

# 适用于无人驾驶车辆的地图构建与道路提取算法研究

于华超

2018 年 3 月



中图分类号：TP391.41

UDC 分类号 : 621.3

# 适用于无人驾驶车辆的地图构建与道路提取算法研究

作 者 姓 名	<u>于华超</u>
学 院 名 称	<u>自动化学院</u>
指 导 教 师	<u>王美玲</u>
答 辩 委 员 会 主 席	<u>孙长胜教授</u>
申 请 学 位 级 别	<u>工学硕士</u>
学 科 专 业	<u>控制科学与工程</u>
学 位 授 予 单 位	<u>北京理工大学</u>
论 文 答 辩 时 间	<u>2018 年 3 月</u>

# **Research on map building and road extraction algorithm for intelligent vehicles**

Candidate Name: Hao Li

School or Department: School of Automation

Faculty Mentor: Yi Yang

Chair, Thesis Committee: Prof. Changhai Sun

Degree Applied:

Major: Control Science and Engineering

Degree by: Beijing Institute of Technology

The Date of Defence: March, 2018

适用于无人驾驶车辆的地图构建与道路提取算法研究

北京理工大学

## 研究成果声明

本人郑重声明：所提交的学位论文是我本人在指导教师的指导下进行的研究工作获得的研究成果。尽我所知，文中除特别标注和致谢的地方外，学位论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京理工大学或其它教育机构的学位或证书所使用过的材料。与我一同工作的合作者对此研究工作所做的任何贡献均已在学位论文中作了明确的说明并表示了谢意。

特此申明。

签名：

日期：

## 关于学位论文使用权的说明

本人完全了解北京理工大学有关保管、使用学位论文的规定，其中包括：①学校有权保管、并向有关部门送交学位论文的原件与复印件；②学校可以采用影印、缩印或其它复制手段复制并保存学位论文；③学校可允许学位论文被查阅或借阅；④学校可以学术交流为目的，复制赠送和交换学位论文；⑤学校可以公布学位论文的全部或部分内容（保密学位论文在解密后遵守此规定）。

签名：

日期：

导师签名：

日期：

## 摘要

这里是摘要。

关键词:

## **Abstract**

This is abstract.

**Keywords:**

## 目录

第 1 章 绪论 ······	1
1.1 研究背景与意义 ······	1
1.2 国内外研究现状 ······	1
1.2.1 地图构建方法研究 ······	1
1.2.2 道路提取方法研究 ······	1
1.3 主要研究内容 ······	1
1.3.1 二维与三维地图构建 ······	1
1.3.2 二维图像中的道路提取 ······	1
1.4 本章小结 ······	1
第 2 章 基于全景图像的高精度地图构建 ······	2
2.1 引言 ······	2
2.2 全景图像与定位信息采集 ······	2
2.3 全景图像处理 ······	2
2.4 地图的构建与使用 ······	6
2.4.1 图像拼接 ······	6
2.4.2 基于GIS数据库生成无人车行驶地图 ······	7
2.5 本章小结 ······	8
第 3 章 基于无人机航拍图像的三维构建 ······	9
3.1 引言 ······	9
3.2 相机模型 ······	9
3.3 相关坐标系 ······	11
3.4 特征点提取 ······	15
3.5 特征点匹配 ······	19
3.6 多视图重建 ······	22
3.6.1 两视图重建 ······	23
3.6.2 2D-3D位姿求解 ······	23
3.6.3 三角定位法 ······	23
3.7 三维重建中的优化 ······	23
3.7.1 光束平差法 ······	23
3.7.2 最小化相机中心位置误差 ······	23

3.8 稀疏电云的渲染 ······	23
3.8.1 点云稠密化 ······	23
3.8.2 点云网格渲染（三角剖分） ······	23
3.9 本章小结 ······	23
第 4 章 二维图像中的道路提取 ······	24
4.1 引言 ······	24
4.2 图像二值化 ······	24
4.3 非道路区域移除 ······	24
4.3.1 开闭运算 ······	24
4.3.2 轮廓提取 ······	24
4.4 拓扑处理 ······	24
4.4.1 拓扑细化 ······	24
4.4.2 去除小枝丫 ······	24
4.5 本章小结 ······	24
第 5 章 实验结果分析 ······	25
5.1 引言 ······	25
5.2 全景图像构建地图实验结果 ······	25
5.2.1 全景图像地图实验平台搭建 ······	25
5.2.2 全景图像地图构建效果及精度分析 ······	25
5.3 地理空间三维重建实验结果 ······	25
5.3.1 地理空间三维重建实验平台搭建 ······	25
5.3.2 三维重建效果及精度分析 ······	25
5.4 道路提取精度分析 ······	25
5.5 本章小结 ······	25
总结与展望 ······	28
参考文献 ······	29
攻读硕士学位期间发表论文与研究成果清单 ······	30

## 第1章 绪论

**1.1 研究背景与意义**

**1.2 国内外研究现状**

**1.2.1 地图构建方法研究**

**1.2.2 道路提取方法研究**

**1.3 主要研究内容**

**1.3.1 二维与三维地图构建**

**1.3.2 二维图像中的道路提取**

**1.4 本章小结**

## 第2章 基于全景图像的高精度地图构建

### 2.1 引言

全景图像（panorama）是指水平角度包含完整 $360^{\circ}$ 的相机得到的二维平面图像。相对于普通单目相机，全景相机可以完整的呈现无人车周围环境。在本章节中，讨论使用全景相机构建车道级别的用于无人车行驶的高精度地图的方法。首先介绍了全景图像与定位信息的采集，而后介绍了全景相机的标定与全景图像预处理，最后阐述了如何通过全景图像构建大范围的高精度地图的方法。

### 2.2 全景图像与定位信息采集

为了同时采集全景图像信息与高精度定位导航系统信息，本文采用将车辆的定位信息编码为大小一定的QR code (Quick-Response code)，放置于每帧全景图像左上角位置，完成系统的信息采集任务。在此介绍一下二维码的编码与解码过程。由于全景图像的帧率为 $7\text{Hz}$ ，惯性导航系统的帧率为 $20\text{Hz}$ ，考虑拼接图像的精度，在没有使用硬件触发的条件下，本文采用最近邻原则，即当接受到一帧全景图像时，编码离当前时刻最近的惯性导航定位信息。由于编码的信息有时间，经纬度，方位角，偏航角与俯仰角6个整数，所以需要的编码信息量很小，本文只采用 $25 \times 25$ 像素的二维码编码定位信息。

QR code作为一种非常流行的矩阵式二维码，具有空间利用率高，存储数据量大，解码速度快，可以包含字符，数字等不同内容等优点。如图 2.1所示，二维码可分为两部分：功能区和编码区。功能区包含用于二维码定位的位置探测区域，用于校正二维码位置的校正图形区域等，在编码区中，包含数据和纠错码等。二维码的纠错级别越高，则实际存储的数据信息就越少。按照QR code的规则，可以很方便的为定位信息编码为固定像素大小的二维码，进而将该大小的区域放置在每帧全景图像的左上角，完成信息的采集过程。当回放该视频时，可以很容易地对大小固定的编码区进行解码，读取到每帧全景图对应的车辆当前的位置与角度信息等，为图像拼接做准备。

### 2.3 全景图像处理

Ladybug全景相机具有6个独立的摄像头，如图 2.2，可以在球形坐标系下将6个相机图像实时合并为一张全景图，呈现在二维平面空间中。ladybug相机提供 $360^{\circ}$

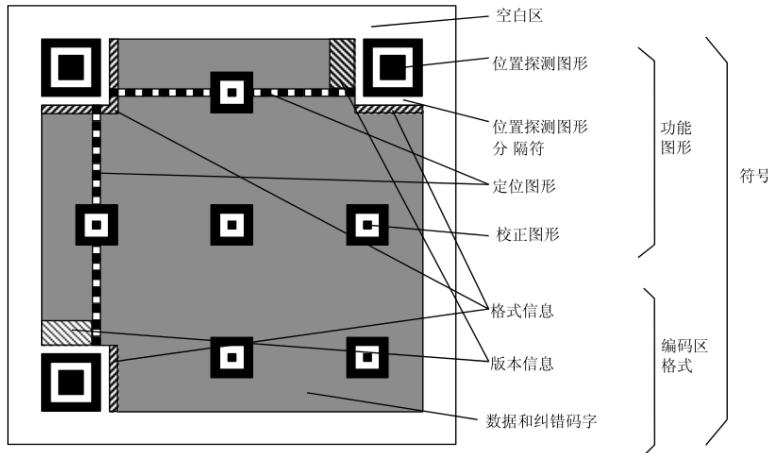


图 2.1 QR code示意图

视频串流功能，可以覆盖90%的可视球面具有图像采集、处理和校正功能。在得到Ladybug相机的全景图后，使用逆透视投影（IPM）将全景图像投影成俯视图。而后将车辆的高精度定位信息编码为二维码，放置于对应的逆透视投影得到的图片的左下角，从而完成图像与对应车体位置的采集与处理，方便后续构建大范围地图。



图 2.2 Ladybug全景相机

由于LadyBug成像在球面坐标系中描述，IPM投影需要在车体坐标系中描述，地图拼接需要在世界坐标系中完成，在此简单介绍全景图像处理过程涉及的相关坐标系的变换关系。如图 2.3所示，代表全景相机的球面坐标系与车体坐标系的关系，其中在车体坐标系中，绕自身x, y和z旋转，分别代表俯仰角，横滚角和方位角，使用道路方向形象地代表车辆的行驶方向；在全景坐标系中，球面狭缝与全景图分块成对应关系，如图中虚线箭头所示，得到的全景图的长宽，如图 2.4所示，图片长

宽分别代表相对于车体的方位角 $\mu$ 与俯仰角 $\nu$ 。为了得到全景图像的IPM俯视图，需要标定全景相机相对于车辆的外参：俯仰角 $\theta$ ，横滚角 $\phi$ 和方位角 $\varphi$ ，以及Ladybug相机光心与地面的高度 $h$ ，同时假定车辆周围的环境与地面高度一致。现定义车体坐标系为 $\{F_\omega\} = \{X_\omega, Y_\omega, Z_\omega\}$ ，坐标原点位于车辆后轴中点；定义全景相机坐标系为 $\{F_c\} = \{X_c, Y_c, Z_c\}$ ；定义全景图像坐标为 $\{F_i\} = \{\mu, \nu\}$ 。如图 2.3所示，车体坐标系中任一点与全景相机坐标系中的点的对应关系为：

$$P_c = R_\phi R_\theta R_\varphi^T (P_\omega + T) \quad (2.1)$$

其中

$$R_\phi = \begin{bmatrix} \cos\phi & 0 & -\sin\phi \\ 0 & 1 & 0 \\ \sin\phi & 0 & \cos\phi \end{bmatrix} \quad (2.2)$$

$$R_\theta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \quad (2.3)$$

$$R_\varphi = \begin{bmatrix} \cos(\varphi - \frac{\pi}{2}) & \sin(\varphi - \frac{\pi}{2}) & 0 \\ -\sin(\varphi - \frac{\pi}{2}) & \cos(\varphi - \frac{\pi}{2}) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.4)$$

$$T = \begin{bmatrix} 0 \\ 0 \\ -h \end{bmatrix} \quad (2.5)$$

进而可以得到全景图像坐标系中的坐标 $P_i = [\mu_i, \nu_i]^T$ ：

$$P_i = \begin{bmatrix} \arctan \frac{y_c}{x_c} \\ \arccos \frac{z_c}{\|P_c\|} \end{bmatrix} \quad (2.6)$$

其中 $\|P_c\| = \sqrt{X_{P_c}^2 + Y_{P_c}^2 + Z_{P_c}^2}$ ，即 $P_c$ 的二范数。通过以上算法可以将全景图像（图 2.4）变换得到车体坐标系下的IPM图，显示效果如图 2.5所示

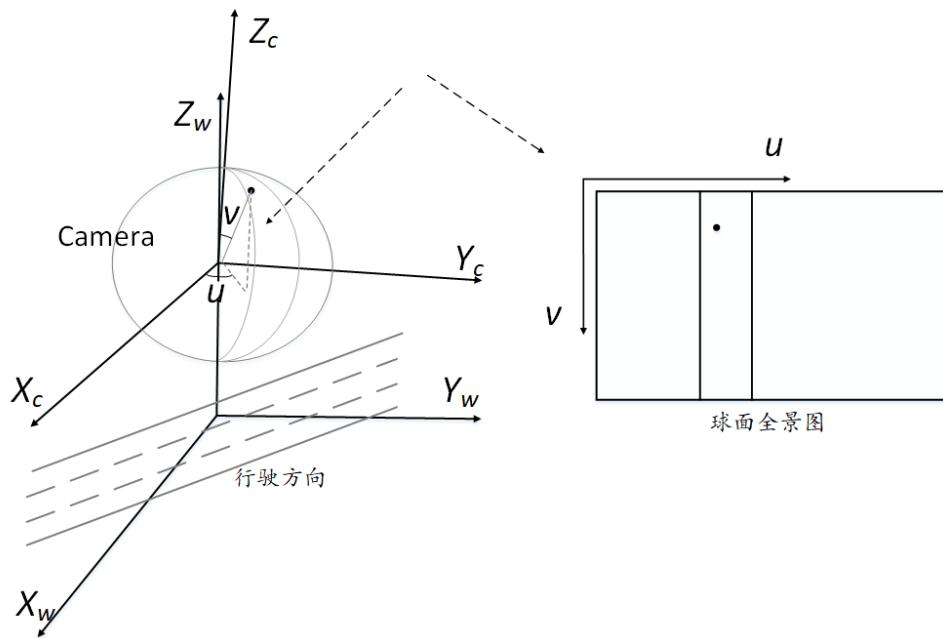


图 2.3 逆透视坐标变换关系



图 2.4 基于球面坐标的全景图

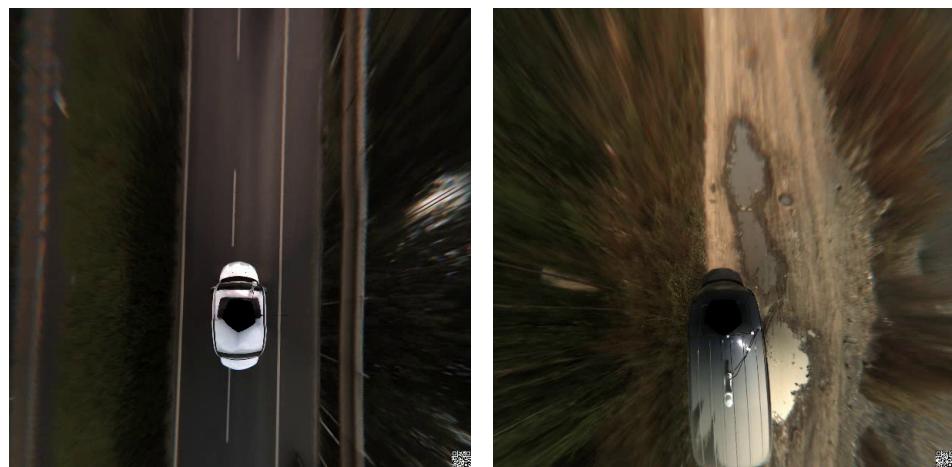


图 2.5 逆透视变换结果

## 2.4 地图的构建与使用

使用以上的方法可以得到全景图像的逆透视投影，以及每张投影图对应的此时车辆位置与方位等信息。进而我们可以得到大范围地图信息，制作大范围行驶地图，为无人车规划与感知提供先验信息。

### 2.4.1 图像拼接

图像拼接的历史可以追溯到上个世纪，使用的方法也不尽相同，总体上可以分为两类，一类是基于特征匹配的图像拼接方法，另一类是基于图片相互间位置关系的拼接方法。由于本文采用的惯性组合导航系统，在无遮挡，多路径效应较弱的空旷场地，实验精度达到 $5cm$ 以内，所以在此本文采用基于图片之间相互位置关系的方法。具体就是将视频流中每帧逆透视投影图片按照二维码解算的航向旋转一定角度，然后按照解算的位置平移一段距离，对所有视频按照上述方法处理，最终得到完整的城区地图。

对于每帧IPM俯视图，车身为无效信息且占据图像大量位置，对地图拼接造成干扰，所以在拼接地图前，需要将车身移除。由于全景图变换为IPM投影图时，将IPM俯视图变换到车体坐标系中，所以任何一张IPM俯视图的车辆位置与方位是固定的：车辆后轴中心位于图片横轴的中心，纵轴靠下的 $\frac{1}{3}$ 位置处，航向朝上。如图2.6所示。据此设计一个mask，在拼接过程中将车辆去掉，只拼接剩下的像素不为零的部分。

在拼接过程中，因为所需拼接的地理空间过大，图片过多，而内存限制了单张图片的大小，所以本文设计了区号标记的方法。具体做法为：将视频流中图片按照位置和航向拼接为多张 $28000 \times 7000$ 像素的图片，如图2.6中橙色方框所示，由于每个像素代表实际地理空间的大小为 $0.06m$ ，即每张拼接后的图片代表的物理空间大小为 $1680m \times 420m$ 。设置整个地图的坐标原点（图2.6红色圆点所示）为某一固定经纬度坐标，IPM图对应的经纬度与该点对比后，计算出该帧IPM图应该放置于哪个区，而后可以拼接成多张橙色大小的图片。一个橙色方框代表一个区域，对应一个区号，这种存在重叠区域的区号法则有效的解决了图像拼接过程中出现的边缘缺失的问题，所以即使橙色方框代表的区域出现边缘缺失现象，由于重叠区域的存在可以有效解决该问题。最后只需要保留图2.6的绿色框表示的区域和该区域的区号，这样就可以按照区号和图2.6的样式将拼接的图片在Photoshop中拼接为更大的图片，为后

续GIS数据库中构建无人车底图服务。

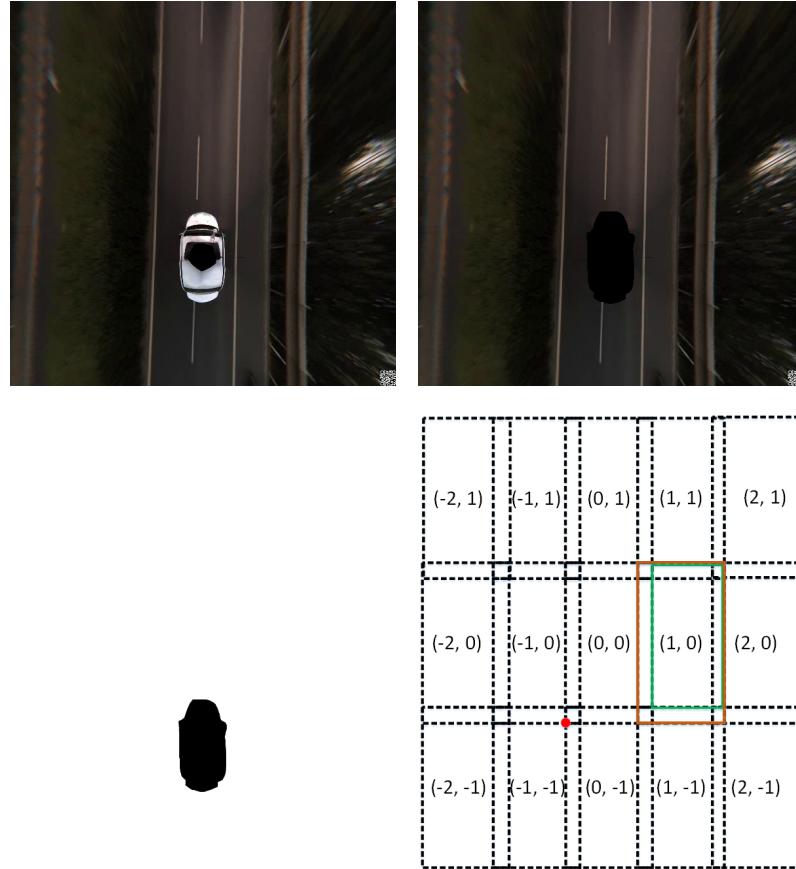


图 2.6 去掉车辆

## 2.4.2 基于GIS数据库生成无人车行驶地图

在得到数张Photoshop拼接的更大的图片的同时，按照图 2.6的分区原则也很容易得到每张图片四个顶点的地理坐标。依托Supermap(GIS)软件，可以很容易完成金字塔地图的生成，作为GIS数据库的一个图层。如图 2.7所示。这样构建的地图并不能直接用于无人车导航或规划，我们将得到的底图作为GIS数据库的一个图层，根据底图可以得到车道级别的道路拓扑结构，该拓扑作为GIS数据库的另一个图层，如图 2.7中的绿色线段，在线段与线段的节点上添加各种属性，例如车道数量，限速，红路灯和路口等，即可实现全局路径规划等功能，这些功能不是本文讨论的内容。

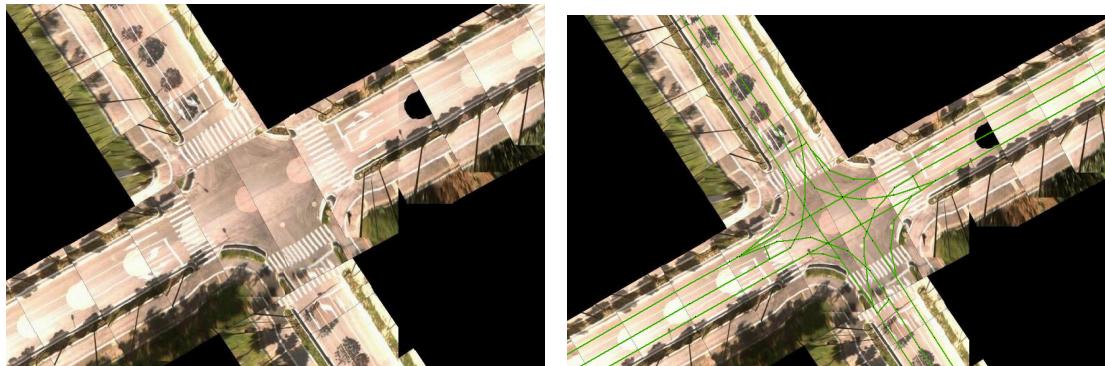


图 2.7 拼接的路口展示

## 2.5 本章小结

本章介绍了如何通过Ladybug全景相机生成的球面全景图与高精度定位导航组合，构建车道级别地图的过程。根据Ladybug与车体的固定位置关系，将全景图像旋转与平移转换到车体坐标系中，并完成逆透视投影的过程，并将每个IPM 投影的图像通过车辆的高精度定位导航信息，旋转平移拼接为完整的道路地图，最后将地图生成GIS数据库的图层，完成道路拓扑构建与道路属性添加的任务，即可将该地图用于无人驾驶导航与规划等任务中。

## 第3章 基于无人机航拍图像的三维构建

### 3.1 引言

大部分无人车辆行驶依赖环境地图导航与感知<sup>[1]</sup>，这限制了无人车在未知环境中的行驶。为了解决该问题，当前许多技术利用车载传感器可以构建很大范围的地图，例如SLAM(Simultaneous Localization and Mapping) 和第二章提到的用车载Ladybug构建地图的方法。然而车载传感器仅能获取车辆周边环境信息且易被阻挡，而且在野外或灾害地区等环境中，车辆可通行区域变少。然而无人机拥有比无人车更广阔的视野，不受地面障碍物的影响，可以灵活地从空中俯视地面环境。本章介绍了应用无人机构建三维环境地图，为无人车导航提供先验信息。

三维重建涉及到的内容可以分为三部分：稀疏点云的获取，点云稠密化与点云的网格渲染。其中稀疏点云的获取使用的是Structure from Motion(SfM)，在技术层面上与当前热门的Visual-SLAM(VSLAM)比较类似，只是SfM侧重于三维重建，而SLAM侧重于实时定位，并需要预测下一时刻的位置。SfM可以分为视觉前端与优化后端，前端涉及到的是相机模型，坐标转换，特征提取与匹配以及多视图几何相关的内容，最后得到粗略的相机位姿与3D点坐标，后端涉及到的是优化前端得到的相机位姿与3D点坐标，去除错误的结果等。点云稠密使用的技术称为Multi-View Stereo(MVS)，MVS的方法很多，本文使用的是深度图融合的方法。网格渲染主要用的技术为三角剖分，网格渲染后的三维重建结果比稠密点云存储量小，更易于后续的仿真与处理。

### 3.2 相机模型

数码相机拍摄的过程中，实际上是一个光学成像的过程，这涉及到摄像机最基本的原件——透镜——的成像原理，如图 3.1 所示，这是最基本的透镜成像原理：Z 是物体距透镜光心的距离，简称物距；f 是焦距；b 是相距，即成像平面与透镜光心的距离。三者满足式 3.1

$$\frac{1}{f} = \frac{1}{Z} + \frac{1}{b} \quad (3.1)$$

在机器视觉中，利用摄像机可以将三维场景记录在二维图形上，不同的相机模型导致不同的成像效果，比如第二章中Ladybug使用的球面成像模型，使不同相机的拼

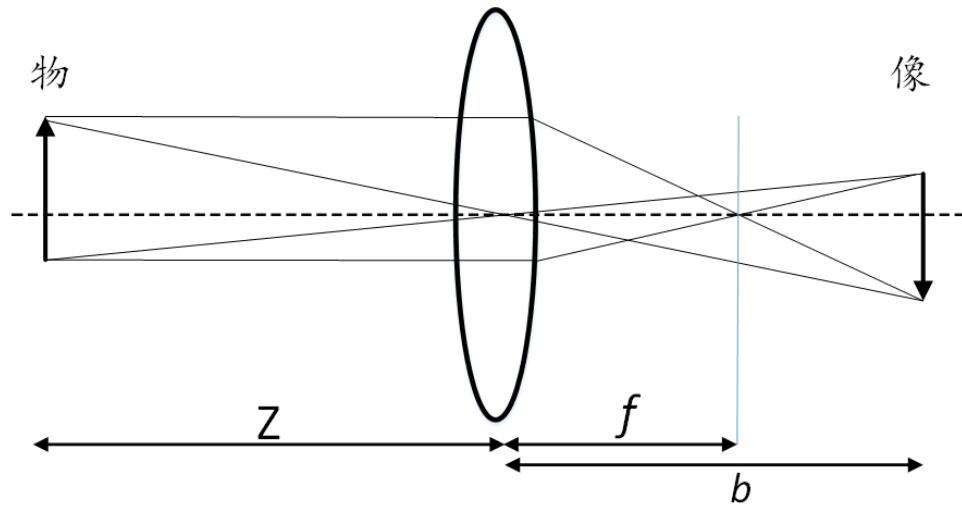


图 3.1 透镜成像原理

接简化为绕光心的旋转。但是常见的相机使用的模型还是小孔成像模型，如图 3.2 所示。数码相机的镜头相当于一个凸透镜，感光元件就处在这个凸透镜的焦点附近，将焦距近似为凸透镜中心到感光元件的距离时就成为小孔成像模型。这可以类比为艺术家画一幅画的过程，将眼睛放在图 3.2 光心 C 处，在物体与眼睛之间放置一个画布（图 3.2 中虚像位置），按照光线直线传播的原则，则物体反射的光线与眼睛的连线交画布一点，如此物体可以在画布上成像。只是相机的成像在光心后面，所以说存在一个虚拟成像平面。

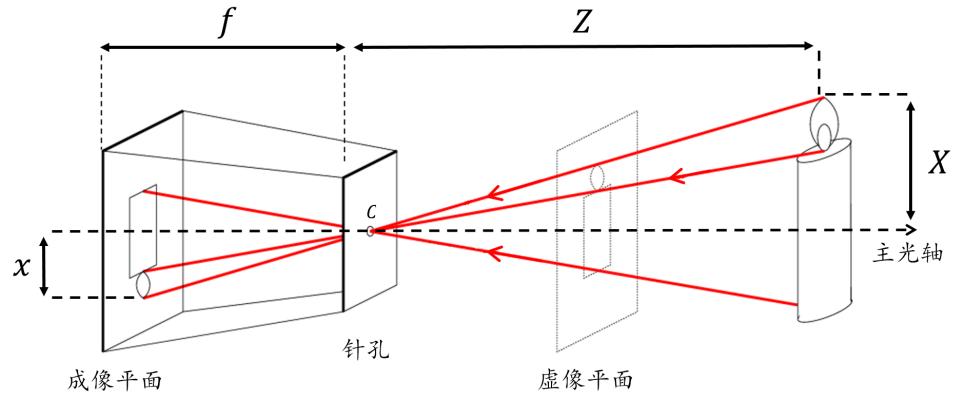


图 3.2 小孔成像模型

在三维空间中，物体反射的光线经过小孔形成倒立二维图像，根据其数学模型，可以得到

$$\frac{-x}{X} = \frac{f}{Z} \quad (3.2)$$

其中,  $f$ 是小孔到成像平面的距离, 即焦距;  $Z$ 是物体距光心的距离, 简称物距;  $x$ 是物体在成像平面的投影长度;  $X$ 是物体实际长度。为了简化以上模型, 用图 3.3 表示小孔成像的等价模型, 其中  $C$  是相机中心,  $P(X, Y, Z)$  为空间中的 3D 点,  $p(x, y, 1)$  表示 3D 点成像在感光元件上的点, 小孔成像近似后, 成像平面与相机中心的距离为焦距  $f$ 。该图也表示了两个坐标系, 分别为图像坐标系和像素坐标系, 在下节中详细介绍。

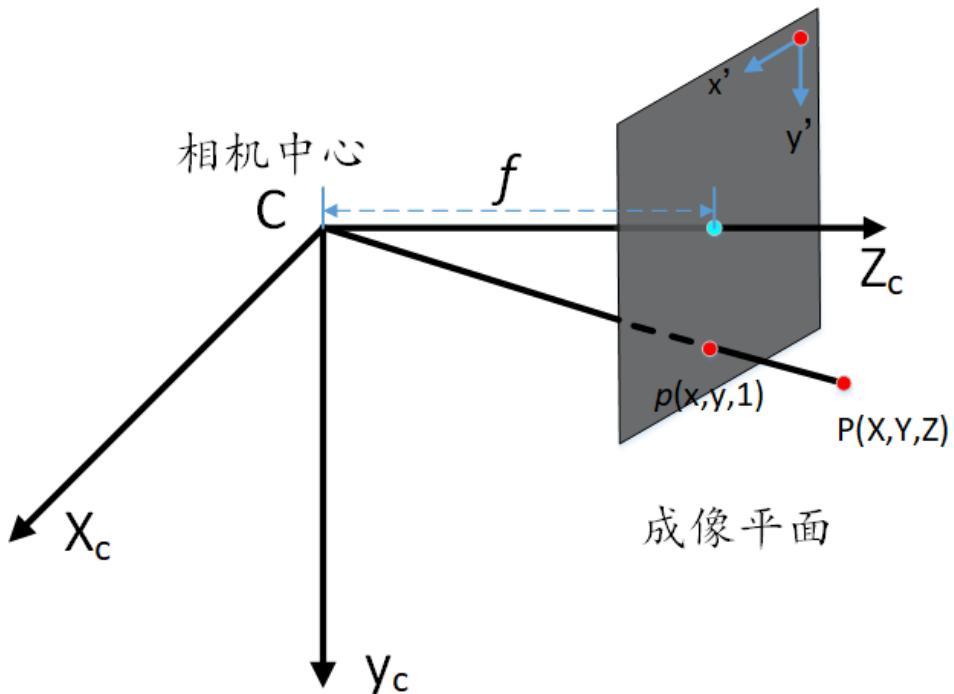


图 3.3 小孔成像等价模型

### 3.3 相关坐标系

从上一节知道, 仅仅相机内部就存在多个坐标系, 那么摄像机在空间中涉及到的坐标系转换问题将在这节详细探讨。相机成像的过程是一个 3D 物体显示在 2D 成像平面的过程, 在此涉及物体, 真实空间与相机三者的位置关。物体在真实空间的位置即是物体在世界坐标系中的位置坐标表达, 在图 3.3 中,  $(X, Y, Z)$  为 3D 点  $P$  在相机坐标系  $(X_c, y_c, Z_c)$  下的坐标表示, 这里就存在两个三维坐标的转换, 即同一个点在世界坐标系与相机坐标系下表达方式的转换关系, 如式 3.3, 该变换可以看做三维坐标系的刚体变换。

$${}^cP = {}^wR_w^cP + {}^cT_w \quad (3.3)$$

其中 $c$ 表示相机坐标系,  $w$ 表示世界坐标系,  ${}^wP$ 表示世界坐标系下 $P$ 点3D坐标,  ${}^cP$ 是相机坐标系下 $P$ 点3D坐标,  ${}^cR_w$ 和 ${}^cT_w$ 分别表示由世界坐标系到相机坐标系的旋转矩阵和平移矩阵。

将世界坐标系下的3D点表示为齐次坐标 $(X, Y, Z, 1)$ , 式3.3可以改写为:

$${}^cP = \begin{bmatrix} {}^cR_w & {}^cT_w \\ 0^{1 \times 3} & 1 \end{bmatrix}_{4 \times 4} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (3.4)$$

其中 ${}^cR_w$ 是 $3 \times 3$ 矩阵, 三个列向量俩俩正交;  ${}^cT_w$ 是 $3 \times 1$ 矩阵, 代表两个坐标系原点连线的向量,  ${}^cP$ 坐标表示为 $(x_c, y_c, z_c)$ 。

从图3.3中可以看到成像平面内的二维坐标系 $x', y'$ , 该二维坐标系称为图像坐标系, 可以推导相机坐标系与图像坐标系(3D到2D)变换的公式表示:

$$x' = f \frac{x}{z} \quad (3.5)$$

$$y' = f \frac{y}{z} \quad (3.6)$$

改为矩阵表达形式为

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \quad (3.7)$$

相机在生成图片的过程中, 由于成像平面由许多CCD感光元件组成, 所以成像的横纵坐标是不连续的, 引入像素的概念, 同时也存在从图像坐标系到像素坐标系(2D到2D)的转换关系:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{s_x} & s & p_x \\ 0 & \frac{1}{s_y} & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \quad (3.8)$$

其中,  $dx$ 是u轴方向单像素的宽度,  $dy$ 是v轴方向单像素的宽度, 当前大部分ccd的像素宽度,  $dx = dy$ , 只有以前老式电视机使用的是矩形ccd, 导致 $dx \neq dy$ 。 $p_x, p_y$ 是主点(图3.3)与图像左边界的距离与 $p_y$ 是主点与图像上边界的距离。 $s$ 是衡量主光轴与成像平面的倾斜程度, 如果主光轴垂直于成像平面, 则 $s = 0$ 。这些参数被称作相机内参, 一般可以通过查询相机手册得到包括焦距在内的相机内参。

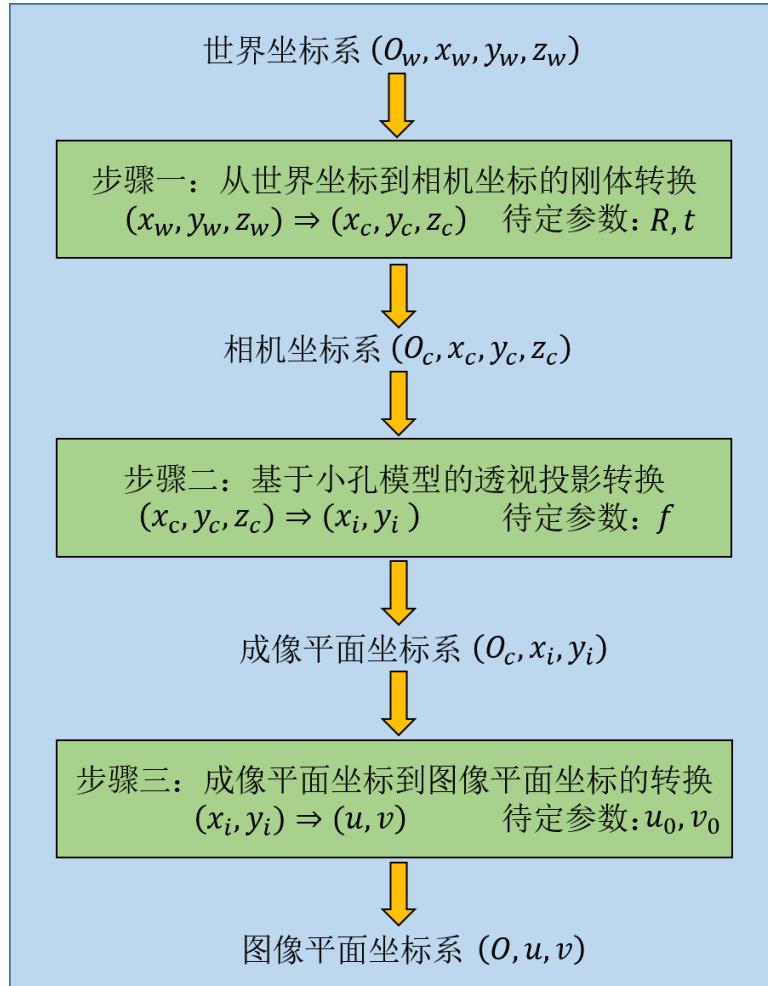


图3.4 坐标转换流程图

至此我们可以得到从世界坐标系到像素坐标系变化过程, 如图3.4和3.5。图3.4过程用公式表达为3.9

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{s_x} & s & p_x \\ 0 & \frac{1}{s_y} & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} {}^cR_w & {}^cT_w \\ 0^{1 \times 3} & 1 \end{bmatrix}_{4 \times 4} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (3.9)$$

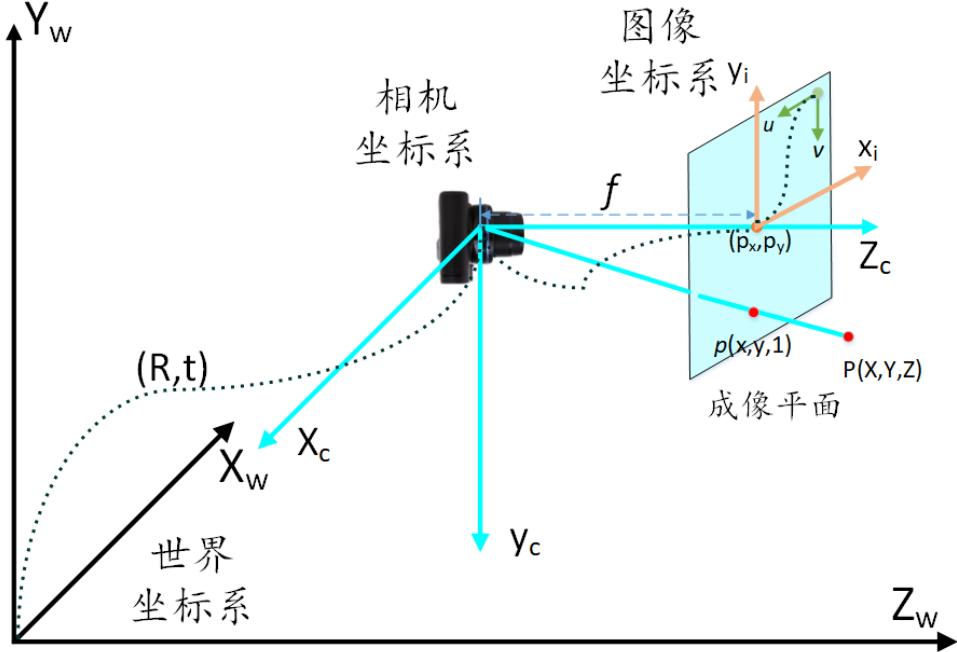


图 3.5 坐标转换关系图

至此我们知道了相机成像过程中某点如何在不同坐标系下转换，从而完成摄像过程。式 3.9 涉及到相机参数  $f, p_x, p_y$  以及空间变换相关的参数（外参  $R, t$ ），实际上在相机成像过程中除了这些参数，我们还会遇到无法线性建模的参数：像 Gopro 等鱼眼镜头中，我们可以看到世界中的直线不再是直线，如图。。。这些弯曲的线有一些特别的特性：他们都是被关于中心扭曲的我们称这种畸变为径向畸变。这意味着像素点是沿着从中心放射的径向成比例扭曲的。为了去除相机的畸变，需要对畸变用多项式进行建模，如式 3.10。

$$u^{dist} = u(1 + k_1 r + k_2 r^2 + k_3 r^3 + \dots) \quad (3.10a)$$

$$v^{dist} = v(1 + k_1 r + k_2 r^2 + k_3 r^3 + \dots) \quad (3.10b)$$

$$\text{where } r^2 = u^2 + v^2 \quad (3.10c)$$

其中， $k_1, k_2, k_3$  是未知参数，需要标定才可以得到，而实际校正径向畸变时一般也只用到 2 个或 3 个参数，如图 3.6 所示，左图为存在径向畸变的图片，右图为经过校正后的图片。



图 3.6 径向畸变的校正

在相机校正的过程中，实际上还需要校正图像的切向失真，这是由于摄像机安装时成像平面与透镜没有平行导致的，如果说径向畸变是坐标点距离中心点的长度放生了变化，那么切向畸变就是坐标点与中心点的水平夹角发生了变化。对切向畸变的建模如式 3.11，一般使用  $p_1, p_2$  两个参数。

$$u^{dist} = u + 2p_1uv + p_2(r^2 + 2u^2) \quad (3.11a)$$

$$v^{dist} = v + 2p_2uv + p_1(r^2 + 2v^2) \quad (3.11b)$$

$$\text{where } r^2 = u^2 + v^2 \quad (3.11c)$$

综上所述，本节详细介绍了相机模型与相机相关坐标系转换关系，以及相机成像的校正等相关内容，这部分属于图像处理的基础，熟悉该部分内容将为下面几节多视图几何相关内容作铺垫。

### 3.4 特征点提取

SfM 首先需要找到图片集中俩俩图片间的几何关系，这样才能估算相机间的运动与成像的三维点坐标。为了找到图片之间的几何关系，现今有许多方法，例如光流法，特征点法等，光流法适用于匀速运送的视频流中，要求帧间运动比较小，从而找出帧间运动关系；特征点法指对图像特征的提取与存储，是图像中比较有代表性的特征。由于本文使用的数据集为无序的航拍图像，特征点法较为适宜使用。特征点提取更通俗的理解为将二维矩阵描述的图像降维为一维特征点描述的图像，突出图像的重点，去除冗余信息的过程。

最简单的图像特征为图像的角点，角点就是图像中线的交点，例如大厦最高处的顶点。两幅图像的角点不会随着图像移动而改变，但是角点会随着离相机的距离太近

而变得无法识别，所以许多研究者设计了精巧而互有优劣的特征检测与描述方法，比较著名的有SIFT(Scale Invariant Feature Transform)<sup>[2]</sup>，SURF<sup>[3]</sup>，ORG<sup>[4]</sup>等。相对于简单的角点，这些特征检测具有以下优点<sup>[5]</sup>：

- 可重复：相同的特征可以在不同的图片中找到
- 可区别：不同的特征有不同的表示
- 高效率：特征点数远远小于像素数

实际上，特征点可以分为局部特征点和全局特征点，像上面举例的三个特征点实际上是局部特征点，局部特征点顾名思义就是基于图像的突出区域明显的区分图像的特征，一般而言，局部特征需要具有旋转不变性，光线明暗的鲁棒性等。因此图像可以被提取到的一系列称为感兴趣区域的特征描述。相反全局特征是将图像表示为一个向量，向量的值为图像的各个方面，如颜色，纹理或形状。例如像区分图像，一些是海洋，一些是森林，一个全局颜色描述子可以很好的对其进行归类。在这种情况下，全局描述子涉及图像所有像素的特定属性，这个属性可以是颜色直方图，纹理，边缘等。如图 3.7 所示，使用哪种特征描述子取决于图像处理的情景。由于地理空间的航拍图像在颜色或者纹理等全局特征上，图片的相似度极高，不易辨识，难以找到图片之间的匹配关系，所以此文选用局部特征描述子：SIFT。

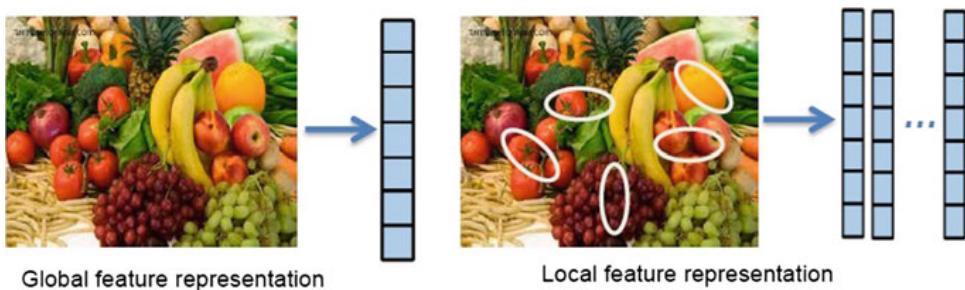


图 3.7 全局特征与局部特征描述

特征点提取包括特征点检测与特征点描述两部分，在此以SIFT特征点为例，简单介绍一下检测子与描述子的使用方法。检测子有很多，使用较多的有Harris Detector<sup>[7]</sup>，FAST Detector<sup>[6]</sup>。Harris检测子结合边缘与角点检测方法得到图像不同方向的自相关系数变化率，例如强度变化率等，该方法将局部特征描述为对称的自相关系数矩阵。Fast检测子考虑某像素周围一定大小的圆，院上像素强度与该像素比较，如果大于或小于某一阈值则定义该像素为角点，这两种检测子对尺度变化都不具有不

变性，不适合直接应用在拍摄高度不同的航拍图形的特征检测中。而SIFT作为一种具有尺度、放缩不变，对不同光线鲁棒的特定，很适合应用于环境复杂的航拍图形的特征特取与描述中。SIFT算法由四个步骤组成：尺度空间极值检测，兴趣点（关键点）定位，方向计算以及兴趣点描述。第一个阶段是使用Difference of Gaussian(DoG)确认潜在的兴趣点，这里使用DoG代替Laplacian of Gaussian (LoG)算子以提高计算速度。

Laplacian of Gaussian (LoG)是二阶导数的线性组合，也是检测斑点(blob)的检测子，给出一个图片 $I(x,y)$ ,尺度空间 $L(x,y,\sigma)$ 是由图片 $I(x,y)$ 与不同大小的高斯核 $G(x,y,\sigma)$ 卷积得到的，表示为

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \quad (3.12)$$

其中，

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (3.13)$$

据此可计算Laplacian算子为

$$\nabla^2 L(x,y,\sigma) = L_{xx}(x,y,\sigma) + L_{yy}(x,y,\sigma) \quad (3.14)$$

这对于大小为 $\sqrt{2\sigma}$ 的斑点有最强的响应，然而该算子严重依赖图像中斑点大小与用来平滑图像的高斯核大小，即大小不同的高斯核函数被用来寻找对应大小的blob。为了在一幅图像中自动检测不同大小的斑点，一种多尺度的方法被提出来<sup>[8]</sup>，其通过尺度不同的归一化的Laplacian算子对图片进行平滑从而在不同尺度空间找到大小不同的blob，尺度归一化的高斯核函数为式 ??。

$$\nabla^2_{norm} L(x,y,\sigma) = \sigma^2 L_{xx}(x,y,\sigma) + L_{yy}(x,y,\sigma) \quad (3.15)$$

大的 $\sigma$ 对应图片的粗略特征，小 $sigma$ 对应图片的精细特征。同时LoG算子是对称的，所以它具有blob旋转对其没有影响，但是LoG算子耗费计算资源，为了加速计算并在尺度空间检测到稳定的兴趣点，Lowe<sup>[2]</sup>提出了高斯差分尺度空间，简称DoG

scale-space。如式 3.16所示，利用不同尺度的高斯差分核函数与图像卷积生成。

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y)) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (3.16)$$

计算DoG的关键步骤是构建图像金字塔：对于某个图像，使用大小不同的高斯核与该图像卷积，可以得到模糊程度不同的但是图像长宽一致的图像，这些图像构成了一个八度(octave)，而后对这一系列图像进行降采样，降采样就是对octave中的图像隔行隔列采集像素，最后图像尺度(scale)变为原来大小的 $\frac{1}{4}$ ，构成下一个octave。这样就避免了对不同尺度空间下的图像进行卷积，减少了计算。为了寻找尺度空间的极值点，在DoG尺度空间中，每个采样点要和其周围所有相邻点（8邻域中，一个像素点周围共有26个像素点）比较，若该点是极大值或者极小值则定位为兴趣点。整个过程可以用图 3.8描述。

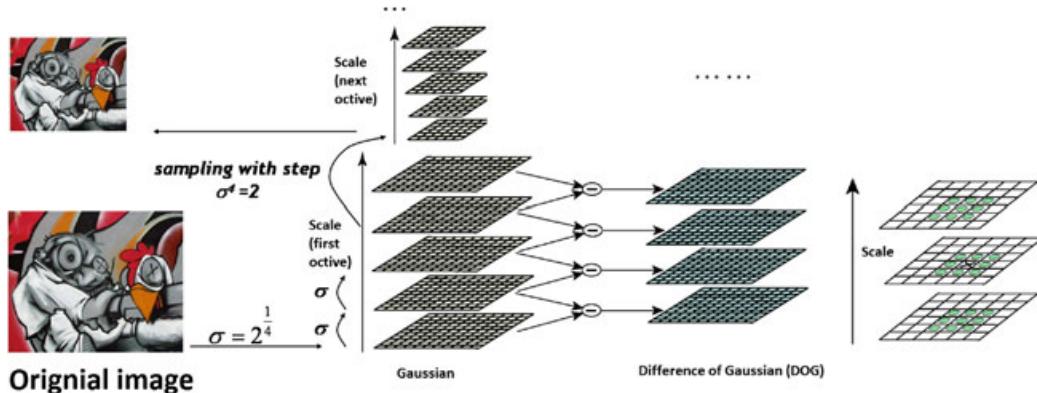


图 3.8 图像金字塔的构建

检测到特征点后，需要定量的表示各个特征点，以为特征点匹配做准备。Lowe在其论文中提出用一个128维度的向量表示每个特征点，具体做法为：如图 3.9，首先对每个关键点周围的 $16 \times 16$  邻域内所有像素计算其梯度的幅值与方向，用直方图统计邻域像素的梯度方向，梯度直方图将 $0 \sim 360^\circ$ 分为8个区间，对得到的直方图进行高斯平滑，以减少突变带来的影响。选取直方图的峰值作为该关键点处邻域梯度的主要方向。所以在每个 $4 \times 4$ 的象限内，将每个像素的主方向加权到直方图的8个方向区间中的一个，计算一个梯度方向直方图，这样对于每个特征点可以形成 $4 \times 4 \times 8 = 128$  维的描述子。

至此基本介绍了本文使用的SIFT特征点提取的方法，详细介绍了特征的提取与描述方法，以上是使用SIFT特征点检测多张图像的结果，其中特征点主方向以及特征点的强度分别表示在图中的。。。

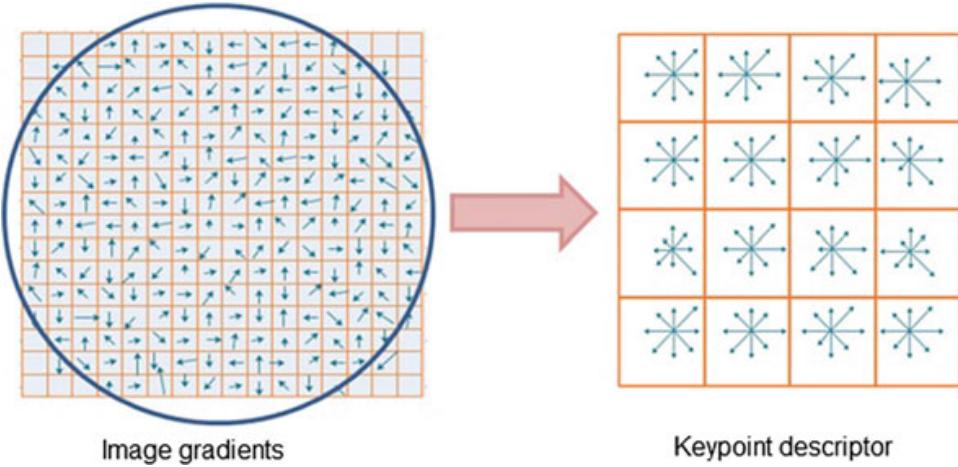


图 3.9 SIFT描述子的构建



图 3.10 SIFT特征点检测结果

### 3.5 特征点匹配

上面提到本文使用SIFT特征点作为关键点检测的方法，特征点提取并描述后，就应该找出各个图像中以特征点为依据的关系，即每张图像中的特征点如何与其他图像的特征点进行匹配，找到相近的特征点并去除错误的匹配的特征点。特征点匹配解决航拍图像之间的数据关联问题，将无序的图片数据集建立关联关系，为下文通过图像之间的关系，计算相机位姿与还原特征点的三维坐标提供基础。很好理解，一对匹配点就是实际空间的一个点，如果将所有图片对应同一个真实点的特征点连在一起，将使数据集中所有无序图片“有序化”。相同的，如果匹配点中有大量的错误匹配，那

对后续的位姿估计与还原三维坐标都产生不利影响，所以如果去除错误匹配更是至关重要的问题。下面就这两方面展开讨论。

假设在图片集 $I = \{I_i | i = 1 \dots N_I\}$ 中找到的局部特征点为 $F_i = \{(x_j, f_j) | j = 1 \dots N_{F_i}\}$ ，其中 $x_j$ 表示特征点位置坐标， $f_j$ 表示特征点描述子，如果是SIFT特征点，则 $f_j$ 是128维向量。对于 $F_i$ 与 $F_j$ 的特征点匹配的方法，最简单的是使用暴力匹配，即每个 $(x_j, f_j) | j = 1 \dots N_{F_i}$ 与 $(x_j, f_j) | j = 1 \dots N_{F_j}$ 中特征点测量两个向量之间的距离，需要进行 $N_{F_i} \times N_{F_j}$ 次距离比较。对于像SIFT这样的浮点描述子，一般采用欧拉距离作为衡量依据，而对于像BRIEF（ORB特征使用的描述子）二进制描述子，则一般使用汉明距离作为度量依据。

但是当图片长宽较大，分辨率较高时，图片特征点会很多，这时采用暴力匹配会导致效率很低，一般实际使用中采用近似最近邻(Approximate Nearest Neighbor, ANN)匹配方法。事实上，无论SIFT特征匹配还是数据库检索本质上是相同的，都是使用距离函数在高维矢量空间中寻找相近对象的过程。常用的方法试K近邻查询，设置查询点与正整数K，从另一个数据集中根据距离公式找到距离最近的K个数据，如果K=1，改方法变为最近邻查询算法。K近邻查询算法是通过构建KD-Tree或者R-Tree等实现的。在此不做详细介绍。

在此简单介绍近似最近邻的快速库(Fast Library for Approximate Nearest Neighbors, FLANN)，FLANN具有一种内部机制，该机制可以根据数据本身选择合适的算法来处理数据集，基于FLANN的匹配非常准确、快速。FLANN在使用的时候需要配置两个参数：indexParams和searchParams。使用FLANN时，indexParams可以选择为LinearIndex、KTreeIndex、KMeansIndex、CompositeIndex和AutotuneIndex，其中KTreeIndex灵活且可被并行处理。对于searchParams，用来指定索引树被遍历的次数，即check值，该值越大则匹配的时间越长，当然也越准确。

当然即使用上面的方法，也无法保证所有的匹配都是正确的，Lowe<sup>[9]</sup>在1999提出一种简单的方法可以剔除大部分的错误匹配。对于某个特征点向量，如果查询数据集中与其距离最近的特征点向量与次近的特征点向量的距离之比大于0.7，可以避免接近90%的错误匹配，但是正确的匹配也因此变少。使用FLANN并采用Lowe提出的方法可以去掉大部分的错误匹配，如图??所示。

以上提到的方法，匹配后仍然存在错误的匹配点，由于刚刚的匹配是单纯基于特征描述子的向量值，无法保证匹配的两个特征点对应相同的场景点。因此SfM使用投

影几何的知识估计两张图像的特征点变换。依赖图像对的空间配置，不同投影描述图片间不同的几何关系。例如，单应性变换（又称射影变换<sup>[10]</sup>）描述相机拍摄的二维图片之间纯旋转和平移变换。对极几何通过基本矩阵（Essential matrix） $E$ 和基础矩阵（Fundamental matrix） $F$ 描述移动相机的关系，并可以通过三焦张量扩展到三视图的关系。无论哪种变换关系，有效的变换可以满足大部分匹配点的几何关系，那么这种变换或者匹配点被认为是有效的。本文在删除错误匹配时，使用的是单应性变换，一下简单介绍一下单应性变换的详细内容。

单应性又称射影变换、保线变换，与坐标系无关。映射 $h$ 是射影变换的充要条件是：存在一个 $3 \times 3$ 非奇异矩阵 $H$ ，使得任意一个矢量 $X$ 表示的坐标点都满足 $h(x) = HX$ 。即

$$\begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (3.17)$$

矩阵 $H$ 是一个齐次矩阵，在 $H$ 的九个元素中有八个独立比率，即交比(cross ratio) $h_{11} : h_{12} : h_{13} : h_{21} : h_{22} : h_{23} : h_{31} : h_{32} : h_{33}$ 不会因为 $H$ 因为乘以非零因子而变化，所以单应性变化有八个自由度，而每个二维特征点可提供2个方程，故需要解算 $H$ 需要至少4对匹配点。已知4对匹配点，使用直接线性法解算 $H$ ，示意图如图 3.11，已知两个视图四对匹配的特征点坐标 $X_1, X_2$ ，求 $H$ 的过程。线性变换表示为 $X_2 = HX_1$ ，

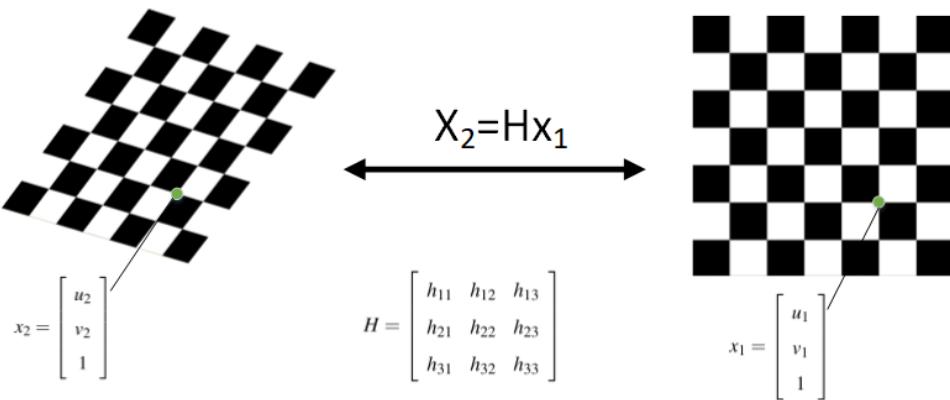


图 3.11 单应性变换

这是齐次矢量方程，三维矢量 $X_2$ 和 $HX_1$ 不相等，实际可以写为 $\lambda X_2 = HX_1$ ，线性变换可以变形为 $X_2 \times HX_1 = 0$ （矢量与自身外积为零），记为 $[X_2] \times HX_1 = 0$ ，设 $h_i$ 为 $H$ 的第*i*行，

$$\begin{aligned}
& \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} \times \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} X_1 = \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} \times \begin{bmatrix} h_1 X_1 \\ h_2 X_1 \\ h_3 X_1 \end{bmatrix} = \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} \times \begin{bmatrix} X_1^T h_1^T \\ X_1^T h_2^T \\ X_1^T h_3^T \end{bmatrix} \\
& = \begin{bmatrix} 0 & -1 & v_2 \\ 1 & 0 & -u_2 \\ -v_2 & u_2 & 0 \end{bmatrix} \begin{bmatrix} X_1^T & 0_{1 \times 3} & 0_{1 \times 3} \\ 0_{1 \times 3} & X_1^T & 0_{1 \times 3} \\ 0_{1 \times 3} & 0_{1 \times 3} & X_1^T \end{bmatrix} \begin{bmatrix} h_1^T \\ h_2^T \\ h_3^T \end{bmatrix} \tag{3.18}
\end{aligned}$$

最后一步是由线性代数的知识得到，即对3维矢量 $x = (x_1, x_2, x_3)$ 的 $3 \times 3$ 的反对称矩阵为式 3.20

$$[a]_\times = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \tag{3.19}$$

合并 3.18 最后一步的前两项得到下式为

$$= \begin{bmatrix} 0_{1 \times 3} & -X_1^T & v_2 X_1^T \\ X_1^T & 0_{1 \times 3} & -u_2 X_1^T \\ -v_2 X_1^T & u_2 X_1^T & 0_{1 \times 3} \end{bmatrix} \begin{bmatrix} h_1^T \\ h_2^T \\ h_3^T \end{bmatrix} = 0 \tag{3.20}$$

上式简写为 $Ab = 0$ ，其中 $A$ 为 $3 \times 9$ 矩阵， $\text{rank}(A) = 2$ ，这是因为每队匹配点只提供两个量： $u$ 和 $v$ ， $b$ 为 $9 \times 1$ 矩阵，由于只需要维持 $b$ 各项的交比不变即可，所以有8个自由度，故需要4对匹配点即可在这个线性系统上获得足够的条件计算得到单应性矩阵 $H$ 的各项。但是如果我们有很多匹配点，这时候

### 3.6 多视图重建

在式 3.9 中， $z_c$ 为每个像素的深度，在单目成像时， $z_c$ 是未知的。所以当知道某点的世界坐标系时，可以很容易得到该点的像素坐标，但是反之无法已知像素坐标得到世界坐标。这将为我们下面

**3.6.1 两视图重建**

**3.6.2 2D-3D位姿求解**

**3.6.3 三角定位法**

**3.7 三维重建中的优化**

**3.7.1 光束平差法**

**3.7.2 最小化相机中心位置误差**

**3.8 稀疏点云的渲染**

**3.8.1 点云稠密化**

**3.8.2 点云网格渲染（三角剖分）**

**3.9 本章小结**

## 第4章 二维图像中的道路提取

4.1 引言

4.2 图像二值化

4.3 非道路区域移除

4.3.1 开闭运算

4.3.2 轮廓提取

4.4 拓扑处理

4.4.1 拓扑细化

4.4.2 去除小枝丫

4.5 本章小结

## 第 5 章 实验结果分析

### 5.1 引言

### 5.2 全景图像构建地图实验结果

#### 5.2.1 全景图像地图实验平台搭建

#### 5.2.2 全景图像地图构建效果及精度分析

### 5.3 地理空间三维重建实验结果

#### 5.3.1 地理空间三维重建实验平台搭建

#### 5.3.2 三维重建效果及精度分析

### 5.4 道路提取精度分析

### 5.5 本章小结





## 总结与展望

### 工作总结

这里是总结。

### 未来工作展望

这里是展望。

## 参考文献

- [1] Mattyus G, Wang S, Fidler S, et al. Enhancing Road Maps by Parsing Aerial Images Around the World[C] IEEE International Conference on Computer Vision. IEEE, 2015:1689-1697.
- [2] Lowe D G, Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
- [3] Bay H, Ess A, Tuytelaars T, et al. Speeded-Up Robust Features (SURF)[J]. Computer Vision & Image Understanding, 2008, 110(3):346-359.
- [4] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C] IEEE International Conference on Computer Vision. IEEE, 2012:2564-2571.
- [5] 高翔, 张涛等, 视觉SLAM十四讲: 从理论到实践[B] 电子工业出版社, 2017-3:138
- [6] Rosten E, Drummond T. Machine learning for high-speed corner detection[J]. Computer Vision - ECCV 2006, 2006: 430-443.
- [7] Harris C, Stephens M. A combined corner and edge detector[C]//Alvey vision conference. 1988, 15(50): 10.5244.
- [8] Lindeberg, T.: Feature detection with automatic scale selection. Int. J. Comput. Vis. 30(2), 79 – 116 (1998)
- [9] Lowe D G. Object Recognition from Local Scale-Invariant Features[C]// iccv. IEEE Computer Society, 1999:1150.
- [10] Multiple View Geometry in Computer Vision Second Edition, Richard Hartley and Andrew Zisserman, Cambridge University Press, March 2004:105

## 攻读硕士学位期间发表论文与研究成果清单

### 项目成果

- [1] 航天八院目标位姿视觉测量项目

## 致谢