

## 【Café 速递】宦麟茜：室内外场景理解与建模及科研经历分享

**核心提示：**什么是室内外场景理解与建模？基于室内外场景影像构建语义三维模型包含哪些问题？如何针对室外场景与室内场景的不同进行语义解译与几何重建？如何对室内外场景理解与建模展开应用研究？报告人宦麟茜将根据个人及课题组的科研实践，介绍室内外语义三维重建中面临的问题及相关科研探索。

主持：董明玥 摄影：许梦子 摄像：冯锐 文字：许梦子，董明玥

### >>>人物名片

宦麟茜，武汉大学测绘遥感信息工程国家重点实验室 20 级博士生，师从龚健雅院士，郑先伟副研究员。以第一作者或学生二作发表 SCI 一区文章三篇，其中 IEEE TPAMI 一篇(IF: 17.86)，ISPRS&RS 两篇(IF: 8.9)。曾获得学业奖学金一等奖，学术创新二等奖。

### >>>报告现场

2021 年 12 月 3 日晚上 7 点，测绘遥感信息工程国家重点实验室 20 级博士生宦麟茜做客 GeoScience Café 第 316 期讲座。宦麟茜根据个人及课题组的科研实践，介绍了室内外语义三维重建中面临的问题及相关科研探索，分享了科研学习过程中的经验与思考，解答了同学们科研方面的问题，同学们受益匪浅。



图 1 宦麟茜作精彩报告

实景三维是国家新型基础设施建设的重要组成部分，为经济社会发展和各部门信息化提供统一的空间基底，在国防工程、大众服务等诸多领域具有广泛的应用。而基于视觉图像的实景三维建模，是将多视图像进行图像匹配后开展运动结构恢复，从而实现视觉定位，然后通过稀疏点云重建和稠密点云重建解析点云语义，从而实现语义三维重建，以便场景理解可视化和后续分析应用。其中，根据类别分为室外场景和室内场景，因此宦麟茜从这两个方面的场景理解与建模展开分享。

## 室外场景理解与建模

宦麟茜首先介绍了课题组在室外场景理解与建模领域的相关工作。

无人机影像覆盖范围大，可以采集到完整的顶面信息，其建立的三维模型顶面完整，立面缺失；而街景分辨率高，立面纹理细节丰富，其建立的三维模型立面完整，顶面缺失。因此如果进行空地影像联合的三维重建，可以实现两类数据优势互补。由于空地影像视角、光照、传感器差异较大，导致特征点相似度低，因此匹配较为困难。考虑到建筑物存在稳定的网格结构，课题组提出了基于点-网格特征互监督的空地影像匹配方法[1]。由于多数场景均不是简单的平面对象，而是复杂多对象组合且表面起伏多变，课题组提出了基于局部非线性仿射验证的多传感器影像匹配方法[2]。

宦麟茜还补充说明了为什么在建模后进行场景理解。从几何建模角度来看，需要提取精细边缘特征，支持构建几何精确、拓扑准确的三维表面模型；从单体建模角度：需要解析场景语义，把对象分割开，并且知道对应的内容，因此涉及到一个关键性内容——语义分割。其中，语义分割中包含两大根本问题，一是尺度问题，即不同尺寸物体难兼顾问题；二是卷积下采样导致边缘细节丢失。应用方向有超高分辨率正射影像语义分割、街景影像语义分割、多模态融合语义分割等。



图 2 宦麟茜讲解室外场景理解与建模

宦麟茜进一步指出,如果在建模之前可以提取带一定语义信息的物体轮廓特征,可以很好地辅助三维重建。当前的深度学习方法由于特征混淆和训练数据类别不平衡,预测的边缘较粗,因此课题组提出了基于卷积特征解混细边缘提取的方法[3]。

最后是直线段匹配的工作。由于常用的仿射变换不能很好的作为直线段间的几何关系约束,因此提出了基于平滑可变投影变换建模的直线匹配方法[4],这个变换可以很好的用于直线段的误匹配剔除。

### 室内场景理解与建模

实景三维在室内有非常多的应用,比如数字博物馆、管道的增强现实展示、智能家居、机器人控制等。这些应用对实景三维建模有着不同的需求,有的需要几何精确,有的还需要语义属性等知识来支撑。

然而，室内空间特点和室外有很大的区别，使得室内实景三维建模产生了许多难点，比如物体封闭遮挡、对象繁杂、动态目标、多楼层、重复纹理、弱纹理等。因此提出一种分治合并的室内图像处理方法，具体做法是按照图片内容进行图片集的分块，减少不必要的匹配和计算；对分块的图片集单独重建，再进行离散点云模型合并，得到完整的室内模型。这样的室内实景三维建模工作框架可分为三层，从底层的语义分类出发，到中层子模型重建，再到顶层的模型拼接，自下而上地完成建模流程。

室内实景三维应用之一是室内视觉定位。宦麟茜在此处以测国重实验室四楼的小实验为例子，行人环绕很多圈后，PDR（Pedestrian Dead Reckoning，行人航迹推算算法）的定位误差会越来越大，通过基于室内三维模型的视觉定位，可以实现轨迹纠偏，获得比较连续的高精度定位。

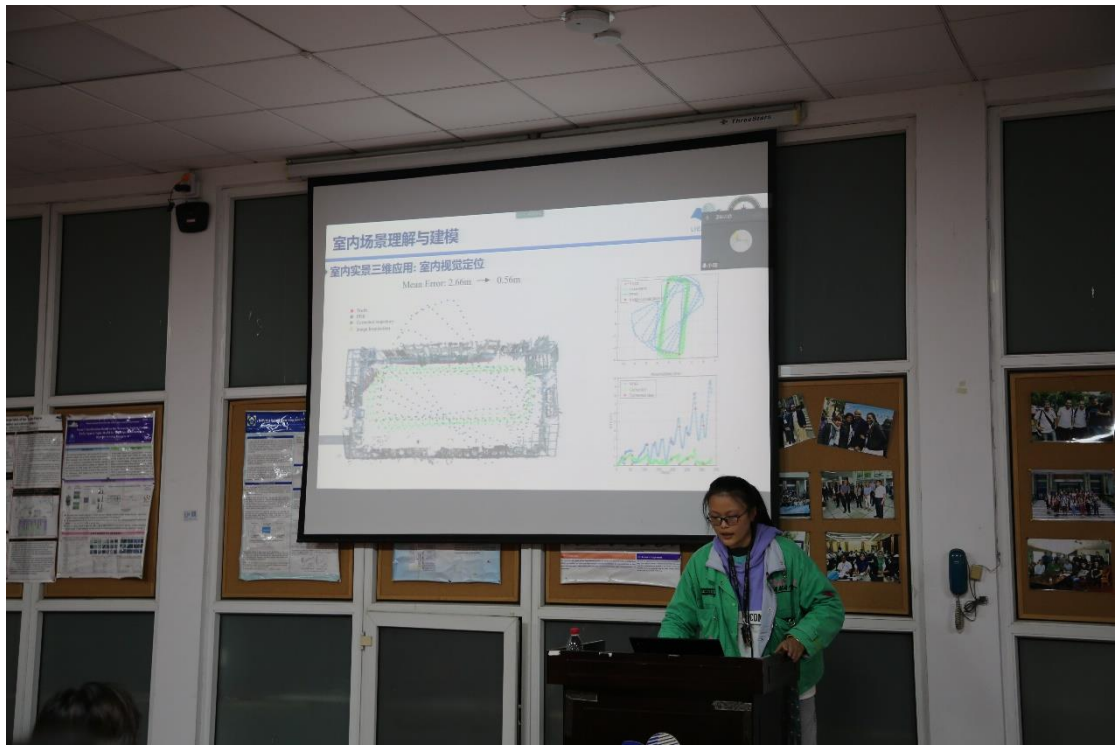


图3 宦麟茜讲解室内实景三维应用中的室内视觉定位

室内应用向精细化、智能化推进，亟需灵活的结构化单体模型支撑，然而室内空间封闭遮挡、布局易变，传统建模思路难以适应需求。如何实现室内高度遮挡条件下的对象几何结构推理成为了关键的难点。因此，课题组设计了基于多任务部件级的室内语义三维重建方法和几何信息增强的知识辅助室内语义三维重建方法，以实现利用低成本RGB-D数据的室内结构化语义三维重建。

## 结语

最后，宦麟茜介绍了未来应用方向——实时视频实景三维融合可视化，达到所见即所得的目的。她希望把团队做的算法集成起来，方便更多人可以通过一个系统来使用，并且总结了以下几点思考：

1. 卷积神经网络结构设计可分两条路线走，一是设计高复杂度、组合型、过拟合的多流网络，二是设计轻量级、可解释、易训练的单流网络。
2. 在卷积神经网络损失设计中，考虑到好的网络模块可以提取更有针对性的特征，好的损失可以引导网络学习，将提取的特征真正在目标任务上发挥作用。
3. 为了更好地进行面向对地观测数据的通用模块与框架开发，需要进一步加强各方向的融合交叉，让经验丰富的神经网络设计人员和各方向的专业人员共同探索数据和任务的特性。

### >>>互动交流

**提问人一：**你说到室外场景需要语义辅助重建，那么具体而言语义信息起到的作用是什么？

**宦麟茜：**当三维场景重建出来以后，我们要的不仅是一个单纯的可视化的三维场景，我们更希望这个场景中的物体具有计算机可识别的语义信息，比如给定一个三维点，计算机应当知道它是属于哪一栋建筑物或者是其它地物；不同的建筑物、建筑物与地面之间应当是可以被区分开来的。目前我们往往需要人工标注来达到这一效果，其实是非常费时费力的，尤其是对于动态的城市场景而言，我们很难做到每一个城市都人工标注完成。所以今后半自动化的城市场景语义建模是一个趋势。

**提问人一：**如何获取三维场景的语义信息呢？

**宦麟茜：**我们组主要是通过视觉的方法来做的三维重建，所以三维点是可以在二维影像上的对应像素的，这意味着我们可以将二维分割得到的语义信息直接传递到三维。对于如何克服多视图语义不一致的问题，我们组之前有师姐做过这个，可以采用一种投票的方式。因为做语义分割，每个像素的类别其实是有一个概率值的。如果一个三维点在多个视角中都有对应的像素，一种方法是如果它在多数影像上都被分到了一个类，那么它就说这个类，另一种则是根据它们分类的概率值，提取概率最高的那个类。当然这两种是最基本的方法，后面也有人提出了一些更复杂的操作。

**提问人二：** 虽然我不是做这个方向的，但我感觉到你对自己做的东西特别热爱，讲的也特别有感染力。另外你做的这个方向跟实际应用之间联系非常密切，你对它的理解也非常深刻。你后面说到要有理想，有蓝图。理想这个大家都有，但是我觉得蓝图很难有：我们应该每一步都很清晰自己的目标，还是自己在做的时候慢慢清晰起来呢？

**宦麟茜：** 首先我们组是希望能和实际应用紧密结合的。其实这些任务工业界也能做，比如开发三维重建的软件，所以我们作为高校，不仅要会做这些工业界做的代码工作，还要能够做他们做不到的事情，例如算法的设计。我们课题组的框架大概就是这个样子的，我在里面做一些具体的工作。

如果说这是一个大的蓝图，我个人的蓝图其实是一个比较曲折的故事。我一开始是做语义分割的，然后在做的过程中，我发现它的边缘不好，同时也有越来越多的人也开始关注这个问题，这说明这个问题的确是非常重要的，于是我就继续做了下去。可能这也和我的性格有关，我比较喜欢岔开来做，看各种领域的东西，寻找它们的共性，去做自己感兴趣的东西。我这个做法并不适合每个人，因为重新开始一个方向的成本很高。

其实你对一个问题的认知是随着实验和文章的推进逐步深入的。比如像我做边缘检测的时候，最开始的时候我只想做一个比较细的边缘，但至于这个边缘能做到多细，我是完全没有概念的，只是觉得可以这么试一下，然后就一路这么做下来了。这个工作断断续续做了一年多，对工作的认知也是一步一步加深。所以你不大可能一开始就知道任务问题的本质是怎样的，只有在积累越来越多的时候，你在面对不同的领域，才会比较敏锐的发现，原来问题的核心在这，并且还会发现很多问题之间其实也是相通的。



图5 同学们积极提问

**提问人三：**师姐您好，我知道您发了 TPAMI 这样特别高影响力的期刊，我想了解一下您在投稿过程中有没有遇到什么挑战？能不能给我们分享一些经验？

**宦麟茜：**你如果想要发一个特别好的文章，比如说期刊质量要求特别高的话，就要把格局打开，不能完全局限在刷精度上。就像我们做语义分割，你加一点上下文，加一点多尺度模块，分是刷上去了，但都是些大家老生常谈的思路。你能不能突然引进一个新的思路，提出一个新的问题，显示出你对这个任务确实有很深入的研究，这是十分关键的。为什么我们在看别人文章的时候，有的文章我们看完了仿佛没看一样，有的文章虽然方法很简单，但是我们会觉得它写得非常的好。这就是因为作者在写的时候，他就告诉我他是从哪个角度切入的。所以一篇文章的 *insight* 是非常重要的，包括你在读文章的时候也要抓住它的 *insight*，并去思考这个 *insight* 有没有吸引到你，带给你了什么启发，有时候到最后甚至方法都不重要了。

这是实验和思路上的经验。另一方面就是写作。不同的期刊对语言风格的要求也是不一样的。像这个期刊它会要求你的语言非常精炼，要读起来句句都是干货，不能显得很业余。我当时是在导师和师兄的帮助下和他们共同完成的，在写作方面受到了极大的锻炼。除了文字部分以外，画图也非常重要。我当时也是被师兄和老师逼着改了很多遍，可能两张图就画了一个多星期。当你经过这么

一个过程以后，再去画就有经验了。

总的来说，要发一篇好文章，中间一定会经历很多的磨难，而且不要怕被人喷，毕竟你的文章说白了是要能展现出去的东西。所以大家要慢着性子，要把每一篇文章当作一个艺术品去对待。不要觉得扣文字，做 PPT 这些是浪费时间，你会发现，你在做的过程中对一些东西的看法会潜移默化的发生变化，它实际上是一个很好的锻炼过程。



图 6 宦麟茜（第一排左四）与部分听众、Geoscience Café 团队成员合影留念

### 参考文献

- [1] X. Zheng, H. Li, H. Xiong, X. J. I. J. o. S. T. i. A. E. O. Xie, and R. Sensing, "Lattice-Point Mutually Guided Ground-to-Aerial Feature Matching for Urban Scene Images," vol. 14, pp. 4737-4752, 2021.
- [2] H. Li, X. Zheng, M. Dong, G.-S. Xia, H. J. I. T. o. G. Xiong, and R. Sensing, "Locally Non-linear Affine Verification for Multi-sensor Image Matching," 2021.
- [3] L. Huan *et al.*, "Unmixing Convolutional Features for Crisp Edge Detection," 2021.
- [4] X. Zheng *et al.*, "Smoothly varying projective transformation for line segment matching," vol. 183, pp. 129-146, 2022.

GeoScience Café 以“谈笑间成就梦想”为目标，于每周五晚 7:00 在实验室四楼休闲厅，邀请 1-4 位嘉宾，为大家带来学术报告或经验分享。报告内容包括摄



影测量与遥感、地理信息系统、导航与定位服务等研究方向，听众可在报告结束后向嘉宾提问、与嘉宾交流探讨，同时每学期还会举办2期人文类讲座和2场导师信息分享会。每期报告会根据嘉宾意愿在B站开设直播，使不能来到现场的听众同步参与。报告PPT和视频会在征得嘉宾同意的情况下在qq群和B站上发布。

更多精彩内容（讲座预告、讲座回顾、报告PPT、报告视频）敬请通过以下方式获取：



**QQ群**



**微信公众号**



**B站直播**