

DS-SLAM

- 论文: [DS-SLAM_A_Semantic_Visual_SLAM_towards_Dynamic_Environments.pdf](#)
- 开源: <https://github.com/ivipsourcecode/DS-SLAM>

主要贡献

- 1、基于ORB-SLAM2 提出了动态环境中的完整语义SLAM系统（DS-SLAM），可以减少动态对象对位姿估计的影响。在TUM RGB-D数据集上评估该系统的有效性。结果表明，DS-SLAM在动态环境中的准确性和鲁棒性方面优于ORB-SLAM2。该系统还与机器人操作系统（ROS）[10]集成，通过在真实环境中对机器人测试DS-SLAM来验证其性能。
- 2、我们将一个实时语义分割网络放在一个独立的线程中，它将语义分割与移动一致性检查方法结合起来，过滤掉场景的动态部分，如走路的人。因此，在动态场景中，提升了定位模块和建图模块的稳定性和鲁棒性。
- 3、DS-SLAM创建了一个单独的线程来构建稠密的语义3D八叉树地图。稠密的语义三维八叉树地图采用优势对数计分法滤除不稳定体素并更新这些体素的语义。

介绍

- 1、以下是DS-SLAM的概述图和框架图，主要是对于五个线程关系的描述。

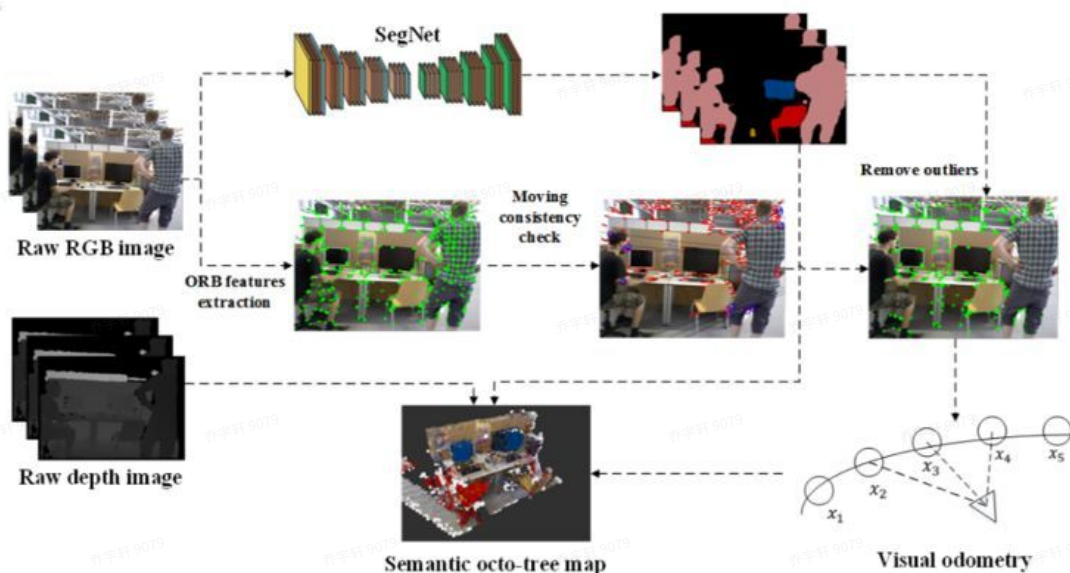


图1 DS-SLAM概述图

原始RGB图像用于同时进行语义分割和移动一致性检查。然后删除异常值并估计位姿。基于位姿，深度图像和语义分割结果，在独立线程中构建语义八叉树地图。

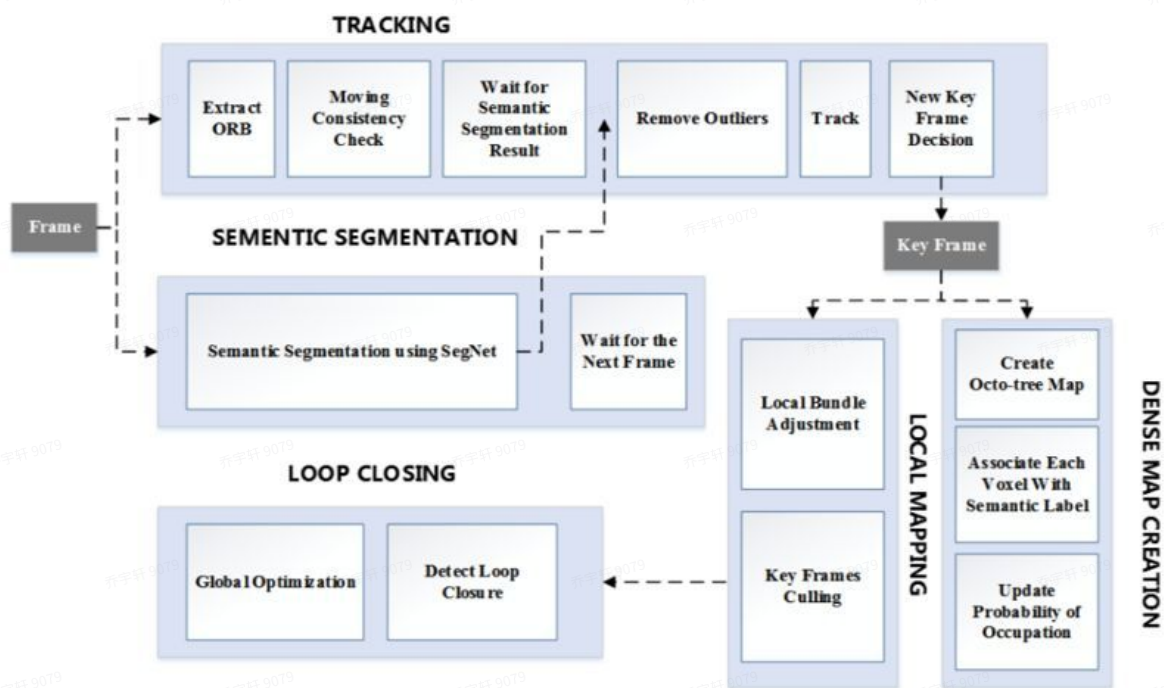


图2 DS-SLAM的框架图

局部地图线程和回环检测线程与ORB-SLAM2相同。前者处理新的关键帧并执行局部束调整以在相机姿势的周围实现最佳重建，而后者搜索回环并在检测到回环时执行图优化。

内容细节

1、框架

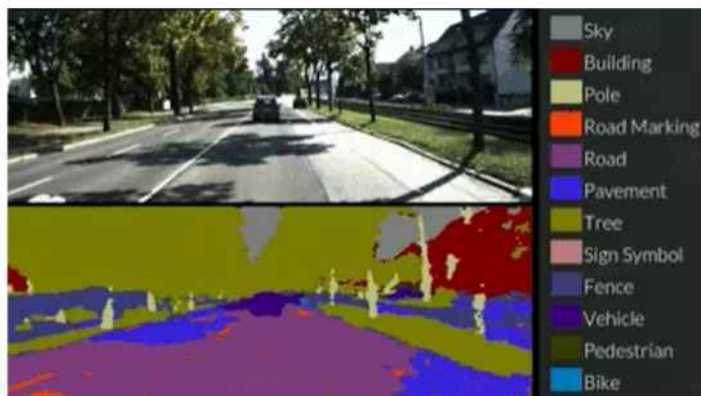
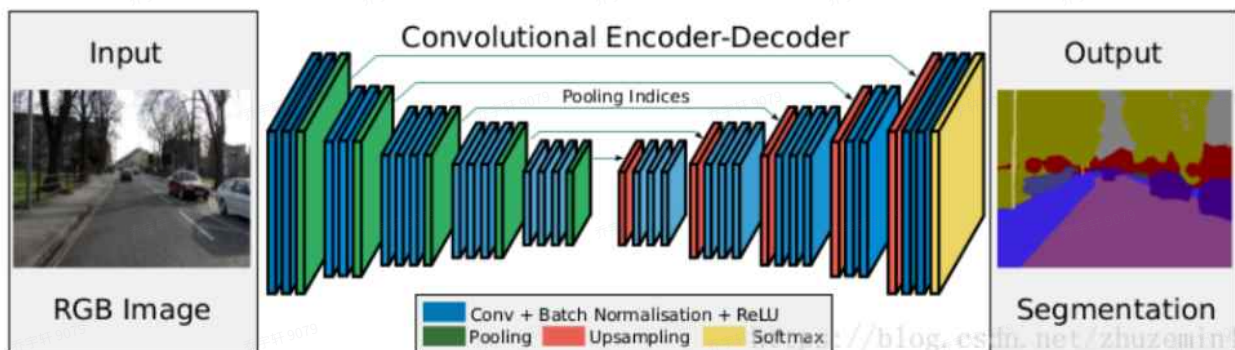
DS-SLAM由五个线程并行：tracking，semantic segmentation，local mapping，loop closing，dense semantic map creation.

ORB-SLAM的位姿估计来自于提取的ORB特征点，这导致了其在动态环境下不够鲁棒。DS-SLAM主要目标在于识别背景中的动态物体并进行剔除后再估计位姿。

DS-SLAM框架是ORB SLAM2，使用Kinect2分别捕捉彩色图和深度图。RGB图送入semantic线程和tracking线程。tracking线程中，首先提取ORB特征点，然后检查这些特征点的移动一致性(moving consistency)，并记录潜在的外点。同时，semantic线程并行运行，其pixel wise的分割结果于tracking线程的ORB特征结合，将移动物体的特征点定为外点。这些外点和上一步的潜在外点都被丢弃。根据剩余的稳定内点，计算转换矩阵。

2、语义分割

在语义（semantic）线程中引入SegNet，SegNet是Cambridge提出旨在解决自动驾驶或者智能机器人的图像语义分割深度网络，开放源码，基于caffe框架。



这个图是SegNet演示效果，来源是作者上传到YouTube的一个视频
(https://www.youtube.com/watch?v=CxanE_W46ts)

3、运动一致性检查

目标

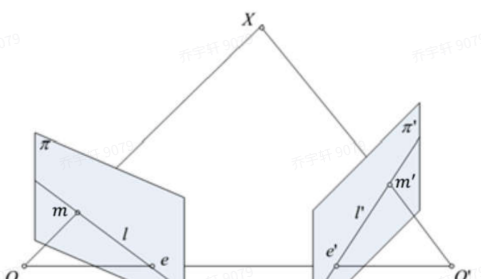
- 在提取了ORB特征点后，判断其中哪些特征点在实际场景下正在运动

实现方法

- 首先需要对每个特征点进行追踪，DS-SLAM采用光流金字塔。(optical flow pyramid)
- 对特征点对进行预处理，丢弃靠近边缘与不合群（如在 3×3 领域内像素差太大）的点。
- 基于剩余点对进行RANSAC计算基础矩阵，并依据点对计算极线。
- 计算配对点到极线的距离，根据预定阈值判断特征点是否属于移动物体。

基本想法

- 根据几何学，若静态物体上的特征点在不同视图下对应点1，点2。理论上，点2必定落在基础矩阵与点1决定的极线上。即如果点2与该极线相距过远，则特征点在运动。（参考下图）



4、剔除运动的特征点

该步主要实现剔除特征点中来自动态物体的部分。注意到语义分割结果告诉我们物体的轮廓，运动一致性检查告诉我们哪些是运动的特征点。结合两者的信息，我们可以实现这一目标。

DS-SLAM算法中将人作为动态物体的代表，实际上动态物体还有很多种。实现过程如下：语义分割后检测是否有人，如无，则直接进行位姿预测；如有，我们想判断其是否静止，根据运动一致性结果，检查其中的特征点是否运动。如果这些特征点没有运动，则认为人是静止不动的，不影响位姿的估计，如果特征点运动了，则认为人是运动的物体，要对轮廓内的特征点都进行丢弃。

5、稠密语义八叉树的建立（三维重建）

完成tracking与semantic segmentation后，我们得到了新的关键帧集与语义分割的结果。依据关键帧处的变换矩阵与深度，可以生成局部点云，进一步保存到全局八叉树中。实际上语义分割信息也可以保存在全局八叉树中，我们只需对每个体素(voxel)赋予一种颜色即可，如红色代表沙发等等。实际上这种赋予是以概率的方式进行赋予，以便更新体素上语义的分布。

问题：考虑到语义分割准确率受限（如物体重叠时，分割结果不可靠），如何对重建优化？

DS-SLAM引入了优势对数积分法(log-odd score method)来考虑这一问题，设 p 为某体素的占有概率（即属于某种颜色的概率），引入如下优势对数积分变换与反变换

$$l = \text{logit}(p) = \log\left(\frac{p}{1-p}\right) \quad (4)$$

The inverse transform:

$$p = \text{logit}^{-1}(l) = \frac{\exp(l)}{\exp(l) + 1} \quad (5)$$

依据变换进行如下递推计算，其中 $L(n|z_{1:t})$ 代表从开始到时间 t 体素 n 的优势对数积分， $L(n|z_t)$ 取值为0或某提前设定的固定值，取决于时间 t 下体素 n 是否被占用。

$$L(n|z_{1:t+1}) = L(n|z_{1:t}) + L(n|z_t)$$

递推结束后，可以通过反变换求的概率 p ，将其与某一阈值比较，决定 n 是否在八叉树中可视化。

引入优势对数积分与递推目的在于，当像素重复被观察到占据时，使积分提高。

主要结果

1、在TUM RGB-D数据集下的评估，本文实验对比比较简单，主要还是与ORB-SLAM2对比，下面是使用相对高动态的fr3_walking_xyz的对比图。其中绝对轨迹误差（ATE）和相对姿势误差（RPE）用于定量评估。

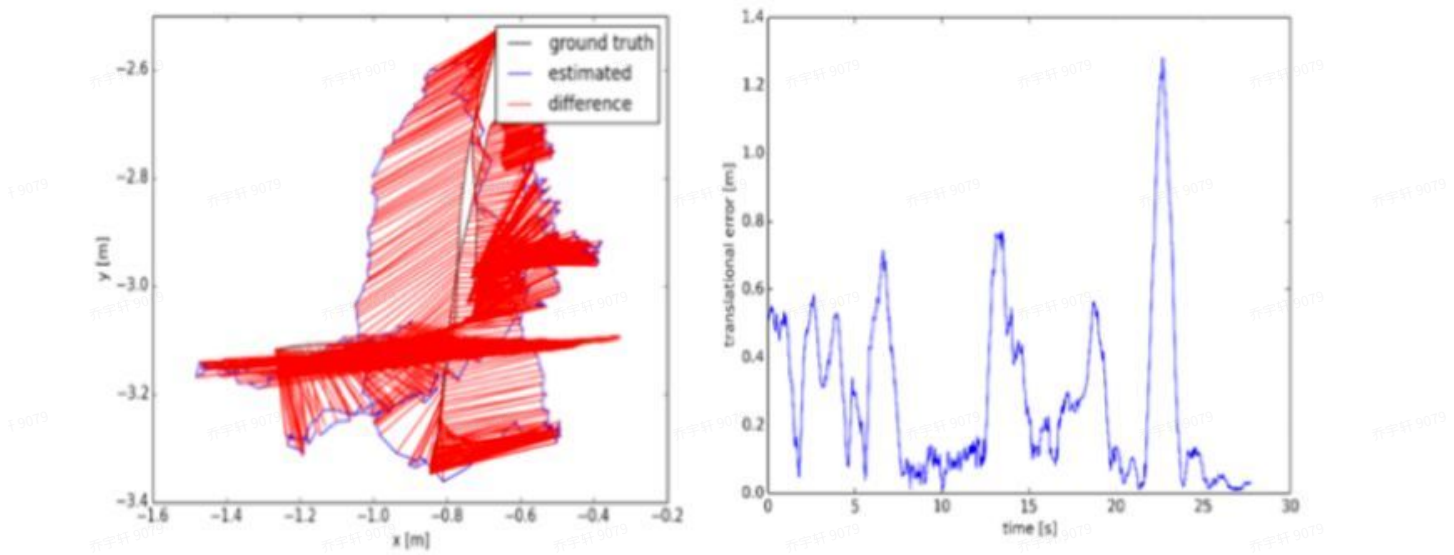


图3 ORB-SLAM2的ATE和RPE

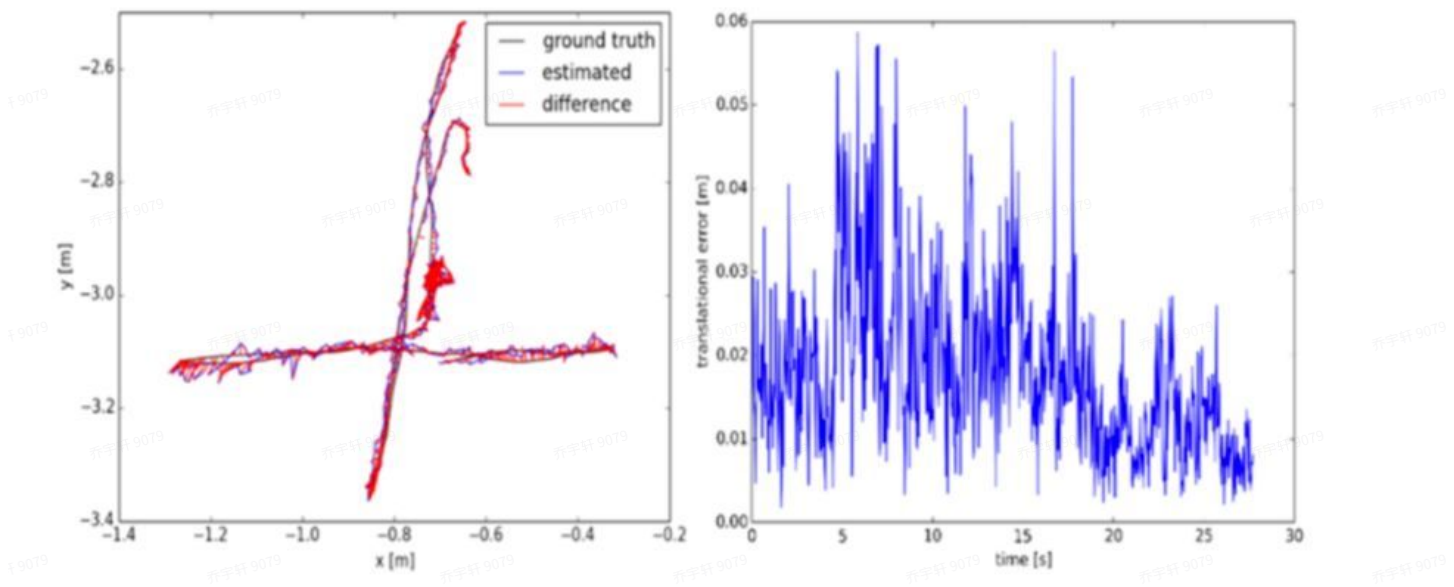


图4 DS-SLAM的ATE和RPE

2、ORB-SLAM2和DS-SLAM在TUM-RGBD数据集下的其他对比的表格，可以看出在动态环境下DS-SLAM提升的效果还是比较明显的。

TABLE I. RESULTS OF METRIC ROTATIONAL DRIFT (RPE)

| Sequences | ORB-SLAM2 | | | | DS-SLAM | | | | Improvements | | | |
|-----------|-----------|------|--------|-----|---------|------|--------|-----|--------------|------|--------|-----|
| | RMSE | Mean | Median | S D | RMSE | Mean | Median | S D | RMSE | Mean | Median | S D |

| | RMSE | Mean | Median | S.D. | RMSE | Mean | Median | S.D. | RMSE | Mean | Median | S.D. |
|--------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| fr3_walking_xyz | 7.7432 | 5.8765 | 4.5340 | 4.9895 | 0.8266 | 0.5836 | 0.4192 | 0.5826 | 89.32% | 90.07% | 90.75% | 88.32% |
| fr3_walking_static | 3.8958 | 1.6845 | 0.3571 | 3.5095 | 0.2690 | 0.2416 | 0.2259 | 0.1182 | 93.09% | 85.66% | 36.75% | 96.63% |
| fr3_walking_rpy | 8.0802 | 5.4558 | 2.7828 | 5.9499 | 3.0042 | 1.9187 | 0.9902 | 2.3065 | 62.82% | 64.83% | 64.42% | 61.23% |
| fr3_walking_half | 7.3744 | 4.5917 | 1.8143 | 5.7558 | 0.8142 | 0.7033 | 0.6217 | 0.4101 | 88.96% | 84.68% | 65.73% | 92.87% |
| fr3_sitting_static | 0.2881 | 0.2598 | 0.2493 | 0.1244 | 0.2735 | 0.2450 | 0.2351 | 0.1215 | 5.07% | 5.70% | 5.68% | 2.36% |

表1 度量旋转角度漂移的结果（RPE）

TABLE II. RESULTS OF METRIC TRANSLATIONAL DRIFT (RPE)

| Sequences | ORB-SLAM2 | | | | DS-SLAM | | | | Improvements | | | |
|--------------------|-----------|--------|--------|--------|---------|--------|--------|--------|--------------|--------|--------|--------|
| | RMSE | Mean | Median | S.D. | RMSE | Mean | Median | S.D. | RMSE | Mean | Median | S.D. |
| fr3_walking_xyz | 0.4124 | 0.3110 | 0.2465 | 0.2684 | 0.0333 | 0.0238 | 0.0181 | 0.0229 | 91.93% | 92.34% | 92.66% | 91.48% |
| fr3_walking_static | 0.2162 | 0.0905 | 0.0155 | 0.1962 | 0.0102 | 0.0091 | 0.0082 | 0.0048 | 95.27% | 90.00% | 47.07% | 97.58% |
| fr3_walking_rpy | 0.4249 | 0.2825 | 0.1487 | 0.3166 | 0.1503 | 0.0942 | 0.0457 | 0.1168 | 64.64% | 66.66% | 69.24% | 63.10% |
| fr3_walking_half | 0.3550 | 0.2161 | 0.0774 | 0.2810 | 0.0297 | 0.0256 | 0.0226 | 0.0152 | 91.62% | 88.16% | 70.74% | 94.60% |
| fr3_sitting_static | 0.0095 | 0.0083 | 0.0073 | 0.0046 | 0.0078 | 0.0068 | 0.0061 | 0.0038 | 17.61% | 17.81% | 17.01% | 16.96% |

表2 度量平移漂移的结果（RPE）

TABLE III. RESULTS OF METRICS ABSOLUTE TRAJECTORY ERROR (ATE)

| Sequences | ORB-SLAM2 | | | | DS-SLAM | | | | Improvements | | | |
|--------------------|-----------|--------|--------|--------|---------|--------|--------|--------|--------------|--------|--------|--------|
| | RMSE | Mean | Median | S.D. | RMSE | Mean | Median | S.D. | RMSE | Mean | Median | S.D. |
| fr3_walking_xyz | 0.7521 | 0.6492 | 0.5857 | 0.3759 | 0.0247 | 0.0186 | 0.0151 | 0.0161 | 96.71% | 97.13% | 97.42% | 95.71% |
| fr3_walking_static | 0.3900 | 0.3554 | 0.3087 | 0.1602 | 0.0081 | 0.0073 | 0.0067 | 0.0036 | 97.91% | 97.95% | 97.82% | 97.74% |
| fr3_walking_rpy | 0.8705 | 0.7425 | 0.7059 | 0.4520 | 0.4442 | 0.3768 | 0.2835 | 0.2350 | 48.97% | 49.26% | 59.84% | 48.02% |
| fr3_walking_half | 0.4863 | 0.4272 | 0.3964 | 0.2290 | 0.0303 | 0.0258 | 0.0222 | 0.0159 | 93.76% | 93.95% | 94.40% | 93.05% |
| fr3_sitting_static | 0.0087 | 0.0076 | 0.0066 | 0.0043 | 0.0065 | 0.0055 | 0.0049 | 0.0033 | 25.94% | 26.87% | 26.29% | 23.15% |

表三 度量绝对轨迹误差的结果（ATE）

| Module | ORB feature extraction | Moving consistency check | Semantic segmentation |
|----------|---------------------------|--------------------------------|--------------------------|
| Thread | Tracking | Tracking | Semantic segmentation |
| Time(ms) | 9.375046 | 29.50869 | 37.57330 |

表四 时间效率上的评估

3、实际环境的实验结果，输入为Kinect V2相机(960*540)。

定性地证明异常值拒绝的结果。如下图从顶行到底行的子图分别是ORB特征提取结果，基于光流的移动一致性检查结果，语义分割结果和异常值去除后的图像。绿点表示ORB特征点的位置，红色短线表示光流的方向。正如我们所看到的，通过移动一致性检查确定该人正在移动，然后有效地移除落入移动人区域中的特征点。

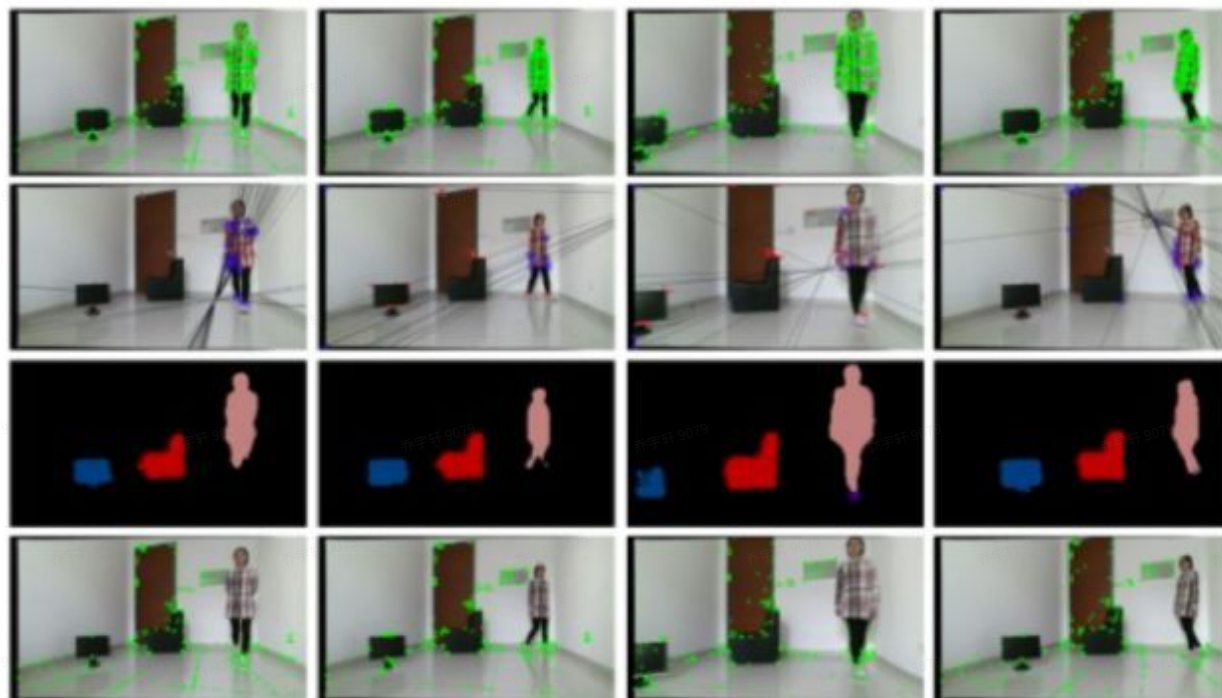


图5 实验室环境的实验结果，从顶行到底行的子图分别是ORB特征提取结果，基于光流的移动一致性检查结果，语义分割结果和异常值去除后的图像。。





图6 稠密的八叉树建图结果，红色体素代表沙发，蓝色体素代表显示器
此外，二维代价地图由八叉树地图生成，可用于导航。

存在的问题

- DS-SLAM中语义分割使用的SegNet支持的最多种类为20。而汽车行驶的道路环境可能非常复杂，可能DS-SLAM无法做到稳定的语义分割。
- DS-SLAM中为了对SegNet的分类上线进行妥协，将运动对象都划分到人这一类别当中。但在我们的感知中，可能应当对动态障碍物进行更清晰的识别。
- 在运动一致性判断中，物体沿极线运动时无法判断出是否运动。

