多源信息融合位姿的精度。

表4.4 II 类异源影像视觉位姿估计精度

路径组	X 轴方向/m		Y 轴方向/m		Z 轴方向/m	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	11.815	13.653	11.849	13.642	3.489	4.307
2	14.788	16.401	15.896	17.589	3.056	3.701
3	13.208	14.726	15.801	17.766	2.507	3.016
4	10.662	12.890	10.809	12.820	3.064	3.798
均值	12.618	14.418	13.589	15.454	3.029	3.706
路径组	**偏航角 ψ /°**		**俯仰角 φ /°**		**滚转角 θ /°**	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	0.799	0.871	0.756	0.838	0.729	0.876
2	0.766	0.826	0.682	0.716	0.601	0.755
3	0.739	0.838	0.716	0.779	0.635	0.710
4	0.427	0.534	0.488	0.603	0.365	0.446
均值	0.683	0.767	0.661	0.734	0.583	0.697

表4.5 II 类异源影像多源信息融合位姿精度

路径组	X 轴方向/m		Y 轴方向/m		Z 轴方向/m	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	8.483	10.511	8.756	10.339	0.399	0.503
2	9.229	10.287	10.931	10.076	0.441	0.557
3	8.794	10.491	11.729	13.794	0.431	0.568
4	8.303	9.781	7.677	9.155	0.400	0.503
均值	8.702	10.268	9.773	10.841	0.418	0.533
路径组	**偏航角 ψ /°**		**俯仰角 φ /°**		**滚转角 θ /°**	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	0.481	0.546	0.526	0.628	0.594	0.669
2	0.528	0.605	0.577	0.632	0.505	0.580
3	0.626	0.650	0.530	0.678	0.587	0.655
4	0.275	0.336	0.310	0.389	0.283	0.343
均值	0.478	0.534	0.486	0.582	0.492	0.562

根据表中均值结果，相比于仅使用视觉位姿估计多源信息融合方案在 X 轴、Y 轴和 Z 轴方向上平均绝对误差分别减少了 31.0%、28.1%、86.2%，均方根误差分别减少

了 30.1%、26.5%和 15.5%。在偏航角、俯仰角和滚转角上平均绝对误差分别减少了 28.8%、29.9%和 85.6%，均方根误差分别减少了 30.4%、20.7%和 19.4%。

表4.6 II 类异源影像视觉定位时间结果

	重定位帧耗时/s	平均定位耗时/s
4 组路径均值	0.184	0.010

II 类异源影像视觉定位时间如表 4.6 所示，相比于实时性要求的 0.04s 每帧，平均定位耗时仅 0.01s 的视觉位姿方案显然可以满足实时性要求。综合实验结果可以得出结论：在基于 II 类异源影像匹配的无人机位姿估计技术基础上，通过 IMU 和高度计等数据进行多源信息融合导航能够进一步提升无人机位姿估计的精度，提升的平均精度约在 29%以上。视觉辅助的多源信息融合导航处理技术可以保障无人机长时间导航的稳定性，满足在 GNSS 拒止条件下的无人机导航需求，达到实验预期。

4.4 高低空无人机协同导航融合实验

本实验选取了乡村村镇以外的区域来进行飞行实验，高低空无人机保持同步按照预设轨迹进行飞行，起点坐标均为（1000，3000），路径总时长为 100s，高空高度 2000m，低空高度 1000m，初始速度为 x 方向 25m/s。

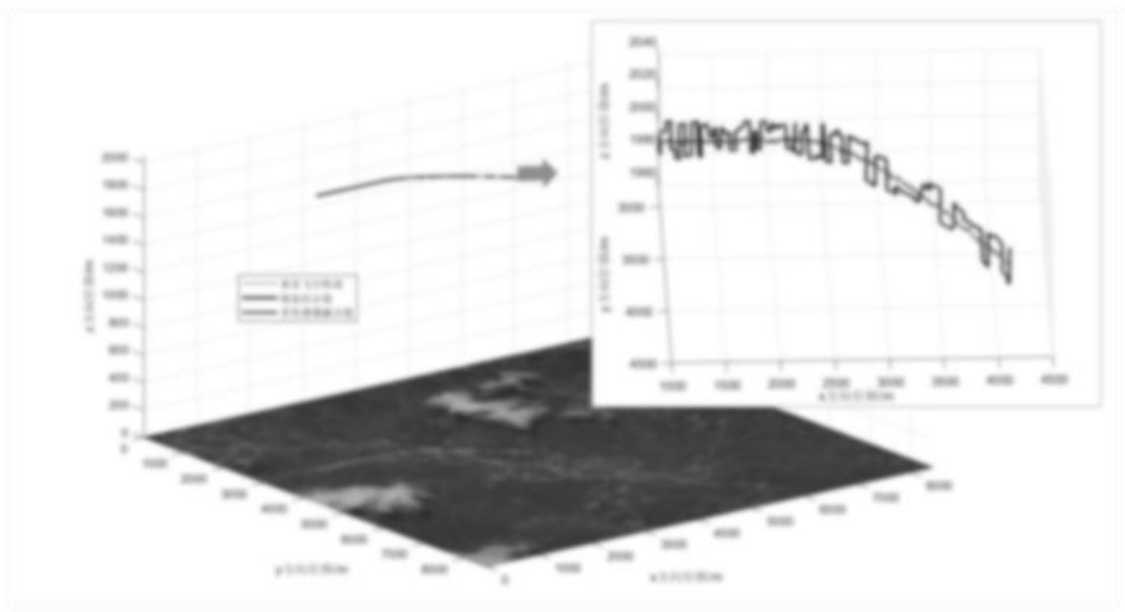
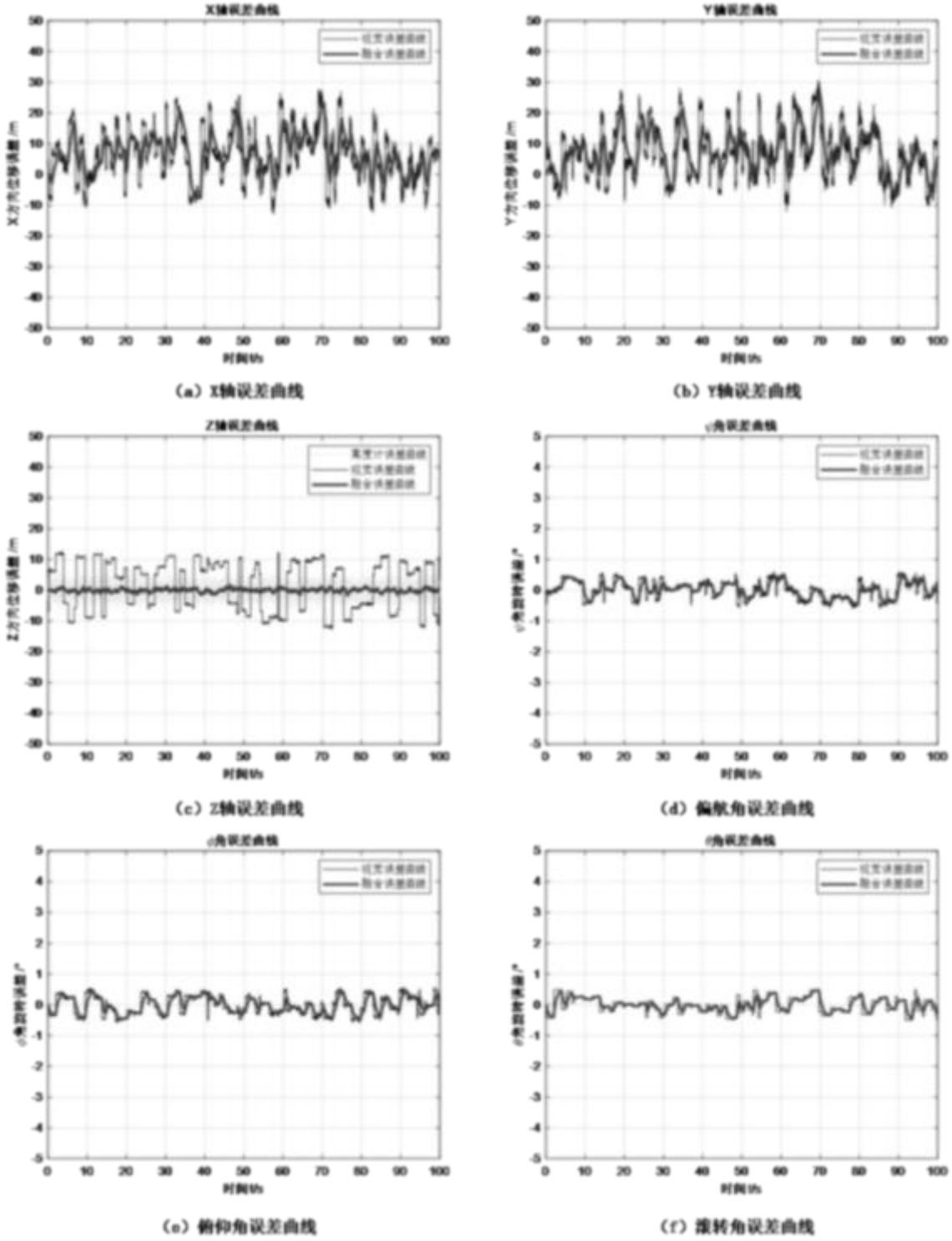


图4.17 高空无人机飞行轨迹示意图

其中高空无人机视觉位姿估计飞行轨迹整体效果如图 4.17 所示，可见高空视觉

曲线较为稳定，呈现振荡但振幅大致相当。误差曲线如图 4.18 所示，蓝色线为视觉估计值，在高空下由于无人机视野范围广，因此能够在视觉位姿估计时得到在真实值周围均匀波动的位姿值。红色线为多源信息融合值，相对于视觉位姿值融合值能够更加平滑。



（a）X轴误差曲线　（b）Y轴误差曲线

（c）Z轴误差曲线　（d）偏航角误差曲线

（e）俯仰角误差曲线　（f）滚转角误差曲线

图4.18　高空无人机视觉与融合导航六轴误差曲线

对于低空无人机，由于属于 II 类异源图像，因此同样采用本研究提出的基于邻域共识图像匹配的无人机位姿估计技术进行实验（基于 SPG-SE 导航定位失败），视觉位姿估计在无协同下路径轨迹图如图 4.19 所示，可以看出在整个飞行路径中，大部分区域都能较为稳定的导航，但在路径的中间和后半部分视觉位姿产生了极大抖动，因此也导致了融合位姿值剧烈变化。这主要是由于无人机飞行高度较低，视野区域有效特征在数字影像地图区域所占比例较小，因此低空下的异源差异导致了算法难以完成视觉位姿估计。

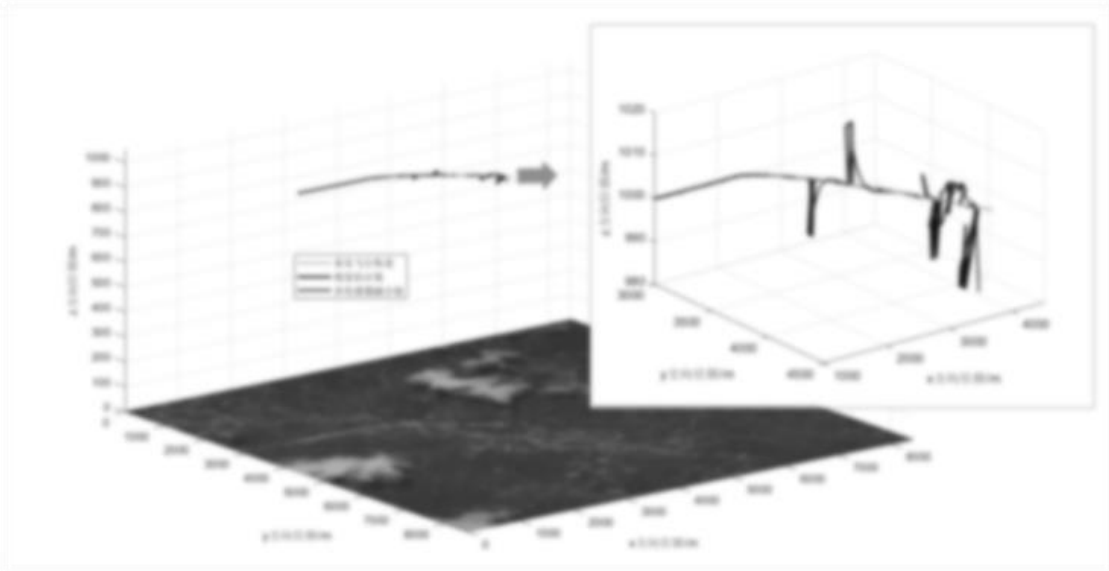


图4.19 低空无人机飞行轨迹示意图（无协同）

以位姿解算误差过大区域帧为例，如图 4.20 所示，低空无人机视野范围内仅包含大量林木区域，相比于数字影像地图，除光照天气变化外，地貌也有较大改变，两个区域相似度较低，因此邻域共识方法也难以寻找到正确有效的匹配。而在高空无人机视野下，虽然局部区域变化较大，但从整体来看，地形轮廓、周边建筑或是整体地貌特征可以找到较多的相似性。

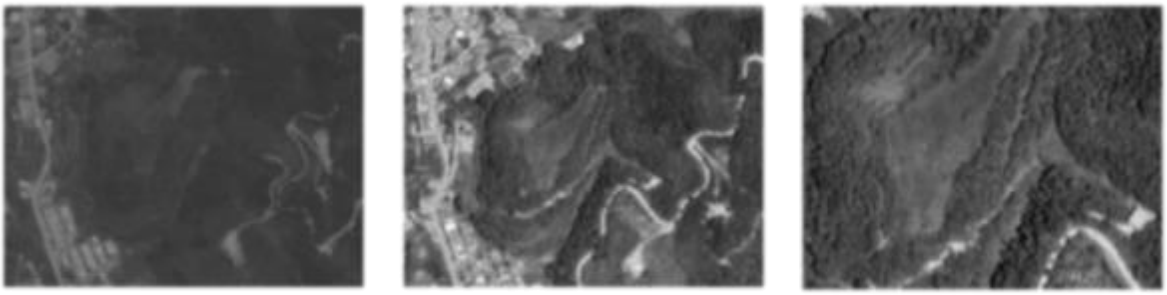


图4.20 低空无人机视觉位姿估计误差过大帧

如图 4.21 所示显示了在高低空协同导航下低空无人机的六轴位姿曲线，可以看出低空无人机视觉位姿估计相比于高空无人机误差稍小，这是因为低空能够获取更多地面细节，往往图像匹配精度更高，且由于高度较低，位姿估计的解算误差也较小。曲线图中紫色倒三角表示高低空位姿协同更新点，可见其协同更新位置基本仅限于低空无人机视觉位姿误差较大的情况下，这能够减少实际运行过程中的资源消耗。。

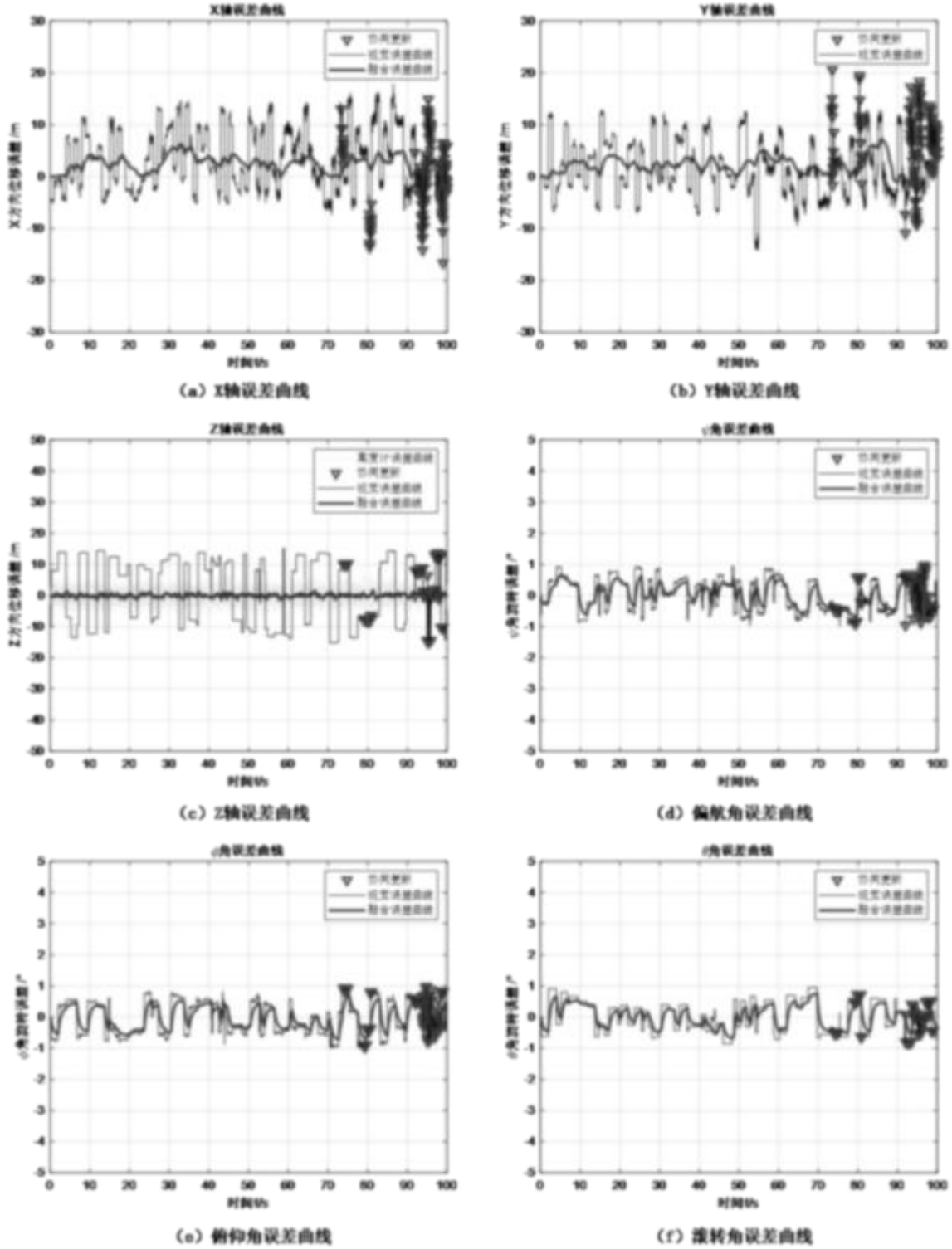


图4.21 低空无人机视觉与融合导航六轴误差曲线

实验过程中低空无人机的视觉精度统计如表 4.7 所示，由于视觉上图像匹配的失败，导致数据产生了部分异常值，因此均方根误差较大，其中 X 轴均方根误差约是平均绝对误差的两倍，Y 轴均方根误差约是平均绝对误差的三倍，而 Z 轴、偏航角、俯仰角和滚转角均达到了五倍。在高空无人机协同导航的支持下，低空视觉位姿上的异常值被消除，各轴的平均绝对误差均有了不同程度的改善，在均方根误差上则产生了显著改善，例如 Y 轴从 22.897m 降低到了 6.175m，大约降低了 73%。

表4.7 低空无人机视觉导航误差结果

		X/m	Y/m	Z/m	偏航角/°	俯仰角/°	滚转角/°
自主	MAE	6.409	7.529	7.704	0.704	0.702	0.676
	RMSE	11.624	22.897	10.313	0.818	0.757	0.785
协同	MAE	5.647	4.881	0.115	0.527	0.522	0.404
	RMSE	6.981	6.175	0.235	0.571	0.575	0.485

可见当低空无人机视觉位姿估计失败时，通过高空无人机数据能够修正低空无人机的位姿估计结果。也就是说在 GNSS 拒止情况下，相较于单独的无人机视觉导航方法，本研究提出的高低空无人机视觉协同导航方案能够满足无人机在差异极大的异源区域完成自主导航和定位。

4.5 本章小结

本章节主要研究了多源信息融合导航处理技术，借助于上一章的视觉位姿估计方案，通过视觉、高度计和 IMU 进行数据融合以获取精度更高且更加鲁棒的导航系统。

本章首先对 IMU 进行了简要介绍，随后针对本文的无人机系统，提出了一种基于九维状态卡尔曼的数据融合技术，并针对各传感器工作频率不同的问题以及视觉数据结果延时的问题给出了解决方案。

最后通过多组实验对 I 类异源影像和 II 类异源影像下的视觉辅助的融合导航方案以及高低空协同导航方案进行了验证，以检测各方案的合理性和精度、稳定性提升效果。实验证明，在两类异源影像区域，基于 SPG-SE 或邻域共识的视觉位姿解算方案融合导航精度相比于纯视觉位姿估计精度能够提高约 1/3，且在 I 类异源区域中基于 SPG-SE 的位姿估计方法显示出了更高的精度，而基于邻域共识的方案显示出了更好的鲁棒性，在几幅具有挑战性的 II 类地图上均表现较好。

此外，高低空协同导航方案针对复杂任务环境图像匹配失败情况导致的位姿解算结果有良好的改善。

第五章　软件设计与实现

本研究在过程中建立了三个分系统，分别是基于点特征的异源影像匹配系统，基于异源影像匹配与多源传感器融合的无人机导航定位系统，高低空无人机协同定位导航系统。为了针对上述方法进行演示处理，本研究设计完成了 GNSS 拒止下的无人机视觉辅助融合导航定位软件。

5.1 软件开发环境

本软件开发基于 PyCharm 2021（Community Edition）平台，借助 pyQt5、QGIS、OpenCV 等开源库完成和实现，本软件开发的具体环境如表 5.1 所示。

其中 pyQt 主要实现软件交互界面的搭建，包括按钮、下拉菜单等控件与槽函数的连接，信号和参数的传递，界面数据显示，弹出式交互窗口等。

QGIS 主要实现超大像素地理数字影像地图的绘制及显示，以及显示后的地图缩放、拖动等功能，飞行路径的实时绘制也借助于 QGIS 的图层实现。

OpenCV 是图像处理的核心开源库之一，借助其实现包括图像读取、图像缩放、图像显示、色彩空间转换、切片、滤波、部分特征提取与匹配、光流跟踪等操作。

表5.1　GNSS 拒止下的无人机视觉辅助融合导航定位软件验证环境

编号	名称/型号	硬件/软件配置	备注
1	CPU	AMD 5800X CPU @ 3.8GHz	-
2	主板	华硕 B550	-
3	显卡	NVIDIA RTX 2080Ti	-
4	内存	32G	-
5	硬盘	1T	-

5.2 功能模块介绍

本软件标题栏包含最小化、最大化及关闭按钮，下方为菜单选择栏，可进行各页面的切换，菜单选择栏包括首页、匹配算法页、融合定位页、协同定位页和结果分析页，各页面及其包含的功能模块结构如图 5.1 所示。

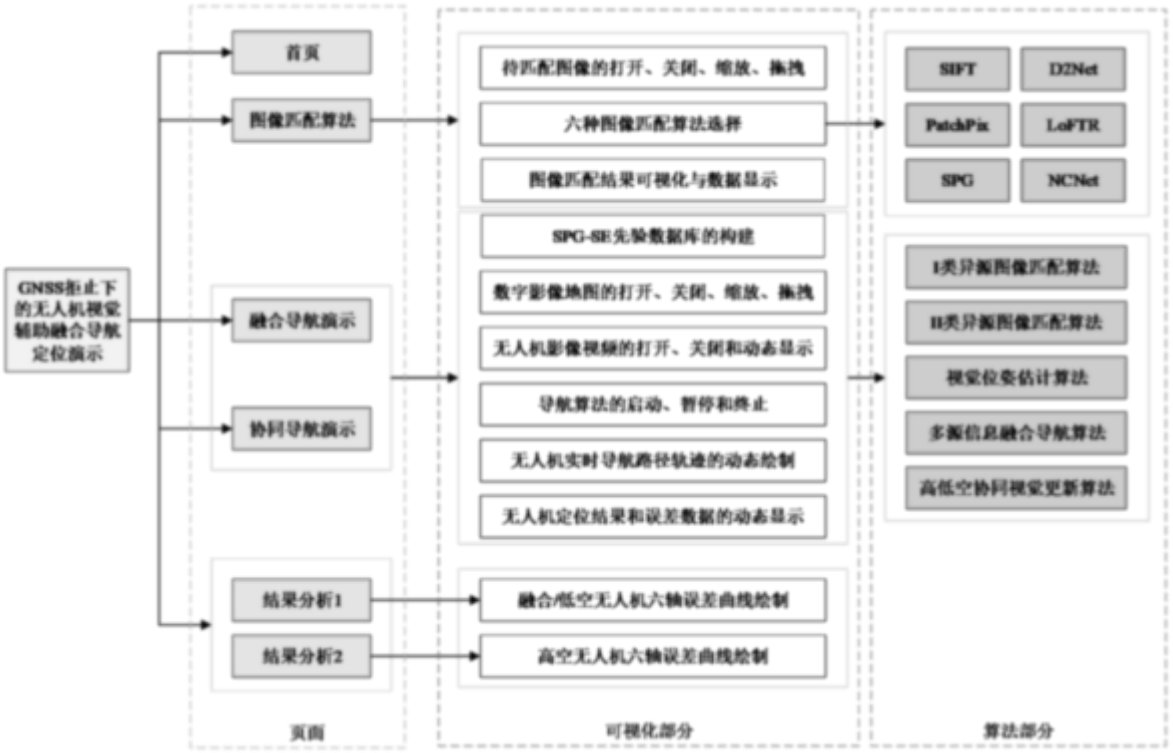


图5.1 GNSS 拒止下的视觉辅助融合导航定位演示软件结构图

5.3 软件界面展示及功能测试

其中首页如图 5.2 所示。

图5.2 首页

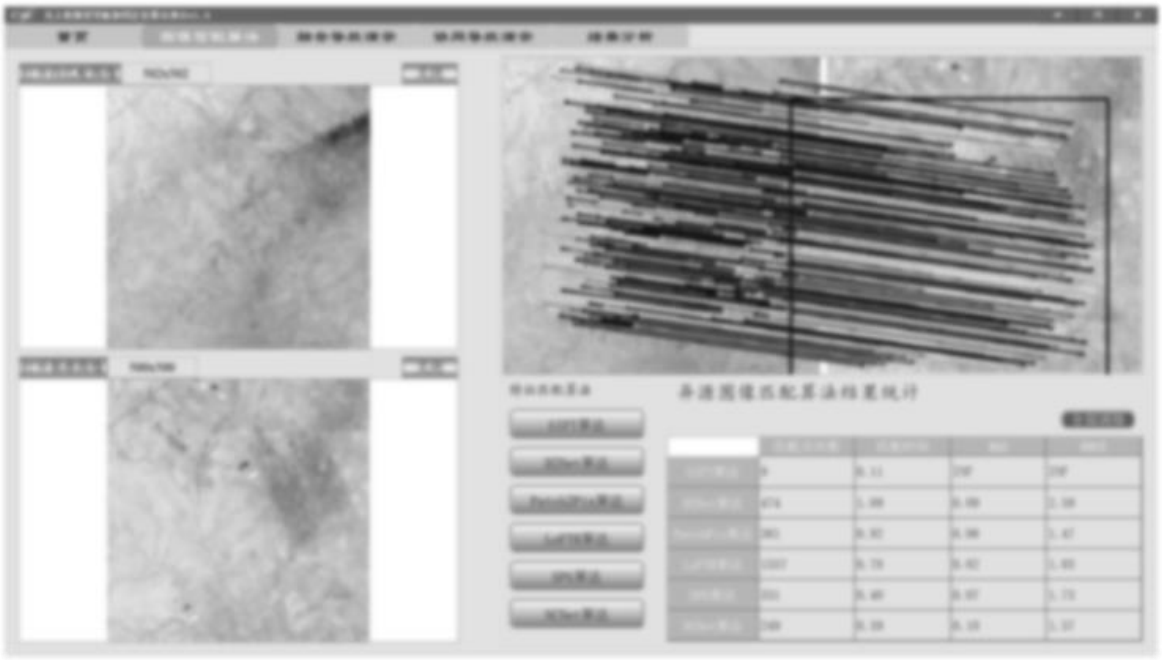


图5.3 匹配算法界面

匹配算法页面如图 5.3 所示，页面左侧用来加载和展示用于异源图像匹配的待匹配图像和基准图像，右侧上半部分用来展示匹配结果，下面部分用来显示对应算法的匹配结果数据，包括匹配点对数、匹配时间、平均绝对误差 MAE 和均方根误差 RMSE，页面集成的算法包括 SIFT、D2Net、Patch2Pix、LoFTR 和本文所使用的 SPG 算法、NCNet 算法。其中各图像展示区域均支持图像的缩放和拖拽。

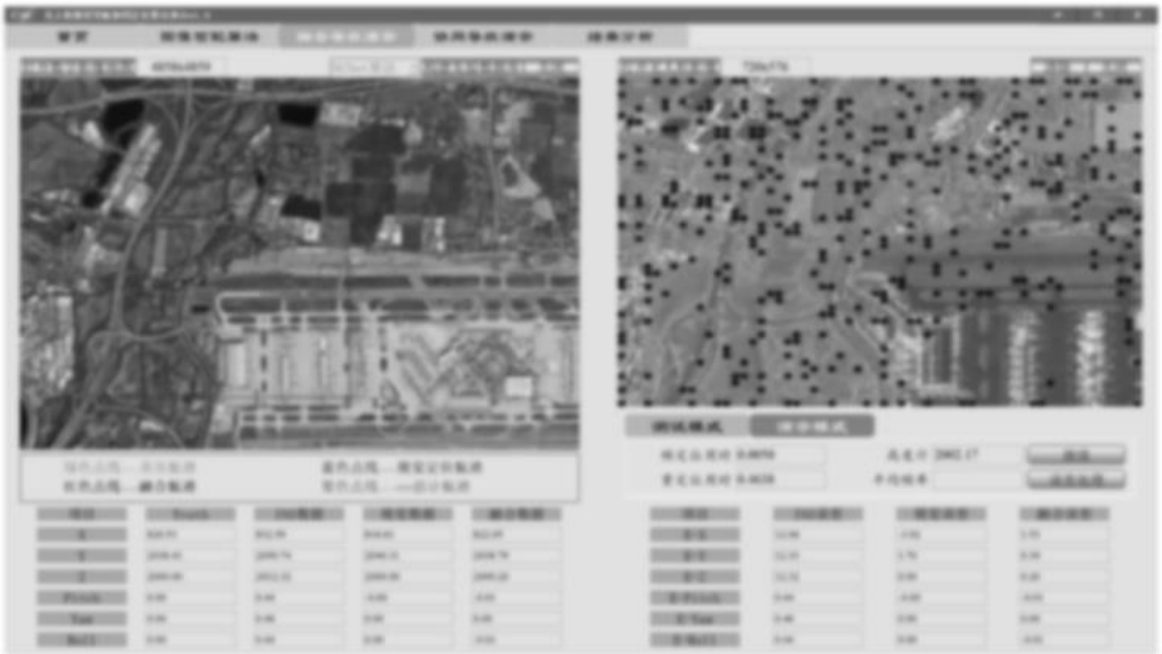


图5.4 融合导航界面

融合导航页面如图 5.4 所示，页面上半部分用来加载和显示超大数字影像地图以及无人机影像，数字影像地图支持缩放和拖动，无人机影像由连续的视频帧构成，在算法运行处理过程中支持视频的暂停和继续，可在数字影像地图显示区域顶部选择要使用的导航方案，支持 SPG-SE 和 NCNet 两种方案。

在算法启动后页面下半部分将实时显示无人机视频每一帧的真值、INS 推算值、视觉位姿估计值和数据融合位姿值，以及对应误差。此外，算法还分为演示模式和测试模式，在演示模式下，将在数字影像地图显示区域实时绘制各数据的轨迹以及当前图像匹配结果（红色框），在无人机影像显示区域绘制当前检测或跟踪到的点特征，以便于观察图像匹配和跟踪的点特性；由于绘制点线较为占用系统资源，因此在测试模式下不再显示区域绘制任何点线，以提升算法的运行速度。

协同导航页面如图 5.5 所示，除将无人机影像替换为高空无人机影像和低空无人机影像区域，以及高低空可分别选择要使用的图像匹配算法外，其他支持的功能与融合导航页面完全一致。

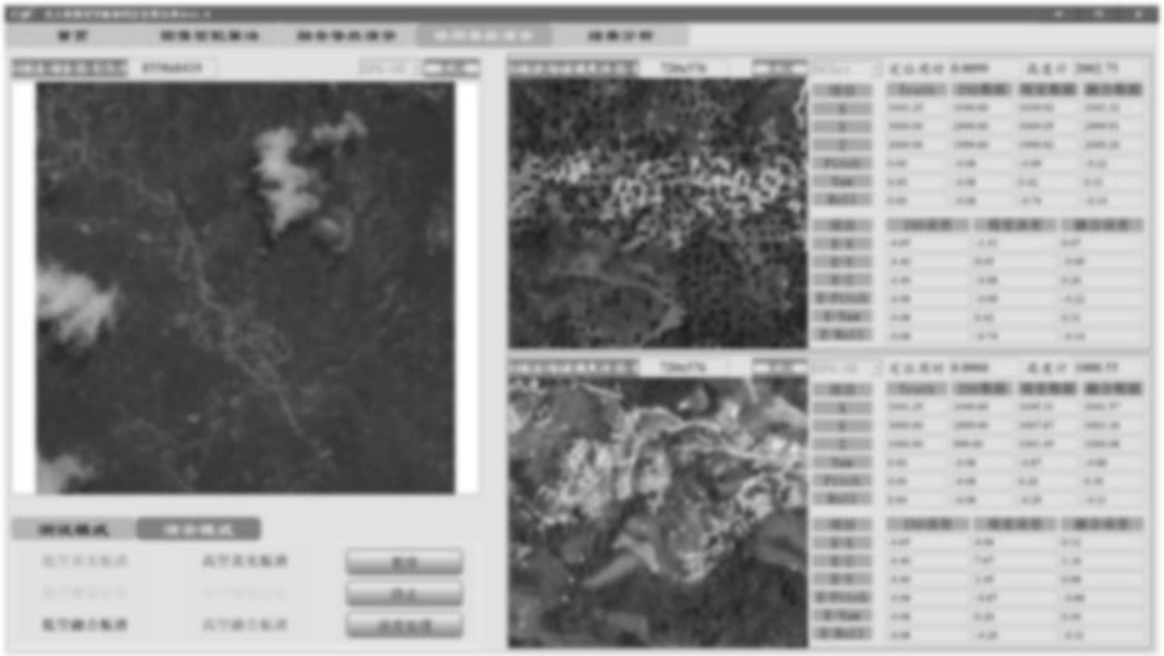


图5.5　协同导航界面

结果分析页面如图 5.6 所示，融合导航演示和协同导航演示的误差处理结果将显示在该页面，融合导航演示默认显示在低空区域，高空和低空显示区域均包含高度计误差、六轴姿态的视觉结果误差和融合结果误差曲线。

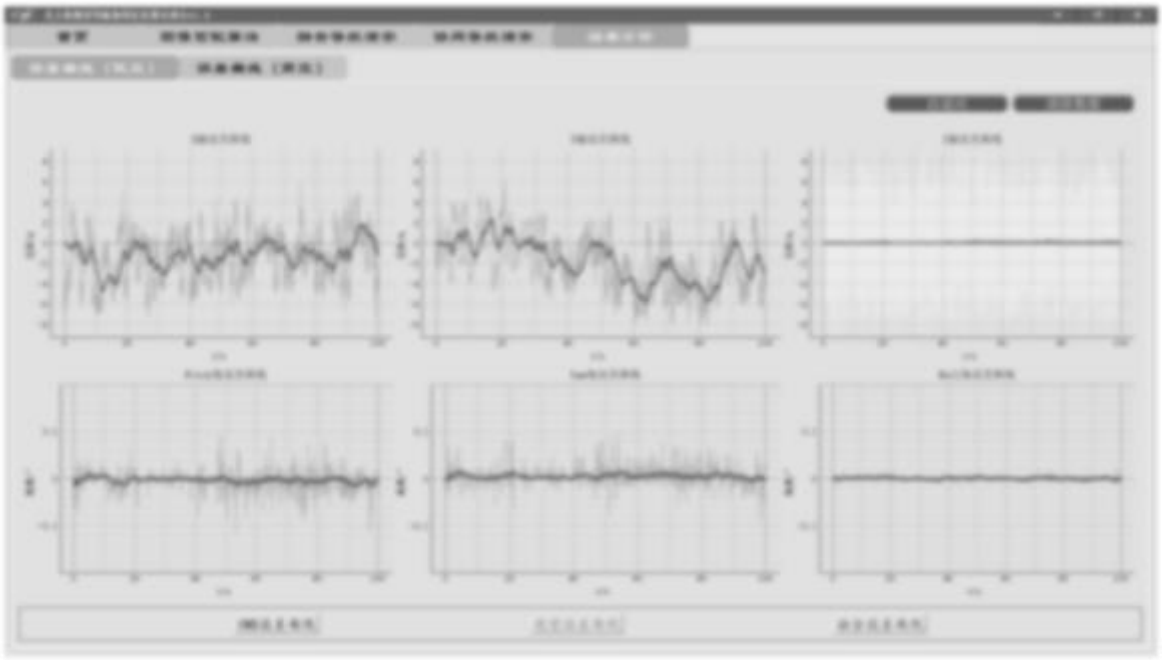


图5.6 结果分析界面

5.4 本章小结

本章主要介绍了 GNSS 拒止下的无人机视觉辅助融合导航定位软件的开发环境、界面设计、具体功能以及操作方式。本软件实现了本研究所展开的技术讨论相关的算法演示，搭建了一个清晰和简单的操作演示平台。

第六章 总结与展望

6.1 工作总结

本文主要研究了 GNSS 拒止情况下的视觉辅助无人机导航技术，具体来说，在 GNSS 拒止条件下，针对遥感数字影像地图和无人机实时拍摄的地面影像存在巨大差异的问题，实现基于图像匹配的无人机视觉位姿估计，并在惯性导航系统和高度计等传感器的辅助下，建立多源信息融合导航系统，实现无人机的长时间高精度稳定导航和定位。

本文主要完成的工作如下：

（1）针对无人机在视觉特征显著区域（I 类异源影像）的成像特点，提出了一种基于 SPG 算法的特征稀疏增强匹配策略，通过无监督训练的 SuperPoint 网络先验提取数字影像地图增强序列中的特征点并建立先验数据库，然后利用 SuperGlue 网络将无人机影像与数据库特征进行匹配，相比于其他算法和直接使用 SPG 算法，本文方法在保障匹配精度的同时，相当程度上提升了匹配速度并兼具了鲁棒性。

（2）针对无人机在弱纹理区域的 II 类异源影像的成像特点，提出了一种端对端的基于邻域共识的图像匹配算法，通过提取密集特征的方式来建立两幅图像之间的最佳匹配关系。实验证明，相比于其他深度网络，该方法能够在 II 类异源图像间实现稳定匹配，鲁棒性较强。

（3）针对 GNSS 拒止情况下的视觉辅助导航，建立了基于视觉的无人机位姿估计框架，通过图像匹配和光流跟踪的方式，能够保证视觉位姿的实时解算。同时为了解决低空无人机视野范围小因此在部分区域视觉匹配可能连续失败的问题，结合高低空无人机各自的优缺点，提出了一种高低空无人机视觉协同导航的方案，并在此基础上进行了实验，该方案能够保证在高空无人机稳定匹配的情况下低空无人机视觉位姿估计的长时间稳定性。

（4）针对视觉位姿估计结果频率较低且定位不稳定的问题，提出了一种基于卡尔曼的多源信息融合导航方案，结合高频的 IMU 数据和高度计数据实现了无人机的融合导航，保障了无人机位姿结果的高频连续性以及导航的精度和鲁棒性，并在后续通过实验进行了验证。

（5）设计搭建了本研究的算法软件，实现了各系统的可视化演示功能。主要包括异源影像匹配系统、融合导航系统和高低空无人机协同导航系统，此外还包括导航飞行轨迹、位姿参数、误差曲线动态显示等功能。

6.2 工作展望

本研究虽然做出了一些成果，但仍然存在一定的局限性，如下：

（1）尽管本研究引入了诸如 SAR 等传感器来作为无人机视觉数据的采集设备，且匹配算法能够在 SAR 和可见光图像间实现稳定匹配，但本文使用的是 GRD 地距多视影像，即已经进行过噪声处理、定标校正、多视处理后的影像，并没有涉及到对 SAR 原始数据的处理，后续可以补充无人机实际飞行过程中源数据的实时处理对图像匹配结果的影响，甚至于探索将 SAR 原始数据处理过程与图像匹配过程融合在一个深度网络中的方案。

（2）进一步探索深度学习在无人机导航中的应用，例如基于深度学习模型的无人机位姿估计技术，相比于通过单应性变换来约束不在同一平面上的匹配点对，深度模型或许可以从无人机视频帧直接估计特征点的三维信息，从而建立包含更多信息的 PnP 模型。

（3）多无人机协同乃至编队飞行已经成为未来发展的趋势，如何实现多无人机之间的协同位姿估计和定位，更充分的利用无人机编组间的信息也是值得探索和研究的方向。

参考文献

[1] 赵春晖,周昳慧,林钊等.无人机景象匹配视觉导航技术综述[J].中国科学:信息科学, 2019, 49(05): 507-519.

[2] 李磊,王彤,蒋琪.从美军 2042 年无人系统路线图看无人系统关键技术发展动向[J].无人系统技术, 2018, 1(04): 79-84.

[3] XIJIA L, XIAOMING T, YIPING D, et al. Visual information assisted UAV positioning using priori remote-sensing information[J]. Multimedia tools and applications, 2018, 77(11): 14461-14480.

[4] LINDEBERG T. Feature detection with automatic scale selection[J]. International Journal of Computer Vision, 1998, 30(2): 79-116.

[5] LOWE DG. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[6] MIKOLAJCZYK K, SCHMID C. Scale & Affine Invariant Interest Point Detectors[J]. International Journal of Computer Vision, 2004, 60(1): 63-86.

[7] BAY H, TUYTELAARS T, VAN GOOL L. SURF Speeded Up Robust Features[J]. In Proceedings of the European Conference on Computer Vision, 2006, 3951: 404–417.

[8] 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.

[9] DAVID M, ARTURO B, TOMASZ A, et al. DARTS: Efficient scale-space extraction of daisy keypoints[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2010, 2416–2423.

[10] MAINALI P., LAFRUIT G, YANG Q, et al. SIFER: Scale-Invariant Feature Detector with Error Resilience[J]. International Journal of Computer Vision, 2013, 104(2): 172–197.

[11] ALCANTARILLA P F, BARTOLI A, DAVISON A J. KAZE features[C]. In Proceedings of the European Conference on Computer Vision, 2012, 214–227.

[12] 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.

[13] SALTI S, LANZA A, DI STEFANO L. Keypoints from Symmetries by Wave Propagation[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, 2898–2905.

[14] DIVYA S V, PAUL S, PATI U C. Structure tensor-based SIFT algorithm for SAR image registration[J]. IET Image Processing, 2020,14(5): 929-938.

[15] VERDIE Y, YI K, FUA P, et al. TILDE: A Temporally Invariant Learned Detector[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, 5279–5288.

[16] LENC K, VEDALDI A. Learning Covariant Feature Detectors[C]. In Proceedings of the European Conference on Computer Vision, 2016, 100–117.

[17] SAVINOV N., SEKI A, LADICKY L, et al. Quad-Networks: Unsupervised Learning to Rank for Interest Point Detection[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, 1822–1830.

[18] ZHANG L, RUSINKIEWICZ S. Learning to detect features in texture images[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, 6325–6333.

[19] YI K M, TRULLS E, LEPETIT V, et al. LIFT: Learned Invariant Feature Transform.[C]. In Proceedings of the Computer Vision and Pattern Recognition, 2016, 467-483

[20] PAUTRAT R, LARSSON V, OSWALD M R, et al. Online Invariance Selection for Local Feature Descriptors[J]. In Proceedings of the Computer Vision and Pattern Recognition, 2020

[21] TIAN Y, FAN B, WU F. L2-Net: Deep Learning of Discriminative Patch Descriptor in Euclidean Space[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, 6128-6136

[22] MISHCHUK A, MISHKIN D, RADENOVIC F, et al. Working hard to know your neighbor's margins: Local descriptor learning loss[J]. Advances in neural information processing systems, 2017, 30.

[23] ZHANG W. Robust registration of SAR and optical images based on deep learning and improved Harris algorithm[J]. Scientific Reports, 2022, 12(1): 5901.

[24] MIKOLAJCZYK K, SCHMID C. A performance evaluation of local descriptors[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1615-1630.

[25] 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.

[26] DETONE D, MALISIEWICZ T, RABINOVICH A. Superpoint: Self-supervised interest point detection and description[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, 224-236.

[27] ONO Y, TRULLS E, FUA P, et al. LF-Net: Learning local features from images[J]. Advances in neural information processing systems, 2018, 31.

[28] SHEN X, WANG C, LI X, et al. RF-NET: An end-to-end image matching network based on receptive field[C]. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

Recognition, 2019, 8132-8140.

[29] LUO Z, ZHOU L, BAI X, et al. ASLFeat: learning Local Features of Accurate Shape and Localization[M]. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, 6589–6598.

[30] MOO Y K, VERDIE Y, FUA P, et al. Learning to Assign Orientations to Feature Points[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 107–116.

[31] REVAUD J, WEINZAEPFEL P, DESOUZA C, et al. R2D2: Repeatable and Reliable Detector and Descriptor[J]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.

[32] DUSMANU M, ROCCO I, PAJDLA T, et al. D2-Net: A trainable CNN for joint detection and description of local features[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, 8092-8101

[33] CHOY C B, GWAK J, SAVARESE S, et al. Universal correspondence network[C]. In Advances in neural information processing systems, 2016, 30: 2414–2422.

[34] ROCCO I, CIMPOI M, ARANDJELOVIC R, et al. NCNet: Neighbourhood Consensus Networks for Estimating Image Correspondences[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(2): 1020-1034

[35] HAN K, REZENDE R S, HAM B, et al. Scnet: Learning semantic correspondence[J]. In Proceedings of the IEEE International Conference on Computer Vision, 2017, 1831–1840.

[36] PLÖTZ T, ROTH S. Neural nearest neighbors networks[C]. In Advances in Neural information processing systems, 2018, 1087–1098.

[37] CHEN Y C, HUANG P H, YU L, et al. Deep semantic matching with foreground detection and cycle-consistency[J]. In Proceedings of the Asian Conference on Computer Vision, 2018, 347–362.

[38] KIM S, MIN D, LIN S, et al. Discrete-continuous transformation matching for dense semantic correspondence[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(1): 59–73.

[39] WANG Q, ZHOU X, DANIILIDIS K. Multi-image semantic matching by mining consistent features[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, 685–694.

[40] ZHOU Q, SATTLER T, LEAL-TAIXE L. Patch2Pix: Epipolar-Guided Pixel-Level Correspondences[J]. Computer Vision and Pattern Recognition, 2021, 4669-4678

[41] SUN J, SHEN Z, WANG Y, et al. LoFTR: Detector-free local feature matching with

transformers[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021, 8922-8931.

[42] SHAN M, WANG F, LIN F, et al. Google map aided visual navigation for UAVs in GPS-denied environment[C]. 2015 IEEE international conference on robotics and biomimetics, 2015: 114-119.

[43] SHAN M, GAO Z, TANG Y, et al. Google Map Oriented Robust Visual Navigation for MAVs in GPS-denied Environment[M]. Autonomy of Single UAV. 2020.

[44] PATEL B. Visual Localization for UAVs in Outdoor GPS-denied Environments[M]. University of Toronto (Canada), 2019.

[45] JUREVIČIUS R, MARCINKEVIČIUS V, ŠEIBOKAS J. Robust GNSS-Denied Localization for UAV Using Particle Filter and Visual Odometry[J]. Machine Vision and Applications, 2019, 30(7): 1181–1190.

[46] CHEN Y, JIA Q, ZHAO Y. Rough Waypoint Extraction Algorithm Based on Mean Shift Clustering Saliency Analysis[C]. 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). IEEE, 2020, 1: 1345-1349.

[47] VOLKOVA A, GIBBENS P W. More robust features for adaptive visual navigation of UAVs in mixed environments[J]. Journal of intelligent & robotic systems, 2018, 90(1): 171-187.

[48] ZHAO S. Multi-sensor Image Registration for Precisely Locating Time-sensitive Objects[C]. 2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC). IEEE, 2019, 144-148.

[49] TIAN X, SHAO J, YANG D, et al. UAV-Satellite View Synthesis for Cross-view Geo-Localization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 1-1.

[50] HAI J, HAO Y, ZOU F, et al. A Visual Navigation System for UAV under Diverse Illumination Conditions[J]. Applied Artificial Intelligence, 2021, 35(2): 1-21

[51] VENABLE D T. Improving real world performance of vision aided navigation in a flight environment[R]. Air Force Institute of Technology WPAFB, 2016.

[52] YU M, CUI H, LI S, et al. Database construction for vision aided navigation in planetary landing[J]. Acta Astronautica, 2017, 140: 235-246.

[53] TRAJKOVI´C M., HEDLEY M. Fast Corner Detection[J]. Image and Vision Computing, 1998, 16(2), 75–87.

[54] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An Efficient Alternative to SIFT or SURF[C]. In Proceedings of the IEEE International Conference on Computer Vision, 2011, 2564–2571.

[55] BIAN J, LINW Y, MATSUSHITA Y, et al. GMS: Grid-Based Motion Statistics for Fast,

Ultra-Robust Feature Correspondence[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017, 2828-2837.

[56] SARLANP E, DETONE D, MALISIEWICZ T, et al. SuperGlue: Learning Feature Matching With Graph Neural Networks[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, 4937-4946.

[57] 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, 24(6): 381-395.

[58] BAILO O, RAMEAU F, JOO K, et al. Efficient Adaptive Non-Maximal Suppression Algorithms for Homogeneous Spatial Keypoint Distribution[J]. Pattern Recognition Letters, 2018, 106: 53-60.

[59] 武汉大学地像天图课题组.RESEARCH RESOURCES[DB/OL] (2022-8-29) [2022-8-29]. https://skyearth.org/research/.

[60] GAO X S , HOU X R , TANG J , et al. Complete solution classification for the perspective-three-point problem[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2003, 25(8): 930-943.

[61] LEPETIT V , MORENO-NOGUER F , FUA P . EPnP: An Accurate O(n) Solution to the PnP Problem[J]. International Journal of Computer Vision, 2009, 81(2): 155-166.

[62] PENATE-SANCHEZ, ADRIAN, ANDRADE-CETTO, et al. Exhaustive linearization for robust camera pose and focal length estimation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(10):2387-2400.

[63] BARATH D, MATAS J, NOSKOVA J. MAGSAC: Marginalizing Sample Consensus[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 10189-10197.

[64] 邓志红等. 惯性器件与惯性导航系统[M]. 北京：科学出版社, 2012

[65] ZAFER M, MOHAMMAD R A. A perturbation-approach error model derivation of local-level gimbaled INS to determine its oscillatory modes analytically [C]. 2017 Iranian Conference on Electrical Engineering (ICEE), 2017, 573-578.

[66] LEMENAGER E , BOUET T , BRAIBANT V . Kalman filtering: A new approach for building global approximations. Application to the inverse optimization of an explosively formed penetrator (EFP)[J]. Structural Optimization, 1997, 14(2-3):158-164.