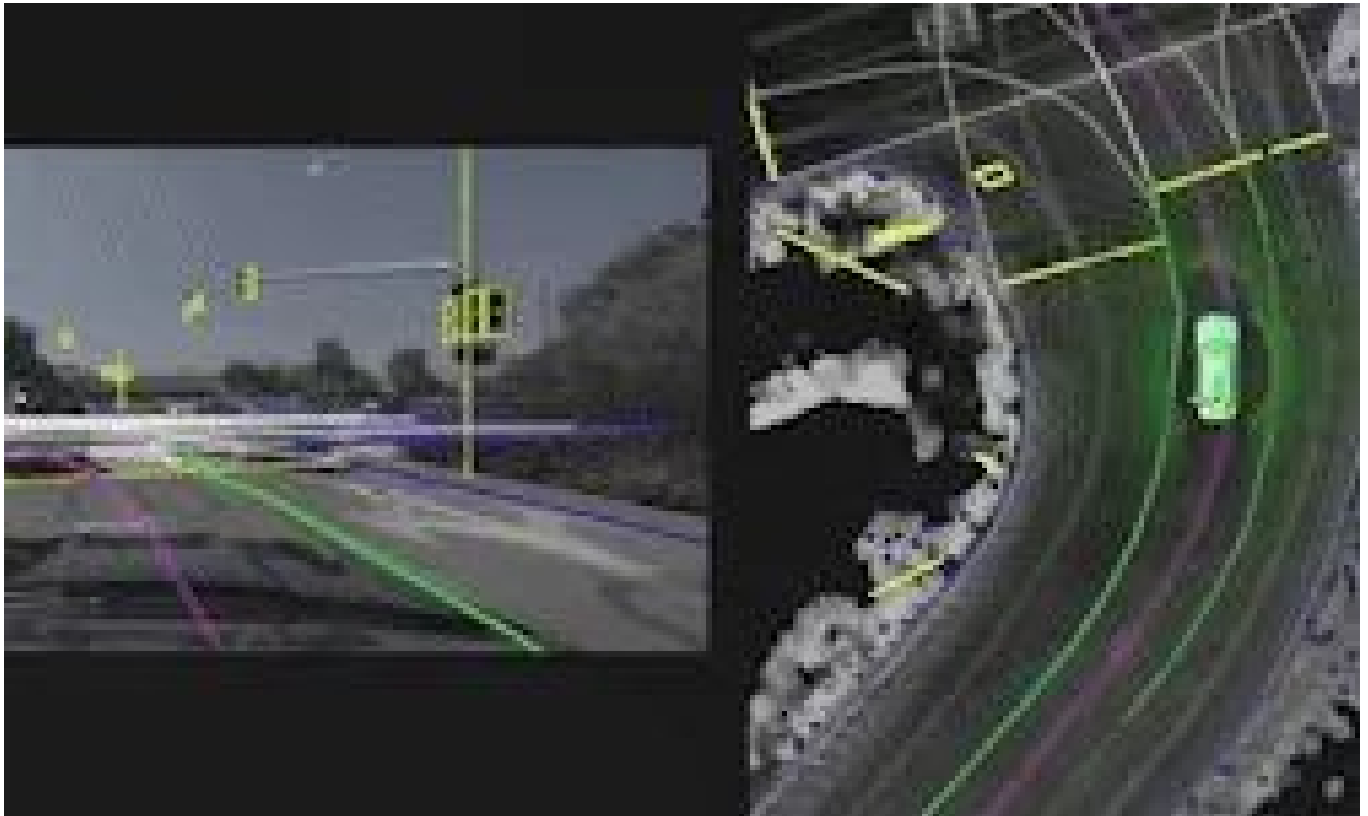


公众号

问题定义

给定一份高精度地图或环境模型，给定相机（不限于单目，双目，多目，针孔或者鱼眼），输出相对于这份地图的高精度的相机的pose,在这个过程中其他传感器信息是可选的(optional)。



<https://news.developer.nvidia.com/drive-labs-how-localization-helps-vehicles-find-their-way/>

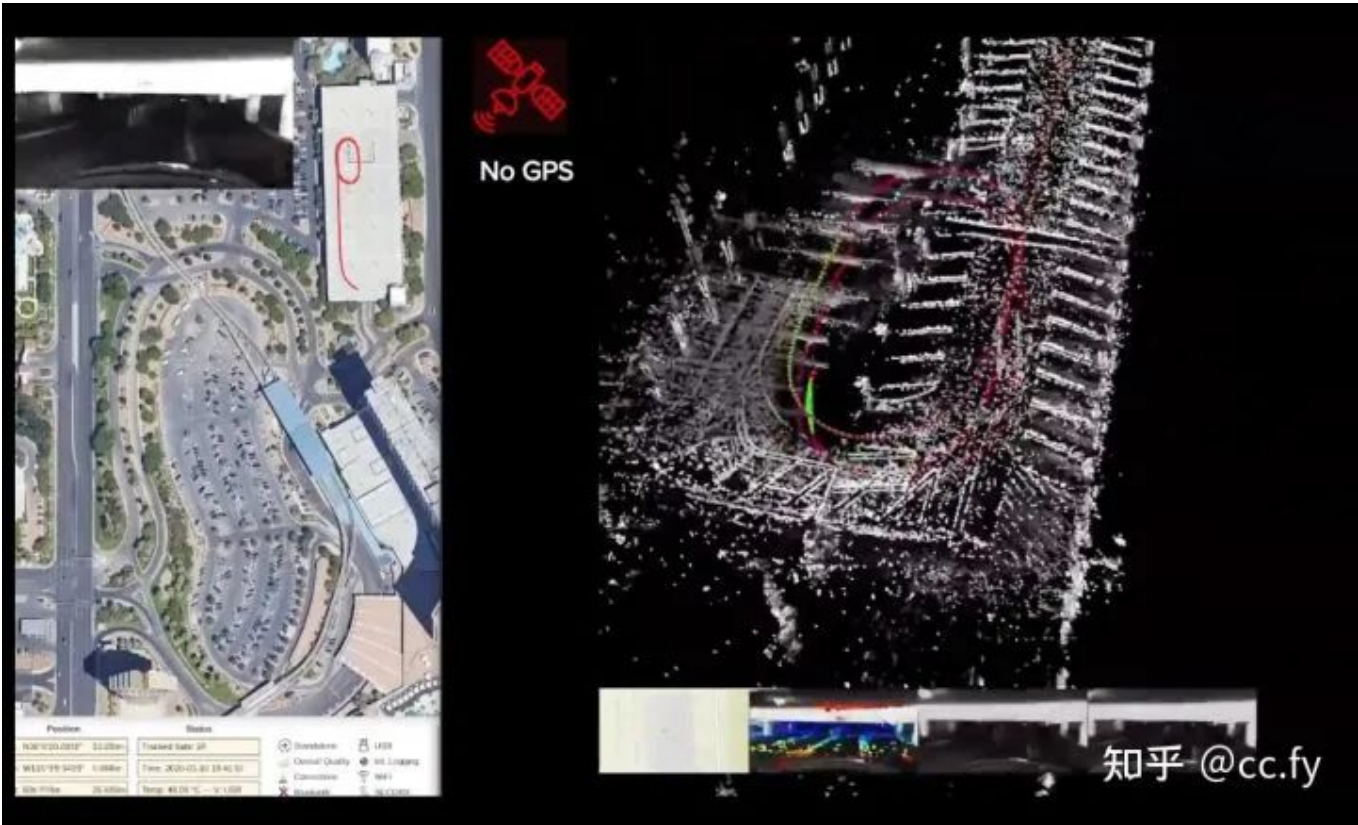
<https://mp.weixin.qq.com/s/04bNTC-CGo1vy6QJF9Jeww>



<https://github.com/ethz-asl/hfnet>

高精地图， 高精定位， 感知的关系

高精地图、高精定位和感知三者相辅相成。假如具备高精地图和感知结果，就可为定位提供先验信息：



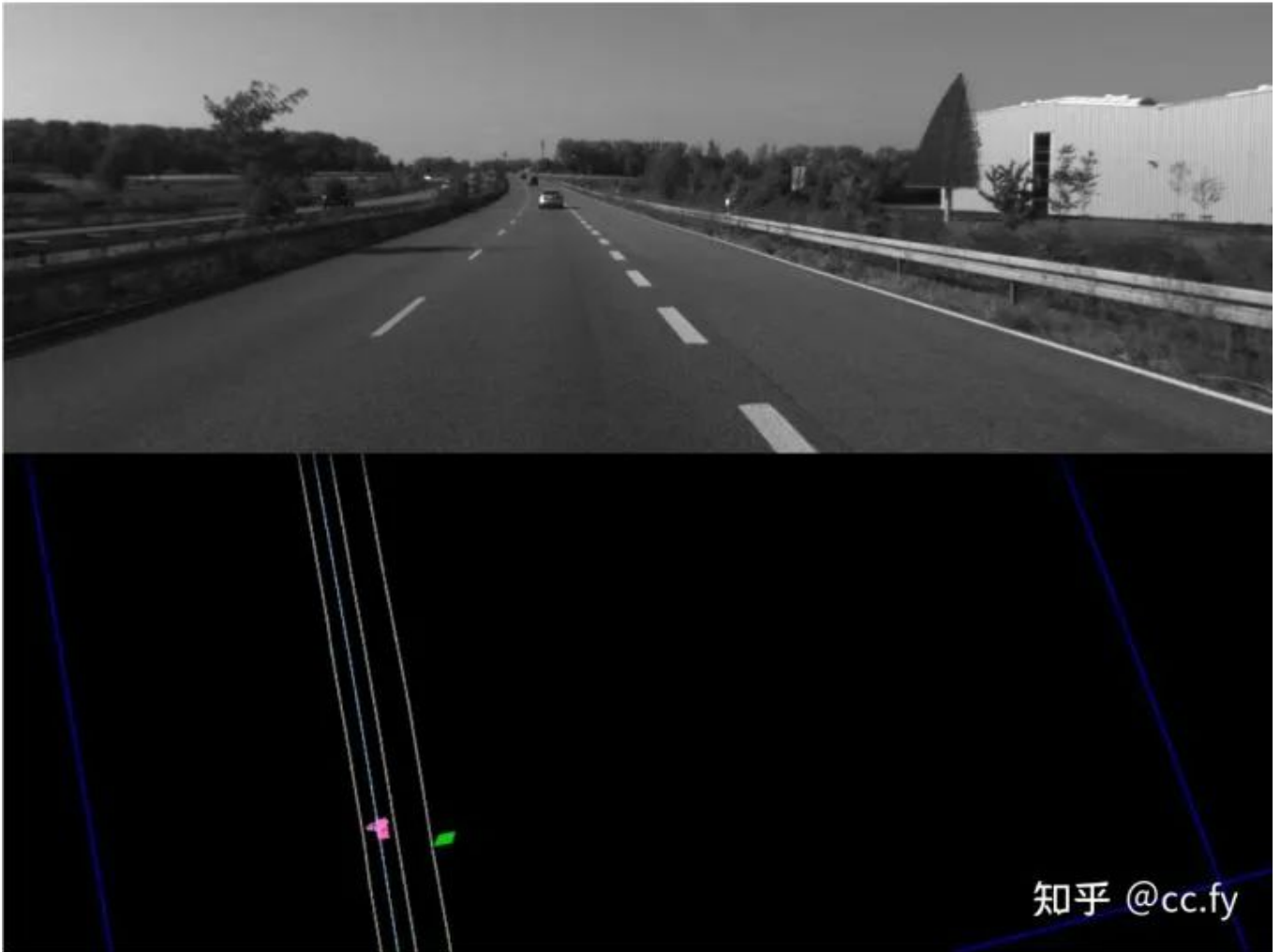
artisense_visual_slam

假如具备高精地图和定位结果，就可把地图元素投影到车体坐标系中，为感知提供先验信息：



<https://www.atlatec.de/localization.html>

假如具备定位和感知结果，就可把感知元素反向投影回地图坐标系，将感知元素与地图已有元素进行比对更新或者构建新的地图：



kitti vector mapping

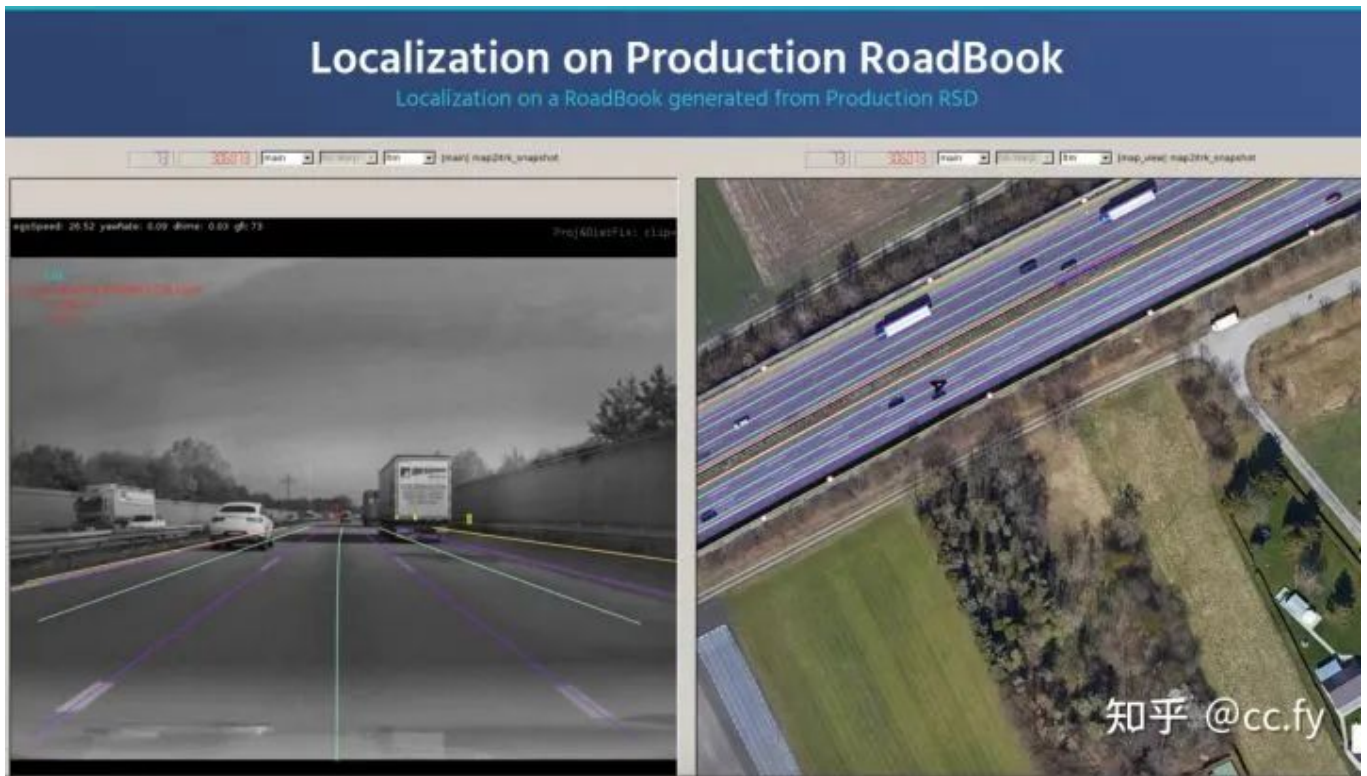
行业现状

NVIDIA

在国内和高德，四维，宽凳合作，在国外和 tomtom, here, zenrin 等合作。DRIVE LOCALIZATION (<https://www.bilibili.com/video/BV1cp4y1e7NM>) , localization within the world robustly and accurately.

Mobileye

Localization in the road book (REM)



上图分别是road book 投影到图像空间上和投影到google earth的结果, 个人觉得mobileye一方面是一家自动驾驶公司, 随着rem系统不断发展, 可能也是一家全球地理数据资产管理公司

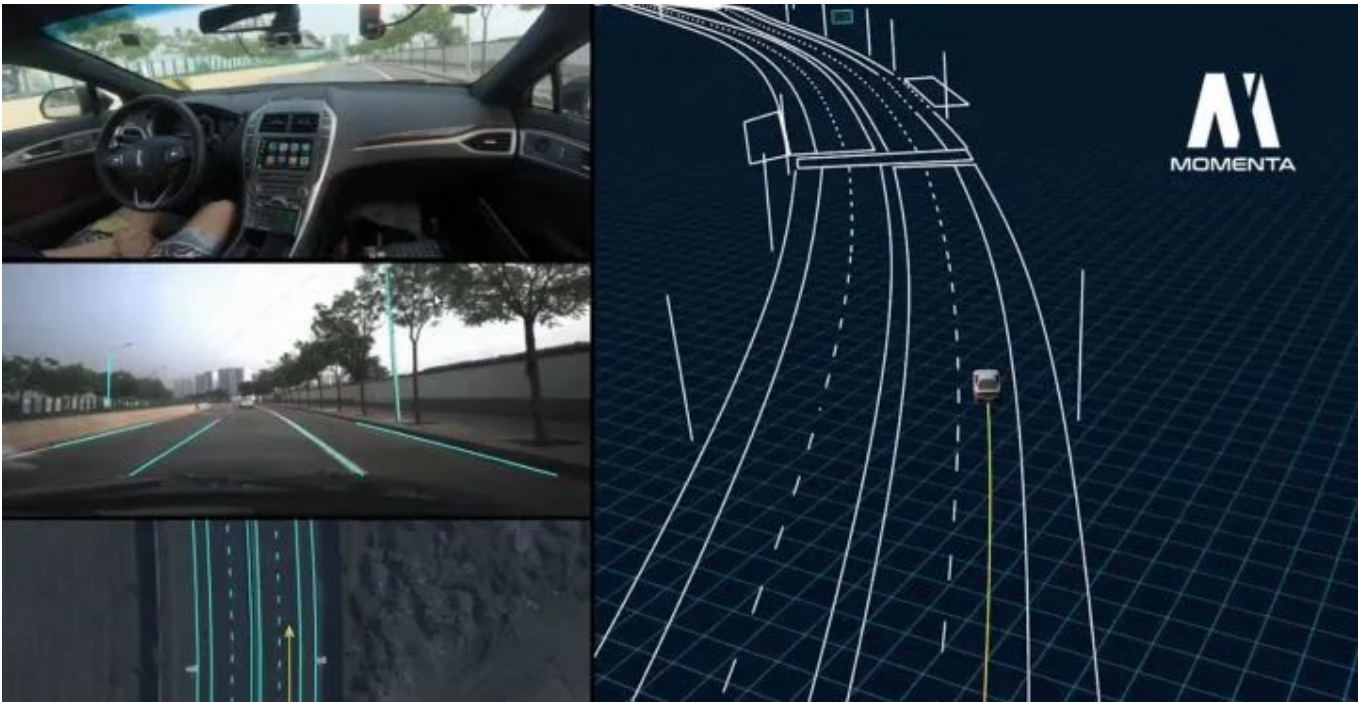
Baidu

DA4AD: End-to-End Deep Attention-based Visual Localization for Autonomous Driving, ECCV 2020

NIO 蔚来

高精地图加持下定位精度20cm: <https://www.bilibili.com/video/BV1Bf4y1D7Lx>

Momenta



Related Paper

1 、 Coarse-to-fine Semantic Localization with HD Map for Autonomous Driving in Structural Scenes (<https://arxiv.org/abs/2107.02557>)

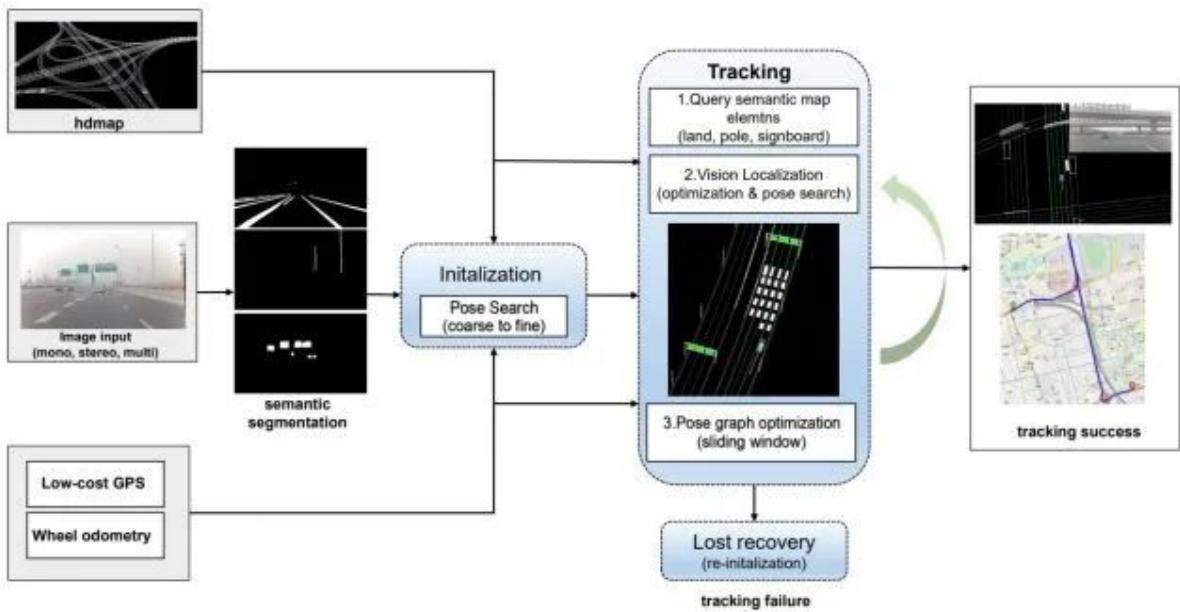


Fig. 2. Overview of the proposed vision localization system. Based on the prior map, low-cost GPS, wheel odometry input and camera signals, the 6-DOF pose can be estimated at centimeter-level accuracy.

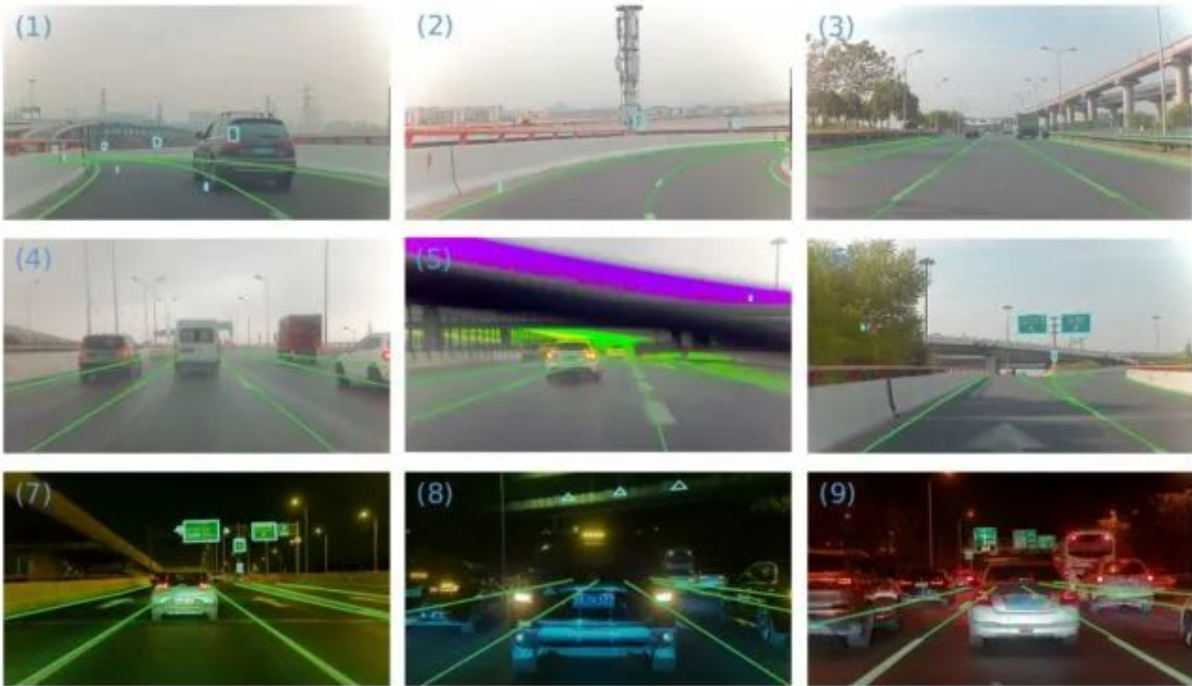


Fig. 6. Qualitative results on Shanghai dataset. Projection results of different scenes: (1-2) curve road; (3) long straight road in sunny day; (4) rainy; (5) windshield wiper blocks part of the image; (6) diverging ramp; (7) low illumination; (8-9) traffic jam.

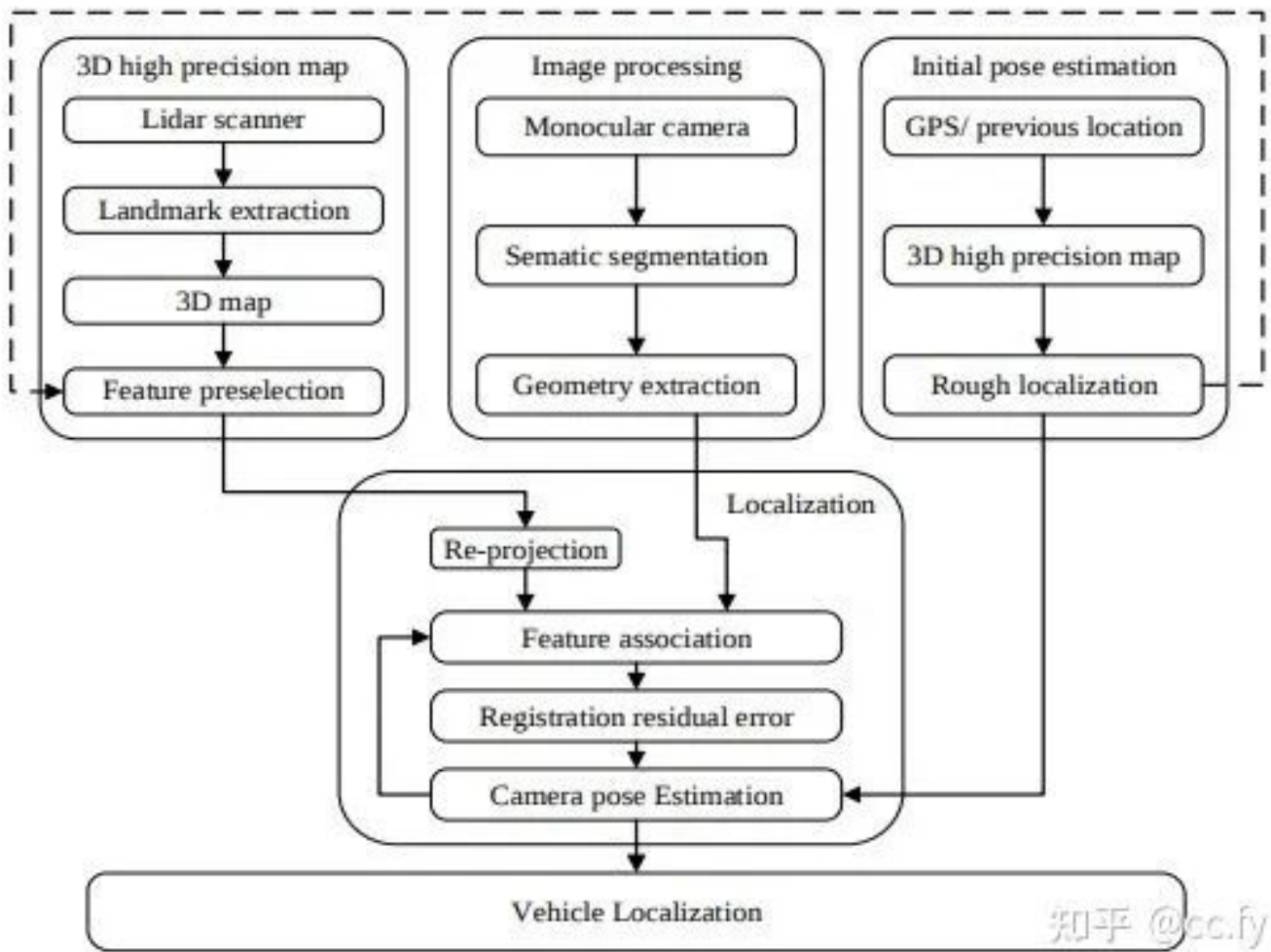


Fig. 7. Qualitative results on Kaist dataset. Left: vector format landmark hdmap and projection result. Middle and right are projection results of two different scenes.

coarse-to-fine semantic localization for autonomous driving

2、Monocular Vehicle Self-localization method based on Compact Semantic Map

Input: 单目相机 + 地图 Output: 6DOF pose, 总体流程图如下：



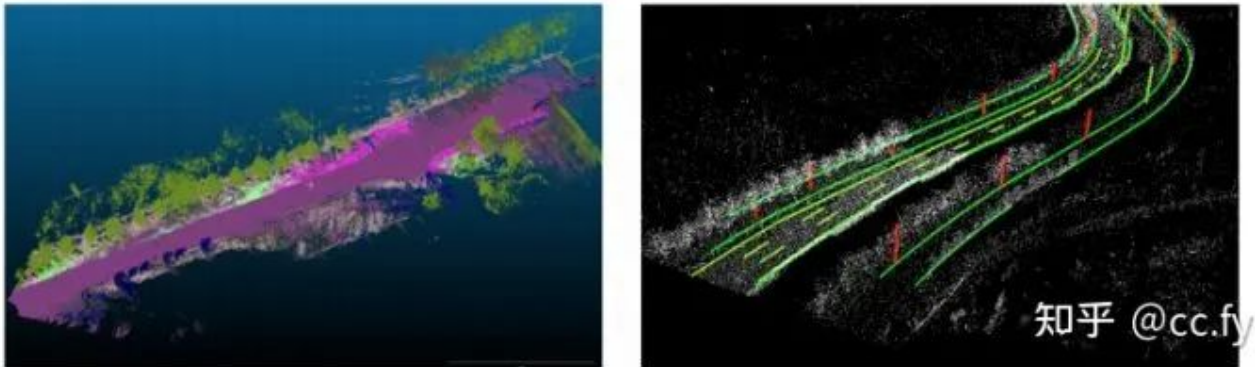
知乎 @cc.fy

3、Monocular Localization with Vector HD Map(MLVHM): A Low Cost Method for Commerial IVs (2020)

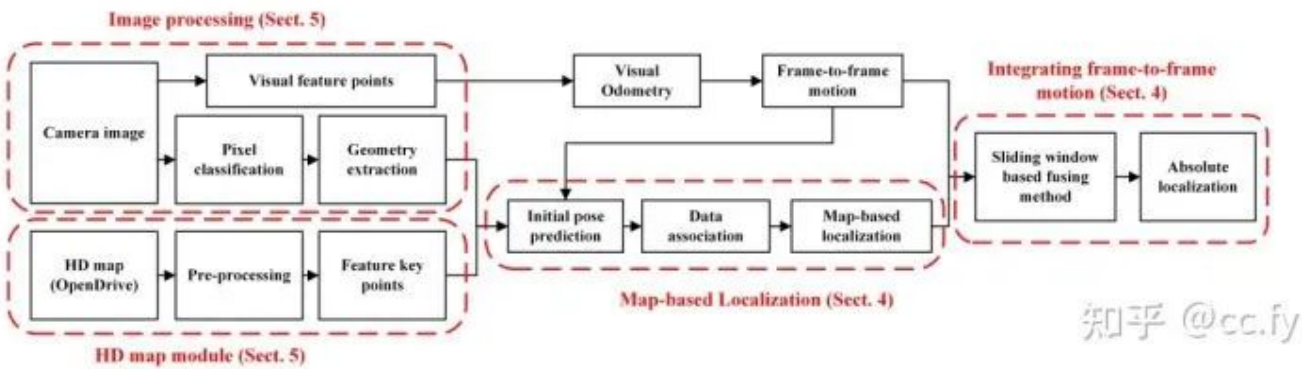
提出了一种耦合单目相机和轻量级矢量地图的定位方法，建立地图特征和图像检测出的相对稳定的语义特征之间的关联，从而进行相机的位姿估计，同时为了增加约束让轨迹更加鲁棒平滑，引入了两帧之间的视觉里程计的约束，最终达到了约24cm的定位的rmse精度。

众所周知，GNSS-RTK + 高质量的IMU + 多线束激光雷达在室内外都可以交付高精度的定位结果，然后由于成本的原因无法大规模商业化使用。然后作者团队也希望能够只需要一个单目相机 + hdmap来完成这件事情，尽管hdmap本身是需要昂贵的传感器来构建，当然也有公司采用众包技术使用消费级传感器就可以进行地图的构建与更新，但是地图一旦构建起来的边际成本接近于0，通过云端分发，分发的车辆越多，其分摊的构建成本也越低，自动驾驶中使用的地图形式也五花八门，如下图，有激光定位需要使用的包含环境原始几何信息的激光点云图，也有如特征点法视觉slam中建立的特征点地图，也有对环境进行拓扑表达的类似于右图中的矢量地图，这些地图都可以完成定位任务，考虑到地图大小，简洁的要求，矢量地图可能是一个更合适的选择，并且在这篇文

章中被使用，当然同样一份矢量地图，国际上存在大量的格式标准，如opendrive, apollo, nds等。如下展示了两两种不同形式的地图：



由于在室外环境中，光照变化，视角变化，遮挡等会频繁发生，所以选择从环境特征中选择一些相对这些条件更加鲁棒的特征是必要的，借助于CNN对图像极强的表达能力，我们可以获得这些稳定的语义特征所在的像素或者描述，而矢量地图中存储的也是这些不变的相对稳定的东西，所以问题很自然的简化为数据关联问题。文章所提出的方法的整体流程如下：



与前作不同的是，同时解算帧间vo（利用了传统视觉特征orb）,并在滑窗中引入帧间的约束来提升鲁棒性。在影像处理模块，通过分割提取关键像素，并将其拟合成point feature(sign)和line feature(lane and pole),同时为了进行帧间vo的计算，会进行orb特征的提取；在基于地图的定位模块会进行初始pose估计，利用初始pose通过随机采样的方式确定最优的匹配，然后根据匹配关系就可以进行pose的优化求解；最后为了确保定位输出的鲁棒，增加了滑窗的位姿优化。

数据关联的过程：确定地图上的特征和图像上的特征的对对应关系

1. 先生成可能的对应关系集合，从集合中随机采样三条同语义的线特征的对对应关系来进行相机pose的计算，

2. 根据计算出的相机的pose投影地图元素，如果和图像feature距离低于设定的阈值则认为是inlier，然后会计算估计出的pose和初始pose的距离，如果满足阈值，则对应关系会加入假设集合
3. 最终选择inlier数最多的一组对应关系作为最终的对应关系

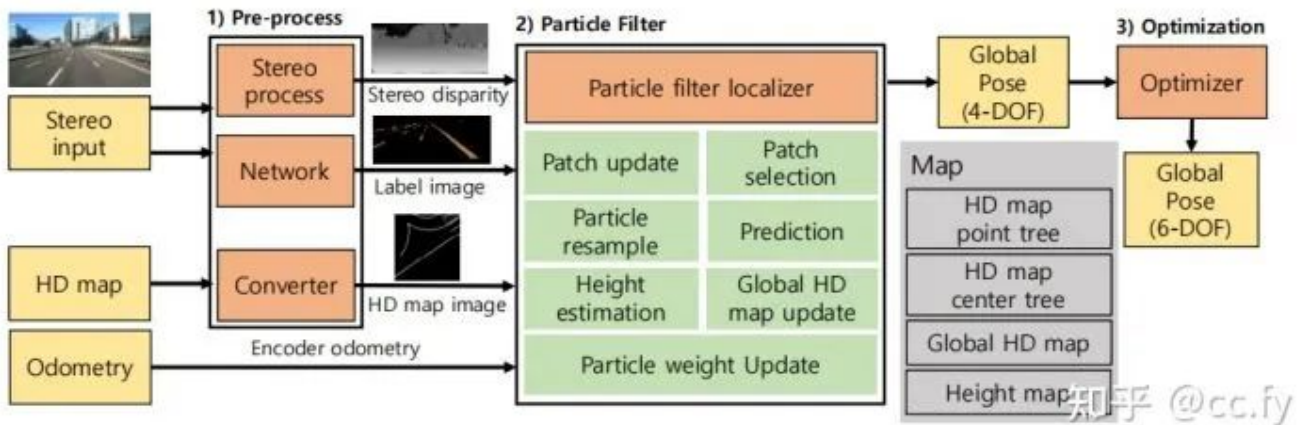
实现细节：

1. 图像特征提取层面，用修改版的pspnet来对图像进行分割，然后根据置信度概率选择特定的像素，用区域生长的方法划分为不同的区域，线特征进行最小二乘拟合，从sign区域拟合点特征，从图像域上得到这些特征分别的位置描述
2. compact map的使用，所使用的地图是opendrive格式，sign使用的是质心点，pole使用的是两个端点，lane用0.2m进行采样散点，然后每两个散点构成拟合成一条线来进行使用
3. 优化与初始化, 系统初始化也是使用了低成本的GNSS的测量值作为初值，优化就是标准的LM算法
4. HDMI_LOC: Exploiting High Definition Map Image for Precise Localization via Bitwise Particle Filter (IROS 2020)

文章提出了一种方法，结合双目相机和hdmap来进行6dof的位姿估计，将地图转换为8bit的影像代表，使得query image可以和地图进行bitwise operation(位运算)来进行匹配，使用粒子滤波框架来进行pose优化估计，11km的测试条件横纵向误差约为0.3m,运行速度为10hz,整个定位过程可以分为4步：

1. 图像处理步骤，从双目图像中得到语义分割图和视差图（所以这个方法比较慢）
2. 将hdmap转化为8 bit image
3. 利用粒子滤波估计4dof（三个平移 + yaw）pose, 通过当前影像的patch和地图影像，因为这两个东西都是8bit影像，所以可以进行位与操作来快速计算
4. 通过优化额外计算roll和pitch来估计完整的6自由度pose

整体流程图如下：



5、From Coarse to Fine: Robust Hierarchical Localization at Large Scale (CVPR 2019)

视觉定位是指，在大尺度场景下，已知环境地图，给定任意一张图片，计算出该图片对应的位姿。

在视觉定位领域，基本上有两种主要的方向，一种是所谓的 End-to-end 方法，代表作是 PoseNet，这类方法直接用图片和 pose 真值作为监督学习，对于给定的同一场景的图片可以回归出 6dof 的真值。这种方法优点是数据库很小，一个网络可以覆盖一个很大的场景，而且虽然需要 Pose 真值但并不需要显式的空间三维重建。但是缺点也很明显，这类方法无非是简单地通过学习建立起了图像和pose之间的映射关系，本质上就是一个图像相似度匹配的过程，因此不可能回归出精确地真值以及获得对于未学习场景的迁移能力。目前比较流行的方法是，由粗到细两步定位。先通过图像检索的方式找到最接近的关键帧，再与该关键帧匹配局部特征。由于关键帧位姿已知，所以通过PnP等方法可以估计出当前帧位姿。两步定位可以避免从所有关键帧中直接匹配带来的时间复杂度，同时避免了将整个环境地图加载进内存带来的空间复杂度。文章主要提出了一种由粗到精的hierarchical的定位网络，hfnet, 对输入的query image, 会通过一个单一的CNN同时预测影像的局部特征和全局特征来进行6自由度的精确定位，采用的这样一种coarse to fine的定位范式: 先通过一个全局的retrieval获取候选帧，然后和候选的位置做局部特征匹配来进行定位。类似于人类定位的模式一样，作者希望利用这样一种coarse to fine的定位范，同时又希望能够结合深度学习在图像特征检测和描述领域的一些进展，这些学习到的detector 和descriptor提高了对环境变化的鲁棒性。为了达到上面的目的，并且最大的提升效率，提出了HF-NET (hierarchical feature network) ,同时估计局部和全局特征，并且最大化的共享计算。通过多任务蒸馏的方式进行网络的训练。

PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization

Alex Kendall

Matthew Grimes

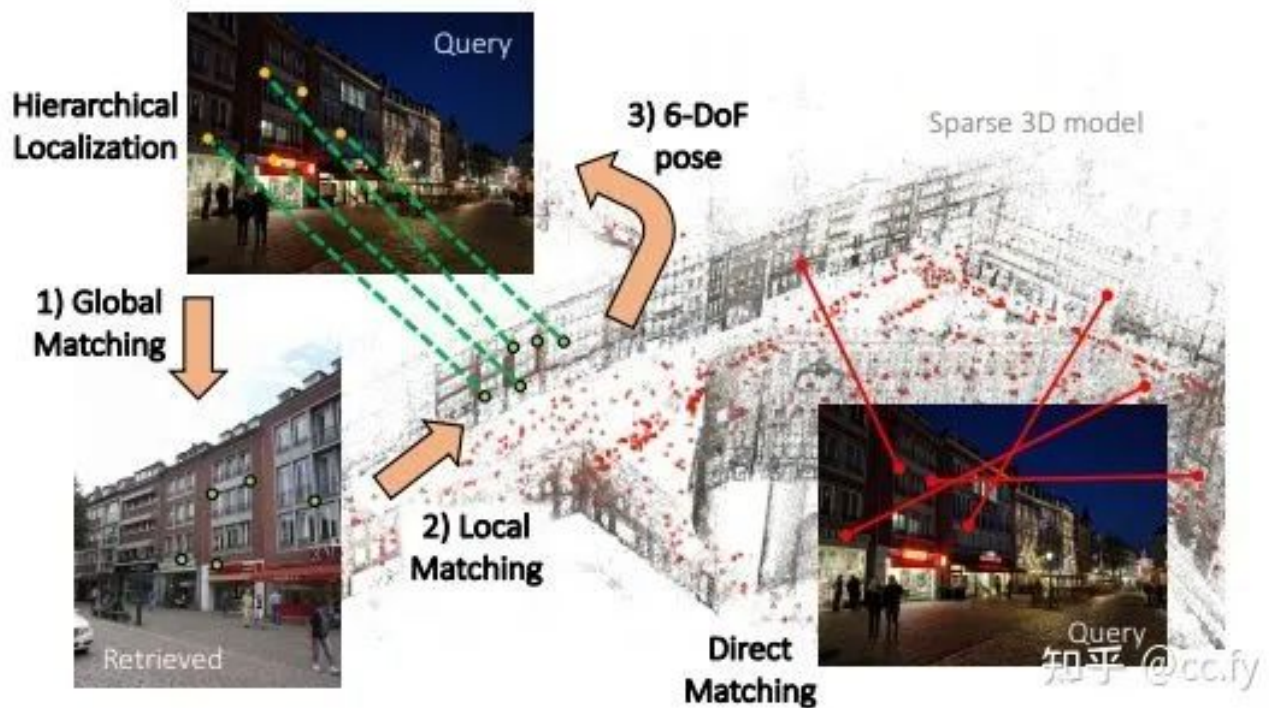
Roberto Cipolla

University of Cambridge

agk34, mkg30, rc10001 @cam.ac.uk

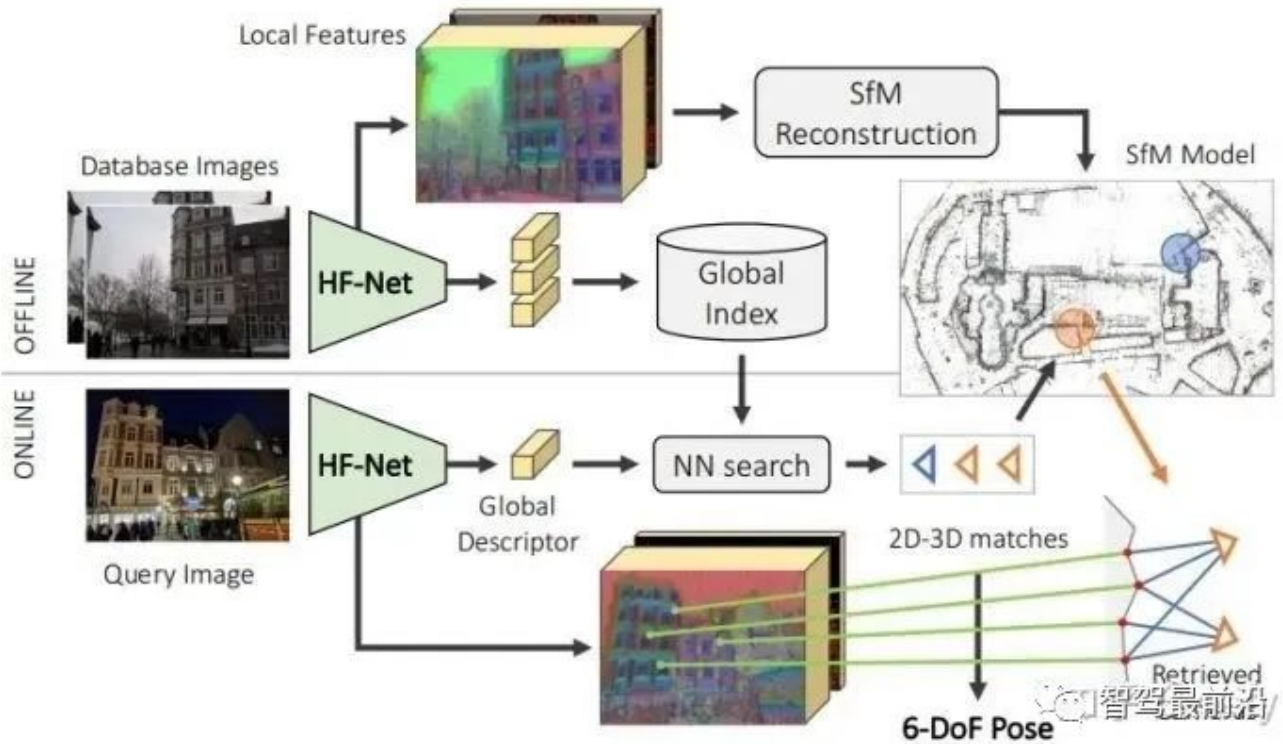


Figure 1: **PoseNet: Convolutional neural network monocular camera relocalization.** Relocalization results for an input image (top), the predicted camera pose of a visual reconstruction (middle), shown again overlaid in red on the original image (bottom). Our system relocalizes to within approximately $2m$ and 3° for large outdoor scenes spanning $50,000m^2$. For an online demonstration, please see our project webpage: mi.eng.cam.ac.uk/projects/relocalisation/



类似于人类定位的模式一样，作者希望利用这样一种coarse to fine的定位范式，同时又希望能够结合深度学习在图像特征检测和描述领域的一些进展，这些学习到的detector 和descriptor提高了对环境变化的鲁棒性。为了达到上面的目的，并且最大的提升效率，提出了HF-NET (hierarchical feature network)，同时估计局部和全局特征，并且最大化的共享计算。通过多任务蒸馏的方式进行网络的训练。

离线模式下，根据db image利用网络提取局部特征和全局特征，局部特征传入到sfm三维重建框架进行环境模型构建，全局特征构建参考图像索引db, 然后在线运行的时候，对query image



做相同的特征提取操作，根据全局描述子利用NN搜索得到候选帧，建立候选帧共视的地图点和query图像特征点的对应关系利用ransac + pnp框架求解pose。

