学士学位论文

融合深度学习特征的无人机影像匹配方法

学　　号：　　20191000466

姓　　名：　　李东泽

学 科 专 业：　　智能科学与技术

指 导 教 师：　　姜三 副教授

培 养 单 位：　　计算机学院

二〇二三年五月

中国地质大学（武汉）学士学位论文原创性声明

本人郑重声明：本人所呈交的学士学位论文《融合深度学习特征的无人机影像匹配方法》，是本人在指导老师的指导下，在中国地质大学（武汉）攻读学士学位期间独立进行研究工作所取得的成果。论文中除已注明部分外不包含他人已发表或撰写过的研究成果，对论文的完成提供过帮助的有关人员已在文中说明并致以谢意。

本人所呈交的学士学位论文没有违反学术道德和学术规范，没有侵权行为，并愿意承担由此而产生的法律责任和法律后果。

学位论文作者签名：李东泽

日 期： 2023 年 5 月 30 日

摘要

无人机影像匹配是遥感图像处理领域的重要研究方向之一。传统的影像匹配方法主要基于手工设计的特征描述符，在小尺度、低变形程度下具有较好的匹配效果，但在复杂场景下的匹配效果不佳。近年来，深度学习在影像匹配中得到了广泛应用。深度学习具有自动学习特征的能力，成为影像匹配的重要方法。

目前，卷积神经网络 CNN 常用于影像匹配任务的特征提取。同时，为了提高匹配的准确性和鲁棒性，融合深度学习特征和经典手工特征成为重要的影像匹配方法。其中，基于 CNN 和 SIFT 的影像匹配方法是最为常见的一种方法。该方法通过将 CNN 提取到的特征向量和 SIFT 关键点进行融合，提高了匹配的准确性和鲁棒性。为此，本文研究了融合深度学习特征的匹配方法，以满足对无人机影像数据进行特征匹配的需求。本论文的主要工作如下：

（1） 本文研究和总结了传统的影像匹配方法与深度学习方法，以及 SFM 三维重建方法。本文研究了传统 SIFT 方法，该方法能够自适应地检测图像中的局部特征点，并提取这些特征点的特征描述子，具有旋转、尺度不变性。但计算量大、参数敏感和特征描述子维数大等缺点也限制了其在实际应用中的推广。通过融合深度学习方法可以进一步改善上述问题。

（2） 本文研究了 Geodesc 方法，该方法使用基于 L2-Net 的网络结构，通过在 GL3D 数据集上进行模型训练。之后我们将 Geodesc 应用于无人机影像匹配任务。通过实验结果的分析和对比，我们证明了 Geodesc 方法在图像匹配任务中的优越性和有效性。

关键词：本地特征；特征描述符；深度学习

Abstract

Unmanned aerial vehicle (UAV) image matching is one of the important research directions in the field of remote sensing image processing. Traditional image matching methods are mainly based on manually designed feature descriptors, which have good matching performance at small scales and low deformation degrees, but poor matching performance in complex scenes. In recent years, deep learning has been widely used in image matching. Deep learning has the ability to automatically learn features and has become an important method for image matching. Currently, Convolutional Neural Networks (CNN) are commonly used for feature extraction in image matching tasks. In order to improve the accuracy and robustness of matching, the fusion of deep learning features and classical manual features has become an important image matching method. Among them, the image matching method based on CNN and SIFT is the most common. This method improves the accuracy and robustness of matching by fusing the feature vectors extracted from CNN and SIFT key points. To this end, this article studied the matching method that integrates deep learning features to meet the needs of feature matching for unmanned aerial vehicle (UAV) image data. The main work of this paper is as follows:

(1) This paper studies and summarizes traditional image matching methods, deep learning methods, and SFM (Structure from Motion) 3D reconstruction methods. This paper studied the traditional SIFT method, which can adaptively detect local feature points in the image and extract the feature descriptors of these feature points, and has rotation and scale invariance. However, its disadvantages such as high computational complexity, parameter sensitivity, and high-dimensional feature descriptors also limit its promotion in practical applications. By combining deep learning methods, the above problems can be further improved.

(2) This paper studies the Geodesc method, which uses a network structure based on L2-Net and is trained on the GL3D dataset. Afterwards, we will apply Geodesc to the task of matching drone images. Through analysis and comparison of experimental results, we prove the superiority and effectiveness of the Geodesc method in image matching tasks.

Key Words: LocalFeatures;Feature Descriptors;Deep Learning

目　　录

第一章 绪论

1.1 研究意义

无人机影像匹配是无人机应用中的一个关键问题，其解决方案对无人机遥感技术的发展具有重要意义。传统的影像匹配方法主要依靠手工选择特征点和设计特征描述符，而这些方法在复杂场景下的匹配效果不佳。近年来，深度学习在计算机视觉领域产生了非常大的影响，其具有自动学习特征的能力，成为解决影像匹配问题的重要手段。

本文研究的融合深度学习特征的无人机影像匹配方法，旨在提高影像匹配的准确性以及鲁棒性。通过将深度学习网络引入到影像匹配过程中，可自动提取影像特征。避免了手工选择特征点和设计特征描述符的过程。在实验中，我们对比了传统的 SIFT[1]与融合深度学习特征的 Geodesc[2]方法对不同无人机影像的匹配效果，计算了两种算法提取的特征点数量与特征点匹配准确度。实验结果显示，融合深度学习特征方法的特征点提取与匹配的效果相对于传统方法有显著提升。

首先，本文所研究的方法对于无人机遥感应用具有重要意义。在地图绘制、资源监测、环境监测、农业智能化等领域，无人机遥感技术都有广泛的应用。而影像匹配是这些应用中的一个关键环节，对于实现精准测量、准确判断与精细化管理具有重要意义。通过融合深度学习特征的无人机影像匹配方法，可以提高影像匹配的准确性以及鲁棒性，为无人机遥感应用提供更加可靠的数据支持和技术保障。

其次，在无人机遥感应用中，三维重建是一个非常重要的环节。而传统特征匹配方法在三维重建应用中存在许多缺点，这些缺点会导致三维重建的精度下降、效率低下等问题。本文所研究的融合深度学习特征的 Geodesc 方法，能够自动学习图像的本质特征，具有较好的鲁棒性，能够有效应对光照、遮挡等变化对特征匹配的影响。与此同时，该方法生成的特征描述符具有尺度不变性与旋转不变性，这意味着无论图像的尺度和旋转角度如何变化，生成的特征描述符都能保持一致。这种方法的优点在于，它能够有效应对尺度变化和旋转变化对特征匹配的影响。

这些优点使得利用 Geodesc 算法生成的特征点描述符在三维重建中更加准确和鲁棒。

最后，本文的研究成果对于无人机应用的发展也具有重要意义。随着无人机相关技术的不断提升以及其应用领域的不断扩大，无人机影像匹配问题逐渐成为无人机应用中的一个重要问题。本文研究的方法为无人机应用提供了一种新的影像匹配技术，可以为无人机应用的发展提供支持与保障。

1.2 国内外研究现状

1.2.1 传统手工特征

在计算机视觉领域中，特征点的提取与匹配是一个非常重要的问题。特征点提取与匹配是指首先从图像中提取出具有鲁棒性与可重复性的局部特征点，随后将这些特征点进行匹配，以实现图像检索、目标追踪、图像拼接等应用与功能。这个领域的研究已经有了相当长的历史，从最早的手工设计算法到现在的深度学习算法，这个领域的研究也是一直在不断发展与进步。

对于特征点提取算法，最早研究出的方法是基于可微分函数的边缘检测算法。这些算法可以对图像中的边缘信息进行检测，但是这些算法的鲁棒性与可重复性相对而言都不是很好。伴随着时间的推移，研究者们研究出了鲁棒性与可重复性更好的特征点提取算法。其中比较著名的算法包括 SIFT，SURF[3]，ORB[4]等。SIFT 算法属于一种以尺度空间极值点检测为基础的算法，它可以提取出具有尺度不变性和方向不变性的特征点。SURF 算法是 SIFT 算法的改进版，它采用高斯差分函数来近似 LoG（Laplacian of Gaussian）函数，通过这种方式可以在不同尺度下检测到图像中的关键点。接下来对每个关键点使用 Haar 小波响应函数计算其特征向量，这些特征向量可以用来进行图像的匹配和识别。与 SIFT 算法相比，SURF 算法的计算速度不但有显著提升，并且该算法在鲁棒性上也是更胜一筹。ORB 算法首先使用 FAST[5]算法检测图像中的特征点，然后通过 BRIEF[6]算法对每个特征点进行特征描述，最后通过方向计算以及旋转操作使 ORB 算法具有旋转不变性和尺度不变性。该算法的运行速度相对更快，并且它还拥有较好的精度。以上所述这些算法在大规模图像检索、目标跟踪、图像拼接等应用中均表现较为出色，并且获得了非常不错的效果。然而，这些算法的性能受到很多因素的影响，如图像亮度、旋转、缩放、遮挡等，当遇到复杂场景时，它们的性能会受到很大影响。

1.2.2 深度学习特征

近年来，伴随着深度学习发展的不断推进，以及其在图像特征提取方面的出色表现，使它吸引了愈来愈多的研究人员的关注与研究。深度学习作为一种自动学习特征的方法，其不需要人工设计特征，因此在一定程度它上可以更好地适应不同的场景。深度学习被应用于特征点提取任务，例如 LIFT[7]，SuperPoint[8]，D2-Net[9]等算法。LIFT 算法由三个相互输入组件组成：探测器、方向估计器和描述符。每个组件都基于卷积神经网络（CNN）。该方法还用 soft argmax 函数取代了传统的非局部最大抑制（NMS）方法。SuperPoint 算法的核心是一个卷积神经网络，它可以将输入图像转换为一个稠密的特征点图。这个特征点图包含了每个像素的位置以及一个描述该位置的向量。这些向量可以用于匹配、跟踪和重建图像。另外，该算法还采用了一种新的非极大值抑制方法，可以在不损失关键点检测质量的情况下大幅降低计算成本，提高算法的效率。D2-Net 使用了密集连接的卷积块（DenseNet）与可变形卷积（Deformable Convolutional Networks，DCN）来提高特征提取的准确性和鲁棒性。在训练过程中，该算法使用了一种称为“多任务学习”的技术，将图像检索、匹配以及定位这三个任务结合在一起，共享模型的特征提取层，从而提高模型的泛化能力和效率。与传统特征提取算法相比，以上算法在准确率和速度上的提升都十分明显。同时深度学习还可以提供更丰富、更高级别的特征，例如语义特征和上下文特征，这些特征可以进一步提高特征点匹配的准确性。

传统特征点匹配算法一般采用手工设计的特征点描述符来完成匹配。这些描述符在图像旋转、缩放等变换中具有一定的不变性，但在遇到复杂场景时，性能仍然不尽如人意。近年来，深度学习在特征点匹配方面也得到了广泛应用， 研究者将深度学习应用于特征点匹配任务，例如 DeepMatching[10]，MatchNet[11]，SuperGlue[12]等算法，在精度与速度上都有显著提升。这些算法首先使用深度神经网络对实验数据中的图像进行特征提取以及描述符生成，然后使用一些特殊的网络结构对这些描述符进行匹配。例如在 DeepMatching 算法中，首先使用卷积神经网络（CNN）对两幅图像进行特征提取，然后将这些特征表示送入一个多层感知机（MLP）中，计算两幅图像之间的相似度得分。同时，为了提高算法的鲁棒性与匹配准确性，该算法还引入了多尺度金字塔与局部相似性约束等技术。MatchNet 使用一个共享网络来提取两个图像中的特征，然后使用一个子网络来计算特征之间的相似度，最后使用 Softmax 层计算匹配概率。该算法的创新之处在

于它使用了独特的三通道架构，其中每个通道都处理输入图像的不同方面，例如颜色、线条和纹理。SuperGlue 是在 SuperPoint 的基础上提出的一种匹配算法，它使用一种名为 SuperGlue 的新型神经架构从预先存在的局部特征中学习匹配过程。使用图神经网络和注意力机制来解决一个任务优化问题，并且能够处理部分点的可见性和遮挡，从而产生部分分配结果。

随着全球无人机市场的迅速扩大，无人机相关技术也不断提升与改进，在这种情况下，无人机影像三维重建技术也得到了越来越广泛的关注与应用。无人机影像三维重建技术是通过无人机获取地面影像数据，并利用计算机算法对影像数据进行处理和分析，最终生成三维模型的一种技术。在此基础上，这项技术被广泛应用于建筑、城市规划、地质探测等领域。而图像特征提取与匹配可用于提取描述场景特征的信息，提高三维重建的精度与准确性。因此本文研究了融合深度学习特征的 Geodesc 算法。该描述符不但在基于图块的基准测试中有良好效果，在图像的三维重建测试中也有更强的泛化能力。

1.3 主要研究内容

本文讨论了一种新的局部描述符学习方法，Geodesc，旨在缓解卷积神经网络（CNNs）在基于图像的三维重建基准测试方面的限制。该方法集成来自多视图重建的几何约束，有利于在数据生成、数据采样和损失计算方面的学习过程。本文展示了该方法在影像匹配上的优越性能。研究具体内容如下：

1.3.1 理论基础研究

本文首先讨论了传统影像匹配算法 SIFT，然后介绍了一种深度学习影像匹配方法 HardNet[13]，最后本文把理论研究重点放在了 Geodesc 上，研究其理论基础、建模流程。接着研究了三维重建方法 SFM[14][15][16][17]，该方法在训练数据生成时起到了重要作用。

1.3.2 基于Geodesc方法影像匹配结果与评估

本文影像匹配方法为 Geodesc。本文使用 GL3D 数据集进行模型的训练，并利用得到模型对无人机拍摄影像进行特征值匹配，并对匹配结果中的特征点数量已经精确度进行了评估，与传统 SIFT 方法进行了对比。

1.4 论文组织结构

本论文一共分为五个章节，每个章节的主要内容分别为：

第一章：介绍了传统影像匹配方法，融合深度学习特征的无人机影响匹配方法等方面的背景和研究意义，然后说明了本文的主要研究内容和组织结构。

第二章：介绍了传统 SIFT 影像匹配方法，以及深度学习影像匹配方法 HardNet。最后介绍了三维重建理论。

第三章：介绍并分析了 Geodesc 的网络结构，几何相似度，损失函数等核心改进。

第四章：介绍了实验基本环境，以及本实验所使用数据集的基本信息，并展示了实验结果并对实验结果进行了分析。

第五章：对全文进行了总结，指出了目前研究中存在的问题，并对今后的研究进行了展望。

第二章 影像匹配理论基础

2.1 SIFT影像匹配

SIFT 算法概述：尺度不变特征转换(SIFT, Scale Invariant Feature Transform)是一种图像处理领域中使用的局部特征描述算法。该算法具有良好的稳定性与不变性，可以适应旋转、尺度缩放、亮度的变化，并且对视角变化、仿射变换、噪声等有一定的抗干扰能力。

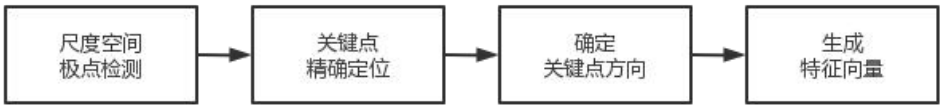

图 2.1 SIFT 算法步骤

2.1.1 尺度空间极点检测

尺度空间的构建旨在寻求在尺度变化中具有不变性的位置，这一目的可以通过对图像进行连续的尺度变化来实现。在尺度空间中，我们需要寻找稳定的特征点，也就是能够在所有可能的尺度变化中保持不变的点，而所选取的这些特征点通常是极点，能够确保图像不会因比例缩放和旋转变化而改变。

这里的尺度空间称为高斯金字塔，并通过采样法模拟图像的远近程度，通过高斯平滑模拟图像的粗细程度。在 SIFT 里，高斯金字塔的层数 S 和组数 0 有着如下设定：

$$O = [log_2 min(M, N)] - 3 \tag{2.1}$$

$$S = n + 3 \tag{2.2}$$

组数的设定是来自于提出 SIFT[5]算法的原始论文给出的经验值，理论上来说只要

$$O \leq [\log_2 \min(M, N)] \tag{2.3}$$

即可，层数的设定则是有着理论依据的，此处的 n 是我们想要提取特征点的图片层数，因为提取出高斯金字塔后需要计算层间差分来获得高斯差分金字塔(DOG, Difference of Gausssian)，所以高斯金字塔层数需要比 DOG 层数多 1，而计算特征值时要求在尺度层面，即上下相邻层间计算，则 DOG 层数要比特征层数多 2，则要求 S = n + 3

在构建高斯金字塔时，为了尽可能多的保留原始图像信息，首先，将原始图像进行升采样，即将图像尺寸扩大一倍，并将其作为高斯金字塔的第一组第一层。在此之后，将对第一组第一层图像进行高斯卷积操作，也就是进行高斯平滑或者说高斯滤波，可以通过这种方法得到第一组金字塔的第二层。金字塔中的每一幅图像都将通过 $L(x, y, \sigma)$进行表示，它的计算公式为：

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{2.4}$$

其中 $I(x, y)$表示图像 $G(x, y, \sigma)$为高斯函数

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{2.5}$$

接着，我们需要将之前得到的第一组第二层图像进行平滑操作，此时平滑因子为原来的 σ 与一个比例系数 k 相乘，这样便得到了新的平滑因子 σ =k* σ ，随后平滑后的结果图像便可以成为第一组金字塔中的第三层图像。然后，我们可以重复这个过程。最终，我们可以得到包含 S 层图像的高斯金字塔，规定

$$k = 2^{O+\frac{r}{n}}, r = 0,1\ldots, n+2 \tag{2.6}$$

高斯金字塔的构建还需要进行以下步骤：首先要将第一组金字塔的倒数第三层进行降采样，使其大小减半，并将得到的图像用作第二组金字塔中的第一层。然后，我们还需要对第二组金字塔中的第一层图像进行高斯平滑操作，其中的平滑因子为 σ 。这样，我们就得到了第二组金字塔的第二层。在此基础上，我们可以重复上述这个过程，对第二组金字塔的图像不断进行降采样以及高斯平滑操作，直到得到所需层数的金字塔图像。在最后，我们可以得到总计 O 组高斯金字塔，其中每组中包含 S 层图像，总共有 O*S 个图像。这些图像一起构成了高斯金字塔。

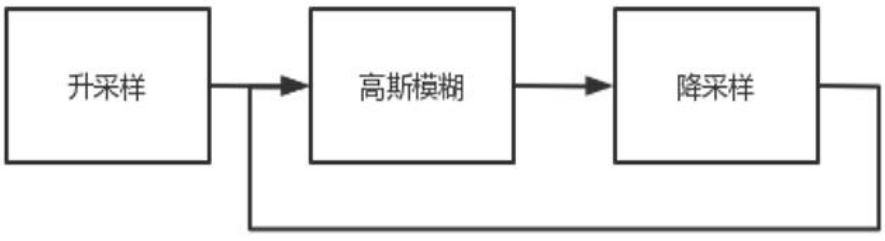

图 2.2 高斯金字塔构建步骤

而通过以前的研究我们可以得知，归一化的高斯拉普拉斯算子的极大值极小值相较于其他特征提取函数可以获得最稳定的图像特征，又尺度归一化高斯拉普拉斯算子和 DOG 函数有着如下关系：

$$G(x,y,k\sigma)-G(x,y,\sigma)\approx(k-1)\sigma^2\nabla^2G \tag{2.7}$$

又对于差分高斯金字塔有：

$$DOG=\frac{G(x,y,k\sigma)-G(x,y,\sigma)}{(k-1)\sigma}\approx\frac{\partial G}{\partial\sigma} \tag{2.8}$$

$$DOG\approx(k-1)\sigma^2\nabla^2G \tag{2.9}$$

在此基础上，利用高斯金字塔，我们可以很容易地构造出高斯差分金字塔。只需要对每一组内的相邻层进行相减即可。通过将高斯金字塔的第一组第二层与其第一组第一层相减就可以生成高斯差分金字塔中的第一组第一层。按照这种方式不断重复，将高斯金字塔每组内所以相邻图层进行相减得到每一个差分图像，最终的高斯差分金字塔便可以由所有差分图像构成。

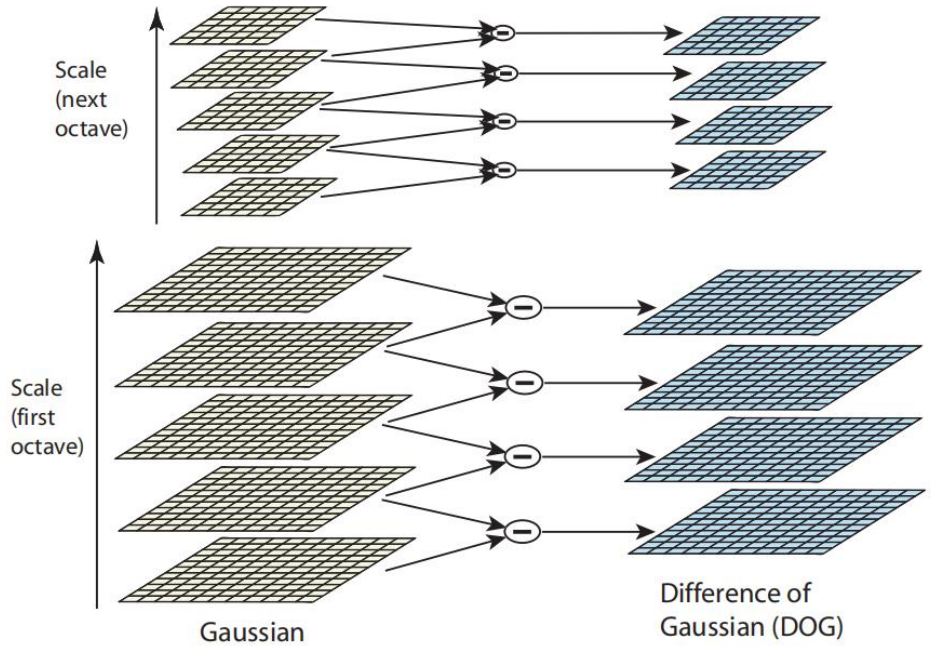

图 2.3 差分高斯金字塔构建示意图

而在 SIFT 算法中，特征点的提取便是在高斯差分金字塔（DOG）上进行的。

2.1.2 关键点精确定位

首先通过阈值化进行去噪处理，去除掉有可能由噪声引起的点。

$$val = \begin{cases} val, \ abs(val) > 0.5\frac{T}{n} \\ 0, \ otherwise \end{cases} \tag{2.10}$$

其中 T 为经验值 0.04，n 为之前提过的提取特征点数目。

随后通过将一个像素（用 X 标记）与 3x3 区域内的 26 个邻居像素（分别为当前图像中的 8 个邻居以及上面和下面尺度上的 9 个邻居）进行比较，检测高斯图像差值的最大值和最小值。将找到的极值点作为备选关键点。

但是因为高斯差分金字塔是离散的，备选关键点可能在真正的关键点附近，需要使用了尺度空间函数 D（x，y，σ）的泰勒展开（直到二次项），进行修正。

$$D(x) = D + \frac{\partial D^T}{\partial x}x + \frac{1}{2}x^T\frac{\partial^2 D}{\partial x^2}x \tag{2.11}$$

$x = (\mathrm{x,y,}\sigma)^T$是从这个点的偏移量。极值的位置，$\hat{x}$，是通过取这个函数对 x 的导数，并将其设为零：

$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2}\frac{\partial D}{\partial x} \tag{2.12}$$

将其带入

$$D(\hat{X}) = D + \frac{1}{2}\frac{\partial D^T}{\partial x}\hat{x} \tag{2.13}$$

若$|\mathrm{D}(\hat{\mathrm{x}})| < \mathrm{T}/\mathrm{n}$,则舍去特征点

由于我们想要提取的特征点为角点而非边缘，而上述方法只能保证选取到灰度值变化剧烈的点，而边缘点也符合这一特征，在边缘处，高斯差分算子的极值容易出现定义不明确的情况。在横跨边缘的地方，极值点有更大的主曲率，而在垂直边缘的方向，极值点的主曲率相对更小。可以根据一个 2×2 的海森矩阵(Hessian Matrix)H 求出主曲率，D 的主曲率与 H 的特征值成正比。令 α 为较大特征值，β 为较小的特征值。

$$H = \begin{bmatrix} D_{\mathrm{xx}} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{2.14}$$

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta \tag{2.15}$$

$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha + \beta \tag{2.16}$$

若 $\mathrm{Det(H)} < 0$ 舍去特征点

若矩阵行列式和迹不满足：

$$\frac{Tr(H)}{\mathrm{Det}(H)} < \frac{(\gamma_0+1)}{\gamma_0} \tag{2.17}$$

舍去该特征点，其中 γ_0为经验值，通常设为 10

2.1.3 确定关键点方向

为了实现对图像旋转的不变性，SIFT 算法给每个关键点赋予一个一致的方向，并使用该方向来描述关键点描述符。该算法将会在差分金字塔中检测出的关键点所在的高斯金字塔图像中的 3 σ 邻域窗口内获取像素的梯度和方向分布特征。在此，梯度的模值和方向可以通过下面的公式来计算：

$$m(x,y)=\sqrt{\big(L(x+1,y)-L(x-1,y)\big)^2+\big(L(x,y+1)-L(x,y-1)\big)^2} \tag{2.18}$$

$$\theta(x,y)=tan^{-1}\big(\big(L(x,y+1)-L(x,y-1)\big)\big/\big(L(x+1,y)-L(x-1,y)\big)\big) \tag{2.19}$$

在 SIFT 算法中，为了确定关键点方向，使用了一种梯度直方图统计法。具体来说，该算法会在以关键点为原点、一定区域内的图像像素点中，统计所有像素点的梯度与方向的直方图。这个直方图将 0 到 360 度的方向范围分为 36 个柱，每柱代表 10 度的方向。其中直方图的峰值方向所表示的为该关键点的主方向，或者说方向直方图中的峰值则表示的为该特征点处邻域梯度的方向。为了进一步提高特征点匹配的鲁棒性，该算法仅将峰值大于主方向峰值 80%的方向进行保留，并把它用作这一关键点的辅方向。在进行统计梯度方向和幅值时，SIFT 算法会在以特征点为圆心、半径为该特征点所在高斯图像尺度的 1.5 倍的圆内，统计所有像素的梯度方向以及梯度幅值，并使用 1.5 σ 的高斯滤波对其进行平滑处理。

2.1.4 生成特征向量

确定计算描述子所需的图像区域，通过计算关键点所在尺度的高斯图像可以得到描述子梯度方向直方图.。图像区域的半径可以通过以下公式进行计算

$$radius=\frac{3\sigma\sqrt{2}(d+1)+1}{2} \tag{2.20}$$

其中 d=4

关键点所在的半径区域，移至关键点方向。随后将区域分割为 4x4 的小块，对每个小块进行 8 个方向的直方图统计，记录每个方向的梯度强度，并将所有结果组合成一个 128 维的向量，以描述每个关键点的位置、尺度和方向。为了实现对光照和视角变化等影响的不变性，需要为每个关键点构建一个描述符。这个描述符使用一组向量来描述该关键点及其周围像素点，从而使其拥有不变性。并且这个描述符必须具有高度的独特性，才能使特征点正确匹配的概率得以提升。

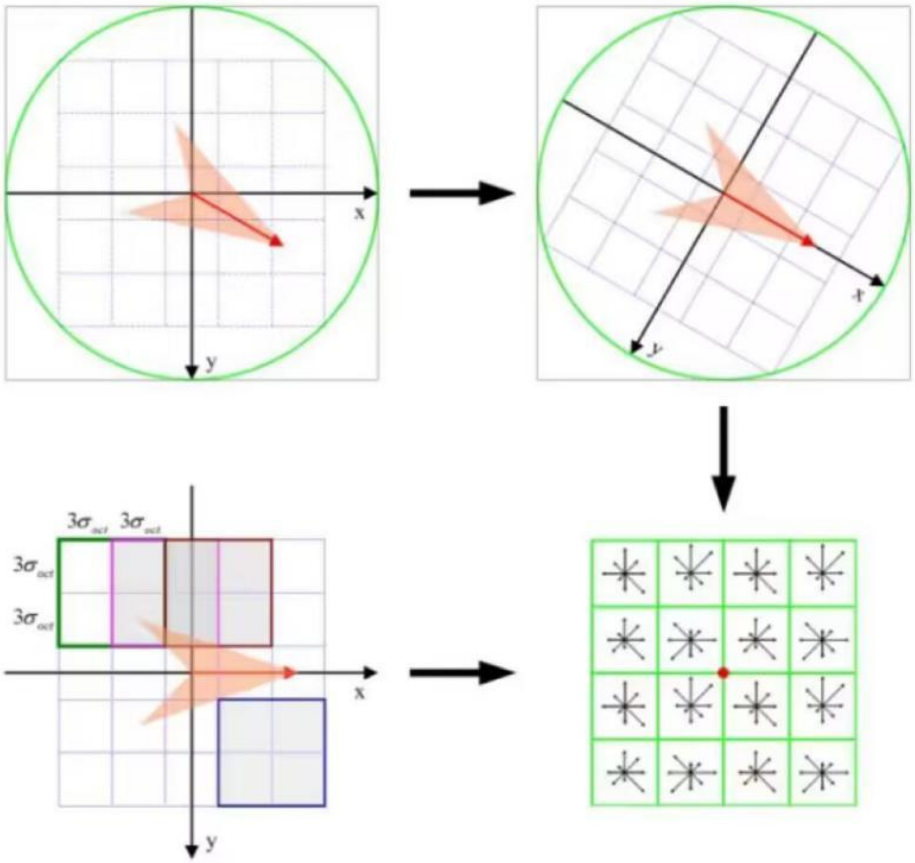

图 2.4 生成特征向量步骤

2.1.5 SIFT优缺点

首先 SIFT 算法能够在不同尺度下对相同的特征点进行检测，从而使其拥有尺度不变性，即在不同尺度下所提取出的图像特征能够具有相同的表达能力，或者是具有相同的特征描述符。这使得该算法在处理具有不同尺度的图像时具有非常优秀的鲁棒性。另外该算法还能够在不同旋转角度下检测到相同的特征点，从而拥有一定程度上的旋转不变性。这使得该算法在处理具有不同旋转角度的图像时具有非常出色的鲁棒性。而且该算法所提取的特征具有很高的独特性，能够有效地区分不同的目标和背景。但是该算法的处理速度较慢，由于该算法需要进行大量的高斯模糊以及差分运算，导致其处理速度相对较慢。并且其参数选择较困难，该算法中的参数较多，包括高斯模糊的标准差、高斯金字塔的层数、DoG 金字塔的层数等，选择合适的参数需要经验和时间的积累。该算法对于大规模图像匹配不够有效，由于该算法需要匹配的特征点较多，处理大规模图像匹配时效率不高。综上所述，SIFT 算法具有很好的尺度和旋转不变性，能够提取具有高独特

性的特征，但其处理速度较慢，对参数选择较为敏感，不太适用于大规模图像匹配。

2.2 HardNet影像匹配

目前研究表明，受到图块数据集大小以及多样性的限制，无法学习得到高质量的特征描述符。因此，HardNet 方法设计了一种新的损失函数用于网络训练，并设计了一种新的采样策略，以获得更高质量的实验数据。

采样策略：首先，批次 X=（A_i，P_i）i=1..n 中包含 n 对图块，随后 X 中的 2n 个图块通过网络传递，得到a_i，p_i分别为A_i，P_i的描述符，进而可以根据所得描述符构建距离矩阵 D，d 的计算公式为：

$$d(a_i, p_j) = \sqrt{2 - 2a_i p_j}, i = 1 \ldots n, j = 1 \ldots n \tag{2.21}$$

距离矩阵 D 的对角线上代表匹配描述符间的距离。随后我们要去找到距离a_i描述符距离最近的非匹配描述符和距离p_i描述符距离最近的非匹配描述符。其中$p_{j_{min}}$为最接近a_i的非匹配描述符，其计算公式为：

$$j\,argmin_{j=1\ldots n, j\neq i}\, d\,(a_i, p_j)_{min} \tag{2.22}$$

$a_{k_{min}}$为最接近p_i的非匹配描述符，其计算公式为：

$$k\,argmin_{k=1\ldots n, k\neq i}\, d\,(a_k, p_i)_{min} \tag{2.23}$$

至此便得到了$(a_i, p_i, p_{j_{min}}, a_{k_{min}})$，如果$d(a_i, p_{j_{min}}) < d(a_{k_{min}}, p_i)$则得到$(a_i, p_i, p_{j_{min}})$,否则得到$(p_i, a_i, a_{k_{min}})$

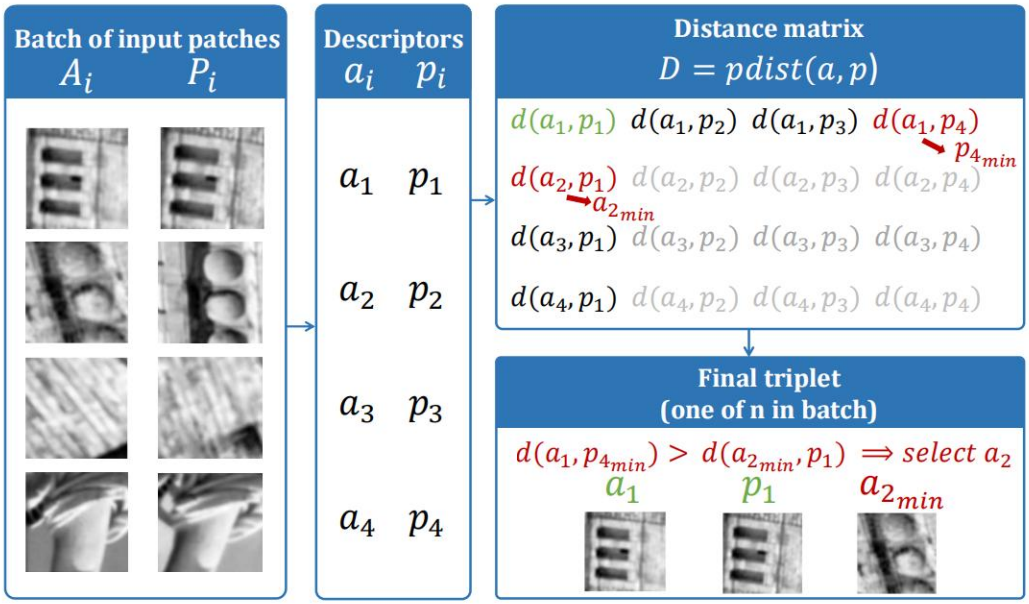

图 2.5 采样步骤

我们的目标是最大化匹配描述符与最近的非匹配描述符之间的距离。这些 n 个三元组距离被输入到损失函数中：

$$L=\frac{1}{n}\sum_{i=1,n}\max\left(0,1+d\left(a_i,p_i\right)-\min\left(d\left(a_i,p_{j_{\min}}\right),d\left(a_{k_{\min}},p_i\right)\right)\right) \tag{2.24}$$

HardNet 使用的网络结构与 L2-Net[18]相同，每个卷积层之后都是批量归一化和 ReLU，除了最后一层。需要在最后一个卷积层之前将正则化去除。

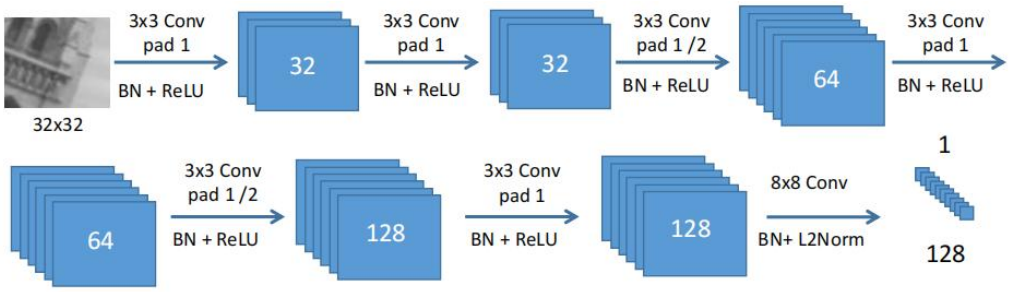

图 2.6 HardNet 网络结构

2.3 SFM三维重建

2.3.1 三维重建基本原理

三维重建旨在通过多个图像或视频来还原三维场景的结构和纹理。目前，三维建模方法可分为基于几何和基于深度学习两种。基于几何的方法较为传统，它是利用多张图像间的几何关系来还原场景的三维结构。其优点在于准确性高、可解释性强，但需要对图像进行精确的标定和配准，对图像质量和拍摄角度有一定要求。而基于深度学习的方法则是近年来发展起来的一种新型三维建模方法，它通过学习深度神经网络来直接预测场景的三维结构。其优点在于不需要对图像进行精确的标定和配准，对图像质量和拍摄角度的要求较低，但准确性与可解释性相对较低。

在三维重建方面，传统的三角测量法、立体匹配法、多视几何法等方法仍然是研究的重点。虽然这些方法准确性高、可解释性强，但需要对图像进行精确的标定和配准，对图像质量和拍摄角度有一定要求。因此，当前的研究热点是如何提高这些方法的自动化和鲁棒性。在深度学习技术不断提升与进步的同时，以深度学习为基础的三维重建方法同样受到了愈来愈多研究者的高度重视。其中，基于卷积神经网络（CNN）的方法、基于生成对抗网络（GAN）的方法等已经成为研究的重要方向。这些方法不需要对图像进行精确的标定和配准，对图像质量和拍摄角度的要求较低，但准确性和可解释性相对较低。因此，当前的主要研究方

向是关于如何进一步提高基于深度学习的三维重建技术的准确性和可解释性。除此之外，三维重建技术还在多个领域的交叉应用中的得到了广泛的应用。例如，在虚拟现实领域中，三维重建技术已经获得愈来愈多的关注，进而成为了研究热点，该技术可用于虚拟场景模拟。在医疗图像领域中，三维重建技术也在医学图像的重建与分析等项目中得到了广泛的应用。以及在文化遗产保护领域中，三维重建技术也被用于数字化文物保护与展示，例如对古建筑、古迹、雕塑等进行数字化重建和保存。

2.3.2 SFM运动恢复结构

标准 SFM[19]是一种经典的三维重建方法，常用于从多张二维图片中还原其三维结构。其基本思想是通过多张图片中的特征点匹配，计算出相机的位姿和三维场景点的位置，从而还原出整个三维结构。该方法的主要步骤如下：首先对于每张输入的图片，利用特定的算法（如 SIFT、SURF 等）提取出其中的特征点以及特征点所对应的特征描述符。然后对于每两张图片之间的特征点，可以通过对特征描述符的相似度进行计算，来找到它们之间的最佳匹配点对。之后通过匹配点对以及相机内参的已知信息，使用 RANSAC[21]等算法估计相机的位姿，即相机在三维坐标系中的旋转和平移。接着通过相机位姿和匹配点对，可以计算出每个特征点的三维坐标，从而还原出整个三维点云。而对于估计出的三维点云，可以采用 BA[22]等优化算法，对点云进行优化，提高精度和稳定性。最后对于还原出的三维模型，可以使用纹理映射等技术，将其贴上原始图片中的纹理，从而得到更加逼真的三维重建结果。

在标准 SFM，相机位姿估计是非常重要的一步，其准确性将直接影响到三维重建的质量。相机位姿估计的方法通常分为基于本质矩阵的方法和基于基础矩阵的方法两种。其中基于本质矩阵的方法应用更加广泛，它的基本思想是通过对相机的内参矩阵进行分解，求解出相机的旋转矩阵和平移向量。而基于基础矩阵的方法则是通过对特征点匹配进行八点法或者最小二乘法求解，从而获得基础矩阵，从而推导出相机的位姿。

标准 SFM 是一种比较经典的三维重建方法，其基本思想是通过多张图片的匹配和相机位姿估计，还原出三维点云，然后再对其进行优化和纹理贴图等处理，得到最终的三维模型。

2.4 本章总结

本章内容是第二章 SIFT 影像匹配与深度学习影响匹配以及 SFM 三维重建。本章节首先介绍了传统的 SIFT 特征描述符，关于其算法流程，核心原理等内容。然后介绍了融合深度学习特征的 HardNet 方法。并在本章的末尾介绍了三维重建与标准 SFM 方法，对该算法的基本流程进行了说明与介绍。

第三章 核心算法

3.1 主要思路

目前，在基于 patch 的数据基准上使用 CNN 网络学习局部特征已经取得了很大的进展。但是，将其应用于具体的 3D 重建等任务的效果仍然不如传统的特征，例如 SIFT，SURF，ORB 等。因此，本文对一种新型的局部特征学习方法 Geodesc 进行了研究，该方法主要是从多视图重建中学习所得的几何约束信息。

3.2 网络架构

该方法借鉴了 L2-Net 中的网络结构，未使用池化层继续降维，并使用步幅卷积层进行网络内部下采样而构建。除了最后一层卷积层外，每一个卷积层之后都跟随着一个批量归一化（BN）层，它的加权和偏置参数都被固定为 1 和 0。最后一个卷积层之后的 L2 归一化层生成最终的 128 维特征向量。

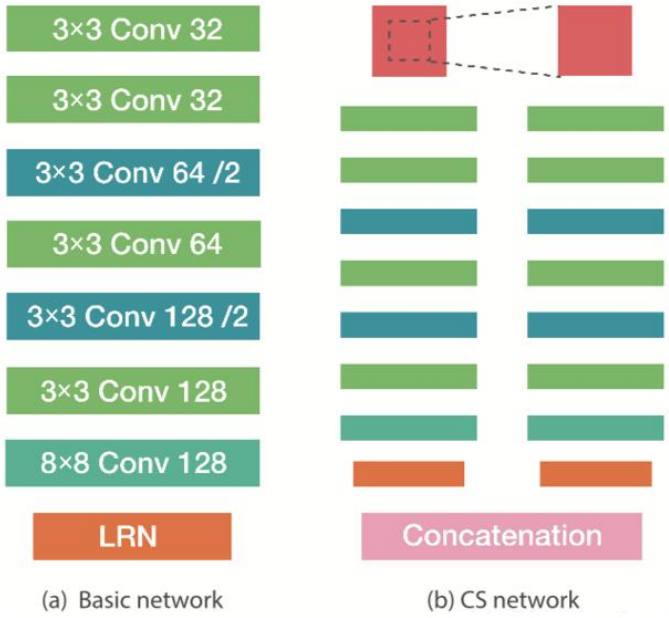

图 3.1 网络结构

3.3 训练数据生成

与 LIFT 类似，我们利用成功的三维重建来自动生成二位对应关系的地面真实数据。首先，利用标准的 SFM 流程进行稀疏重建。在此基础上，将 3D 点云投影到图像平面上来生成二维对应关系。SFM 通常被用来过滤掉大部分不匹配的图像。

通过 SFM 进行验证后，由于图像噪声和错误注册相机的影响，生成的对应关系仍然存在异常值。为了提高数据质量，这里比 LIFT 更进一步，通过使用 3D Delaunay 三角剖分[20]进行可见性检查进行过滤。这种方法在密集立体匹配中被广泛应用于异常值过滤。根据经验，在过滤后将丢弃 30%的 3D 点，只保留高精度的点用于生成地面真实数据。

接下来的步骤，与 LIFT 方法相似，我们通过相似变换对 2D 投影的兴趣区域进行裁剪。其中(x_i^s, y_i^s)，(x_i^t, y_i^t)是输入和输出的正则采样网格，(x, y, σ, θ)是来自 SIFT 检测器的关键点参数（x，y 坐标，尺度和方向）。常数 k 被设置为与 LIFT 相同的值 12，从而产生 12 σ ×12 σ 的图像块。

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \begin{bmatrix} \frac{k\sigma}{2}cos(\theta) & \frac{k\sigma}{2}sin(\theta) & x \\ -\frac{k\sigma}{2}sin(\theta) & \frac{k\sigma'}{2}cos(\theta) & y \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \end{pmatrix} \tag{3.1}$$

3.4 几何相似度

该算法中定义了两种类型的几何相似度，图块相似度和图像相似度，该相似度后续也可用于数据采样与计算损失。图块相似度来衡量具有不同透视变化的图像块对匹配的难度。具体而言，对于给定的图像块对，将其与对应的 3D 轨迹 P 相关联，该轨迹可以由以 Ci 和 Cj 为中心的摄像机看到。接下来，从表面模型中计算在 P 处的顶点法线 Pn。最终得到的图块相似度为：

$$S_{\text{patch}} = s_1 s_2 = g\left(\angle C_i P C_j, \sigma_1\right) g\left(\angle C_i P P_n - \angle C_j P P_n, \sigma_2\right) \tag{3.2}$$

其中，s1 衡量从 3D 轨迹到两条视线之间的夹角$\left(\angle C_i P C_j\right)$，而 s2 衡量从视线到 3D 轨迹上的顶点法线之间的入射角差异$\left(\angle C_i P P_n, \angle C_j P P_n\right)$。

角度度量被定义为

$$g(\alpha, \sigma) = exp\left(-\frac{\alpha^2}{2\sigma^2}\right) \tag{3.3}$$

根据经验，我们选择了σ_1 = 15 度和σ_2 = 20 度。

基于上面所介绍的图像块相似度，这里将图像相似度S_{image}定义为一个图像对的平均图块相似度。图像相似度用于衡量匹配图像对的难度，可以理解为透视变化的度量。

3.5 批次构建

产生有意义的损失以供训练的图像块对通常被称为“有用”的数据。但是由于图像块对的数量非常大，因此很难对这些数据进行有效地采样。之前的工作 L2-Net 和 HardNet 的方法是从整个图块库里面随机选取训练批次。这就造成其中非匹配关系本身的区别度就很大，因此网络可以很轻松地对其进行区分，这就造成了网络学习效率不高的问题。而本文中的图块选取自同一对图像，它们之间具有更高的的相似度，从而进行匹配的难度系数也就更高。示例如下（左图为之前方法的图块图像示例，右图是新方法的图块图像示例）：

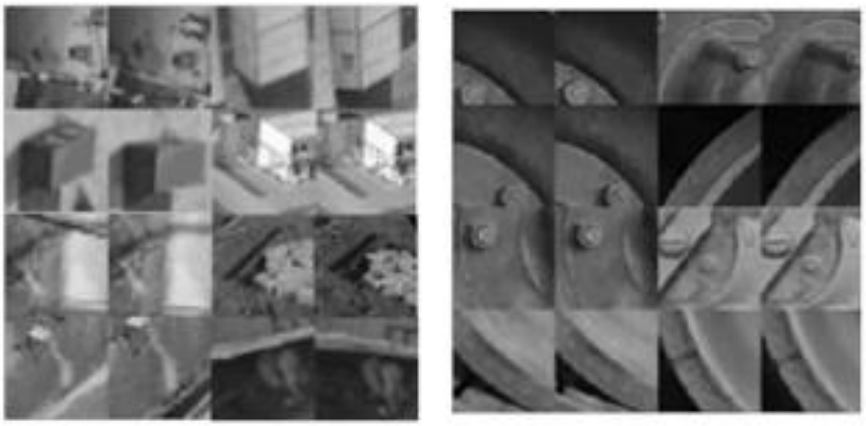

图 3.2 批次示例

3.6 损失函数

L2-Net 和 HardNet 中使用的结构化损失基本上适用于以往方法中构建的批量样本。特别是 HardNet 中基于“批量中最难的”策略和距离边界的公式表明比 L2-Net 中的对数似然公式更有效。然而，当将 HardNet 的损失应用于新方法的批量数据时，却观察到连续的过拟合，由于“批量中最难”这个策略的约束太强。在这种策略中，损失是在产生最大损失的数据样本上计算的，并且设置了一个具有较大数值（HardNet 中为 1.0）的边界，以将非匹配对与匹配对分开。在我们的批量

数据中，已经有效地采样了通常在视觉上相似的“困难”数据，因此强制使用数值较大的边界是不可行的，并且会使学习中止。一个简单的解决方案是减小边界值，但是在本次实验中性能显著下降。

为了避免上述限制并更好地利用批量数据，该算法提出以下损失公式。首先，为一个匹配集计算结构化损失。对于所有(x_i, x_i^+)在匹配集 X 上计算的规范化特征F_1，$F_2 \in R^{N_1 \times 128}$，通过 S = $F_1 F_2^T$得到余弦相似度矩阵

$$E_1 = \frac{1}{N_1（N_1-1)} \sum_{i,\ j} \left(max\left(0, l_{i,j} - l_{i,i}\right) + max\left(0, l_{i,j} - l_{j,j}\right)\right) \quad （3.4）$$

其中$l_{i,j}$是 L 中的元素，$\alpha \in (0，1)$是距离比率。最后，对每个匹配集上的损失取平均值，得到一个训练批次的最终损失，称为结构损失。

虽然结构损失E_1确保匹配的图块对与非匹配的图块对远离，但它没有明确地鼓励匹配的图块对在其度量空间中靠近。为了克服这个问题，根据上文中定义的图块相似性自适应地设置边界，作为最大化正相似性的软约束。称为几何损失，并 将其表示为：

$$E_2 = \sum_i \max\left(0, \beta - s_{i.i}\right), \beta = \begin{cases} 0.7\ s_{patch} \geq 0.5 \\ 0.5\ 0.2 \leq s_{patch} < 0.5 \\ 0.2\ otherwise \end{cases} \quad (3.5)$$

其中β是自适应边界，$s_{i,i}$是 S 中的元素，即图块对(x_i, x_i^+)的余弦相似度，而S_{patch}是(x_i, x_i^+)的图块相似度。我们将$E_1 + \lambda E_2$用作最终损失，并根据经验将α和λ的数值分别设置为 0.4 和 0.2。

3.7 本章总结

介绍了 Geodesc 算法的研究背景与主要思路。详细介绍了该算法的实现流程和算法细节。即根据新的损失函数，通过深度学习网络结构训练训练得到 Geodesc 模型。

第四章 实验与结果分析

4.1 实验环境

本次实验环境为 Windows10，内存 12g，1660ti 显卡，tensorflow 版本 1.14，python 版本为 3.6。

4.2 数据集

4.2.1 训练数据集

本次实验所使用的训练数据集为 GL3D(Geometric Learning with 3D Reconstruction)数据集。GL3D 是一个大规模数据库，该数据集的创建旨在解决 3D 重建问题以及几何相关学习问题。该数据集中包含了 378 个不同场景的 90590 张高分辨率图像。每个场景包含 50 到 1000 张图像，具有大的几何重叠度，覆盖了由多个尺度和角度的无人机拍摄的城市、乡村地区或风景区。它还包含小物体以丰富数据多样性。GL3D 提供了丰富的 3D 上下文信息，例如特征点与轨迹的对应关系、相机位姿、点云数据和网格模型。

图 4.1 展示了 GL3D 数据集中不同类型的场景，右侧为生成训练样本的网格型。

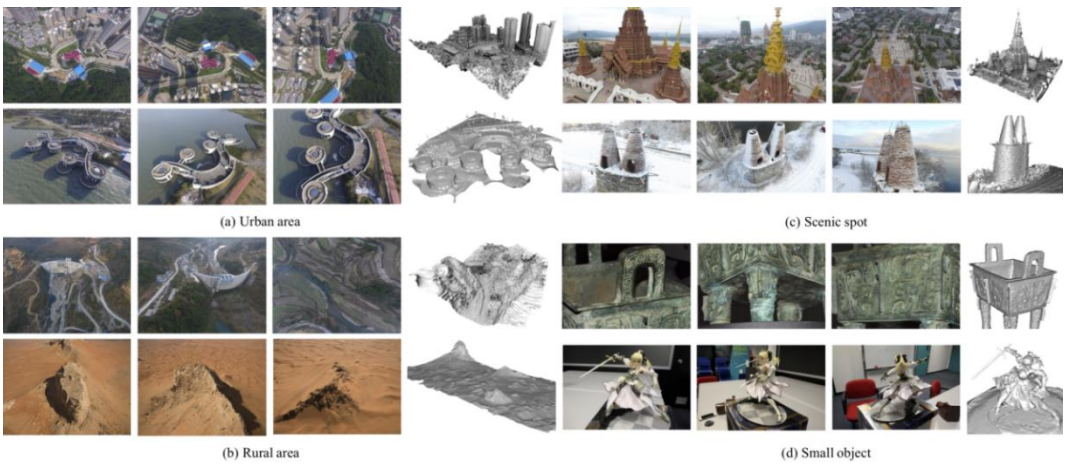

图 4.1 GL3D 数据集示意图

4.2.2 预测数据集

本次实验采用的数据集（a）与数据集（b）分别为使用无人机在城市与郊区拍摄的图像。

如图 4.2 所示，通过在城市区域进行数据采集得到数据集（a）。采集区域内包含运动场、住宅建筑、购物中心等。该数据集的采集基于经典的五视角倾斜摄影测量系统，五台相机中的四台以 45 度倾角同步采集信息，另外一台相机垂直拍摄地面。使用的相机型号为索尼 NEX-7，所有影像的尺寸均为 6000×4000 像素。无人机的飞行高度为 175m，记录图像总数为 750 张。

通过在城市郊区进行数据采集得到数据集（b）。区域内包含一些交叉的铁路轨道。使用的相机型号为一台 Sony RX1R 相机，所有影像的尺寸均为尺寸为 6000×4000 像素。无人机的飞行高度为在 165 米，采集图像总数为 296 张。

图 4.2 左图为预测数据集（a）示意图 右图预测数据集（b）示意图

4.2.3 评价指标

特征点数目：使用对应算法在图像上提取出的特征点数量

特征点匹配准确率：正确匹配的特征点总数占全部特征点总数的比率

4.3 实验结果

在本次实验中，使用 Geodesc 算法与 SIFT 算法分别对同一对图像进行特征点提取与特征图匹配，计算特征点数目并对特征点匹配准确率进行计算。

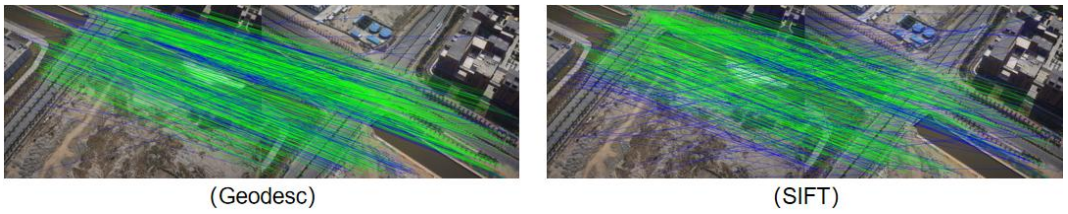

图 4.3 无人机影像匹配结果

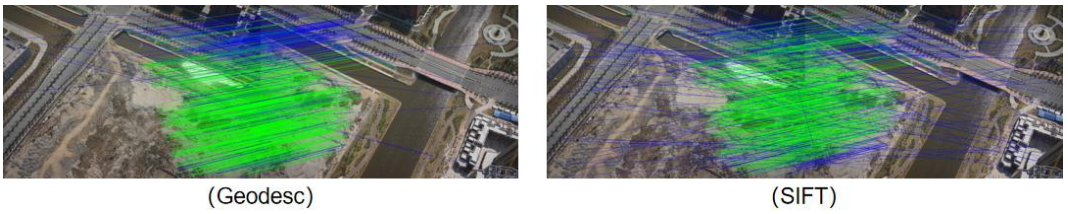

图 4.4 无人机影像匹配结果

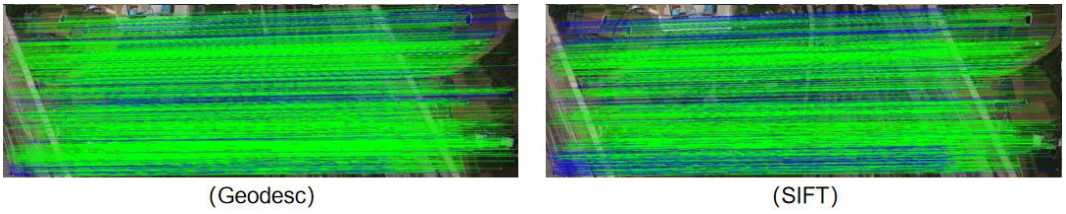

图 4.5 无人机影像匹配结果

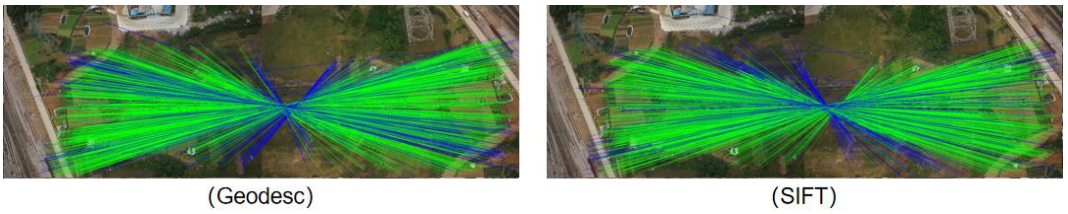

图 4.6 无人机影像匹配结果

如表 4.1 所示，从提取的特征点数量上看，Geodesc 算法相对于 SIFT 算法提取出了更多的特征点。从特征匹配的准确度上看，尽管对图 4.6 中 Geodesc 算法特征匹配的准确度略低于 SIFT 算法，但在其他图像的实验结果中，Geodesc 算法的特征匹配准确度均优于 SIFT 算法。综上所述，我们可以得出结论，本文所研究的 Geodesc 方法相较于传统 SIFT 方法有更好的特征点提取效果。

表 4.1 无人机影像匹配结果量化

	图 4.3 特征点数 [个]	图 4.3 准确度 [%]	图 4.4 特征点数 [个]	图 4.4 准确度 [%]	图 4.5 特征点数 [个]	图 4.5 准确度 [%]	图 4.6 特征点数 [个]	图 4.6 准确度 [%]
Geodesc	946	72.6	1120	79.8	1524	88.1	1294	80.6
SIFT	696	65.8	923	74.3	1168	82.4	1192	82.1

各指标含义：特征点数量，使用对应算法在图像上提取出的特征点数量；

准确度：正确匹配的特征点总数占全部特征点总数的比率；

4.4 本章总结

本章首先介绍了实验相关的其他细节，随后介绍了实验所用的两个数据集，接下来介绍了本次实验的实验结果，分别包括 SIFT 算法与 Geodesc 算法进行特征点匹配的对比与量化结果。

第五章 总结与展望

5.1 研究总结

融合深度学习特征的特征描述符相对于传统方法表现出了更好的结果，但应用于基于图像的三维重建时效果仍然不佳。考虑到传统描述符以及已有的融合深度学习特征的描述符的一些缺点，本文研究了另一种融合深度学习特征的特征描述符，Geodesc，总结如下：

（1）研究总结。对国内外研究人员于图像特征点提取领域的研究进行了学习与总结。学习并总结了传统的 SIFT 特征点提取方法以及融合深度学习特征的 HardNet，学习了标准 SFM 三维建模方法。

（2）基于深度学习的图像特征点描述符 Geodesc 方法研究。研究了 Geodesc 方法在数据处理，批次构建，损失函数等处的研究与创新。并使用无人机拍摄影像，测试了该方法，并将测试结果与 SIFT 方法进行了对比。结果表明，在特征点提取与匹配时 Geodesc 算法表现更好。

5.2 研究展望

针对现有工作所存在的不足,本文在已完成的工作基础上进行了如下展望：

（1）增加与其他特征匹配方法的结果对照，进一步展示该方法的优劣之处。在使用 Geodesc 模型进行图像特征点匹配时，由于时间有限，本文仅将 Geodesc 模型的实验结果与使用 SIFT 算法所得的实验结果进行了对比。在未来的研究中，将增加对照的算法。如 HardNet，L2-Net 等。

（2）增加预测数据。本文中仅选取了部分无人机影像进行预测，为确保实验结果的准确性，将增加预测数据。另外，Geodesc 算法提取的特征维数较高，可能会导致特征匹配的复杂度较高，后续将使用特征降维等技术来解决这一问题。

致谢

四年时光，匆匆而过。纵使一场疫情把大学生活拆的支离破碎，这四年依然美好难忘。一路上，在多位老师的引领教导下，在家人朋友的陪伴支持下，我走过人生最灿烂青春的一段时光，步入人生的下一阶段。

感谢我的导师姜三副教授。从论文选题，到论文撰写，姜老师对学生认真负责，面对学生的问题更是孜孜不倦、循循善诱，在多次与姜老师的交流中，我收获了非常多宝贵的经验与建议。也正是姜老师的指引，使我不再感到迷茫，在学习生活中充满了动力与信心。姜老师细致严谨的工作态度更是令我自惭形愧，诚知自己还有许多要学习改进的地方。

感谢我的父母。父母在我幼时的严格，使我谦虚谨慎，在我成人后的包容，使我自由勇敢。养育之恩，无以回报，只愿父母身体健康。

感谢我的同学朋友。与我分享快乐，帮我度过难关。他们的信任与支持多次帮助我走出低谷。

感谢计算机学院的各位老师以及辅导员。各位老师在授课时倾囊相授，使我受益良多。辅导员的关心与照顾也帮我解决了很多问题。

最后由衷地感谢在场的各位专家评委，请各位老师批评指正！

参考文献

[1] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60: 91-110.

[2] Luo Z, Shen T, Zhou L, et al. Geodesc: Learning local descriptors by integrating geometry constraints[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 168-183.

[3] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features[J]. Lecture notes in computer science, 2006, 3951: 404-417.

[4] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]//2011 International conference on computer vision. Ieee, 2011: 2564-2571.

[5] Rosten E, Drummond T. Machine learning for high-speed corner detection[C]//Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9. Springer Berlin Heidelberg, 2006: 430-443.

[6] Calonder M, Lepetit V, Strecha C, et al. Brief: Binary robust independent elementary features[C]//Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11. Springer Berlin Heidelberg, 2010: 778-792.

[7] Yi K M, Trulls E, Lepetit V, et al. Lift: Learned invariant feature transform [C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14. Springer International Publishing, 2016: 467-483.

[8] DeTone D, Malisiewicz T, Rabinovich A. Superpoint: Self-supervised interest point detection and description[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018: 224-236.

[9] Dusmanu M, Rocco I, Pajdla T, et al. D2-net: A trainable cnn for joint description and detection of local features[C]//Proceedings of the ieee/cvf conference on computer vision and pattern recognition. 2019: 8092-8101.

[10] Revaud J, Weinzaepfel P, Harchaoui Z, et al. Deepmatching: Hierarchical deformable dense matching[J]. International Journal of Computer Vision, 2016, 120: 300-323.

[11] Han X, Leung T, Jia Y, et al. Matchnet: Unifying feature and metric learning for patch-based matching[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3279-3286.

[12] Sarlin P E, DeTone D, Malisiewicz T, et al. Superglue: Learning feature matching with graph neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 4938-4947.

[13] 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.

[14] Zhu S, Zhang R, Zhou L, et al. Very large-scale global sfm by distributed motion averaging[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4568-4577.

[15] Shen T, Zhu S, Fang T, et al. Graph-based consistent matching for structure-from-motion[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14. Springer International Publishing, 2016: 139-155.

[16] Zhang R, Zhu S, Fang T, et al. Distributed very large scale bundle adjustment by global camera consensus[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 29-38.

[17] Zhu S, Fang T, Xiao J, et al. Local readjustment for high-resolution 3d reconstruction[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 3938-3945.

[18] Tian Y, Fan B, Wu F. L2-net: Deep learning of discriminative patch descriptor in euclidean space[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 661-669.

[19] Schonberger J L, Frahm J M. Structure-from-motion revisited[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 4104-4113.

[20] Labatut P, Pons J P, Keriven R. Efficient multi-view reconstruction of large-scale scenes using interest points, delaunay triangulation and graph cuts[C]//2007 IEEE 11th international conference on computer vision. IEEE, 2007: 1-8.

[21] 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.

[22] Triggs B, McLauchlan P F, Hartley R I, et al. Bundle adjustment—a modern synthesis[C]//Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms Corfu, Greece, September 21–22, 1999 Proceedings. Springer Berlin Heidelberg, 2000: 298-372.