Data Association in Monocular SLAM with Dirichlet Process

论文情况

- 标题：Object Clustering With Dirichlet Process Mixture Model for Data Association in Monocular SLAM
- 作者：Songlin Wei, Guodong Chen, Wenzheng Chi, Zhenhua Wang, Lining Sun
- 期刊：IEEE Transactions on Industrial Electronics (TIE), 2023
- 源码：未开源

1. Introduction

将检测到的物体与地图 landmarks 正确关联至关重要，模糊的数据关联会导致定位漂移甚至错误回环检测。在单目 SLAM 中，物体通常表示为长方体，连接长方体的困难是双重的：

- 单目图像长方体检测是一个不适定问题，容易产生次优结果；
- 在存在大规模遮挡的场景中，仅使用 2D IoU 或距离的简单方法可能会产生错误的关联。

为了实现单目 SLAM 的鲁棒数据关联，本文提出了一种基于狄利克雷过程聚类的单目 SLAM 数据关联方法，进一步在数据关联中统一目标的位置、形状和语义标签信息：

1. 在狄利克雷过程混合模型聚类的基础上，将概率长方体测量应用于物体关联；
2. 长方体作为新节点包含在 SLAM 的位姿图中，长方体不仅有助于减小尺度漂移，而且可以提高回环性能。

2. Methods

2.1. 3-D Cuboid Estimation in Monocular Images

在 SLAM 中使用长方体表示 3D 物体，3D 长方体表示为 $C = \{R, L, S\}$。其中 $L = [x, y, z]^T$ 表示长方体中心的位置；物体尺度 $S = [a, b, c]^T$ 为在物体坐标系下沿坐标轴的长度，物体坐标系以长方体中心为原点建立；物体类别通过 2D 目标检测 [1] 获得，物体的位置、旋转、缩放则使用 [2] 获取。[3] 中的物体检测方法也可以使用。

2.2. Cuboid Data Association

2.2.1. Dirichlet Process and DPMeans Method

在语义世界建模中，大多数对象在短时间内不会发生变化，即可以假定大多数对象是静态的。因此，测量的时间序列是不相关的。测量-目标关联问题可以归结为聚簇问题。

设 $\mathcal{O} \triangleq \{(c^k, L^k, S^k)\}_{k=1}^M$ 为一个簇中的所有观测，c 为簇内索引为 k 的观测的物体类别。新物体 (c, L, S) 属于某一确定簇的概率为：

$$\begin{aligned} p(c, L, S|\mathcal{O}) &= p(c|\mathcal{O})p(L|\mathcal{O})p(S|\mathcal{O}) \\ &= p(c|\{c^k\})p(L|\{L^k\})p(S|\{S^k\}) \end{aligned} \tag{1}$$

(1) 右边第一项是给定同一对象的所有过去观测值的类别预测概率，第二、第三项是姿态和尺度的预测概率。设类别集合为 $\mathcal{C} = \{1, \ldots, C\}$，则簇的类别 c' 的后验概率为：

$$p(c'|\{c^k\}) \propto p(\{c^k\}|c')p(c') = \left[\prod_k p(c^k|c')\right] p(c') \tag{2}$$

$p(c')$ 为先验，$p(c^k|c')$ 为类别测量概率。由此：

$$p(c|\{c^k\}) = \sum_{i=1}^{C} p(c|c_i')p(c_i'|\{c^k\}) \tag{3}$$

$p(c_i'|\{c^k\})$ 为从 (2) 得到的后验概率，$p(c|c_i')$ 为类别测量概率。在静态世界假设下，物体位置的测量 $[x, y, z]^T$ 和尺度测量 $[a, b, c]^T$ 服从未知均值和方差的高斯分布。为简化，设位置 x, y, z 和尺度 a, b, c 相互独立：

$$p(L|\{L^k\}) = p(x|\{x^k\})p(y|\{y^k\})p(z|\{z^k\}) \tag{4}$$

$$p(S|\{S^k\}) = p(a|\{a^k\})p(b|\{b^k\})p(c|\{c^k\}) \tag{5}$$

如图-3，每个观测都假设为独立变量的高斯分布。

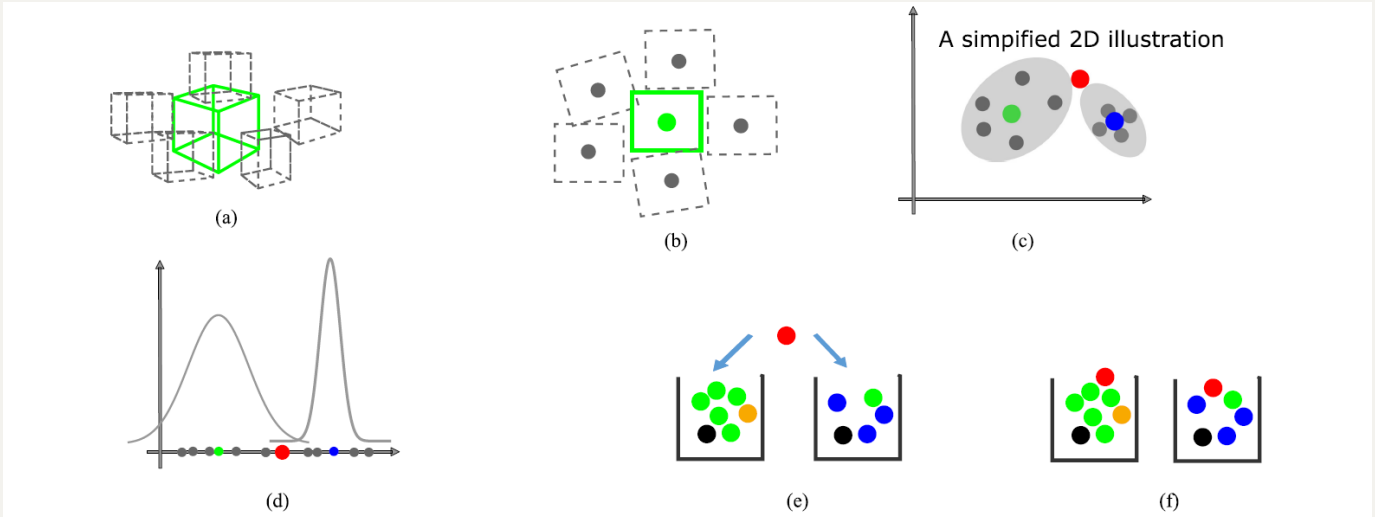

Fig. 3. Here is a brief illustration of how the posterior predictive probability of fitting a new cuboid observation into existing clusters is calculated. (a) Several cuboid observations forming a cluster due to perception noises. (b) Top view of the cluster with dots represents the cuboid center positions. (c) Simplified 2-D illustration of how to fit a new cuboid (represented as a red dot) into each of the exiting two clusters modeled as 2-D Gaussian distributions centered at green and blues dots, respectively. (d) We further assume each component of location (x,y,z) and shape scales (a,b,c) are independent, then each component is fitting into 1-D Gaussian distribution with unknown mean and variance. Finally, (e) and (f) show how the class of the new cuboid is fitting into history observed classes by the analogy of putting a new colored ball into two each urn with noises (represented as black and yellow dots).

与简单假设固定方差不同，对均值和精度未知的高斯噪声模型使用标准共轭先验，即 $NormalGamma(l, \tau; \lambda, \nu, \alpha, \beta)$ 分布。此概率为有两个变量和四个参数的连续分布。均值 l 的边缘分布为非标准的 t-student 分布：

$$p(l|\{x\}; \lambda, \nu, \alpha, \beta) = t\left(l; 2\alpha', \nu', \sqrt{\frac{\beta'}{\lambda'\alpha'}}\right) \tag{6}$$

下一位置观测 x 的后验预测分布：

$$p(x|\{x\}; \lambda, \nu, \alpha, \beta) = \frac{1}{\sqrt{2\pi}} \frac{\beta^{-\alpha^-}}{\beta^{+\alpha^+}} \frac{\lambda^-}{\lambda^+} \frac{\Gamma(\alpha^+)}{\Gamma(\alpha^-)} \tag{7}$$

其中带有 $-$ 上标的超参数是不包括 x 的前一个值，而 $+$ 是 x 的当前 n 个观测值根据 (8)-(11) 中更新规则的更新值，：

$$\lambda' = \lambda + n \tag{8}$$

$$\nu' = \frac{\lambda}{\lambda + n}\nu + \frac{n}{\lambda + n}\hat{\mu} \tag{9}$$

$$\alpha' = \alpha + \frac{n}{2} \tag{10}$$

$$\beta' = \beta + \frac{1}{2}\left(n\hat{s}^2 + \frac{\lambda n}{\lambda + n}(\hat{\mu} - \nu)^2\right) \tag{11}$$

$\hat{\mu}$ 和 $\hat{s}^2$ 为采样均值和方差。时间 t 一个新的立方体测量 k 可以通过将 (7) 的每个分量带入 (4) (5)，将 (3) (4) (5) 带入 (1) 进行计算。将测量 k 分配给具有最大概率的簇 j，如果最大概率低于某个阈值 Υ，则创建一个新的簇。关联结果表示为 $(z_{t,k}, j)$ 并存入数据关联集合 $\mathbf{D}$ 。

2.2.2. Robust SLAM

维护一个目标置信度传播过程，以评估与实际目标相对应的每个测量簇的置信度。直觉上说，一个簇越有可能是一个真实物体，测量结果就越一致。对于每个目标，在给定一组观测的情况下，目标的最终置信度为：

$$bel(\mathcal{O}) = (1 - p(0|\{c^k\})) \prod_{x,y,z,a,b,c} t\left(l; 2\alpha', \nu', \sqrt{\frac{\beta'}{\lambda'\alpha'}}\right) \tag{12}$$

(12) 右边第一部分为真实的检测概率，第二部分来自 (6)。只有置信度超过阈值 Φ 的目标则认为是有效的。

2.3. Pose Graph Optimization

对于已知数据关联 $\mathbf{D}=\{(z_{t,k},j)\}$，所有需要优化的变量都是连续的。给定目标观测长方体 $\mathbf{Z}=\{\{(Z_{t,k},S_{t,k})\}_{k=1}^{K}\}_{t=1}^{T}$（$Z_{t,k}$ 为时间 t 第 k 个测量的长方体中心），执行长方体 landmarks $\mathbf{L}=\{(L,S)_{j=1}^{M}\}$ 和相机轨迹 $\mathbf{X}=\{X_t\}_{t=1}^{T}$ 的联合优化：

$$\mathbf{X},\mathbf{L}=\arg\min_{\mathbf{X},\mathbf{L}} p(\mathbf{X},\mathbf{L}|\mathbf{D},\mathbf{Z}) \tag{13}$$

(13) 等价于 BA 问题，可以使用 g2o 进行优化求解。BA 问题描述为如下的最小二乘问题：

$$\begin{aligned}\hat{\mathbf{X}},\hat{\mathbf{L}}=\arg\min_{\mathbf{X},\mathbf{L}} &\sum_{t,k}\|e(Z_{t,k};X_t,L_j)\|^2 \\ &+\sum_{i,j}\|e(S_j;P_i,L_j)\|^2+\sum_{t,i}\|e(p_{t,i};P_i,X_t)\|^2\end{aligned} \tag{14}$$

$e(Z_{t,k};X_t,L_j)$ 为相机-长方体测量误差；$e(S_j;P_i,L_j)$ 是长方体-点测量误差，将地图点转换到长方体坐标系并当地图点在长方体外时产生惩罚；$e(p_{t,i};P_i,X_t)$ 为 3D 地图点到 2D 关键点的重投影误差。

2.4. Final Algorithm

Algorithm 1: Dirichlet Process Object Clustering For Cuboid Data Association.

Input: Odometry measurements $\mathbf{O}_{1:T}$, Cuboid measurements $\mathbf{Z}_{t,k}$

Output: Poses $\mathbf{X}_{1:T}$, Landmarks $\mathbf{L}_{1:M}$, Data Associations $\mathbf{D}$

Set $\mathbf{X}_0$ to Identity matrix representing world frame, Initialize $\mathbf{X}_{1:T}$, $\mathbf{L}$ with open loop predictions

while $\mathbf{X},\mathbf{L}$ *not converged* **do**

  //phase one, data association:

  set $M=0,\mathbf{L}$

  for *each* t *in* $1:T$, *each* k *in* $1:K_t$ **do**

    Compute the predictive probability of fitting current measurement into each cluster according to (1):

```
    for i in 1 : M do
        pred(i) = −log p(c_{t,k}, X_t, S_{t,k} | O_i)
    Assign D_{t,k} to be cluster m with maximum
     probability or create a new cluster if the best
     predictive probability is below some threshold:
    if min(pred) > ϒ then
        M = M + 1
        create new cluster L_M with current
         measurement, and initialize β_M = β_0
    else
        D_{t,k} = argmin_i pred(i)
        update existing cluster L_i:
        O_i = O_i ∪ Z_{t,k}
        β_i(c_{t,k}) = β_i(c_{t,k}) + 1
        update hyperparameters λ, ν, α, β according
         to (8)-(11)
        update bel(i) according to (12)
//phase two, joint optimization of landmark
 positions and trajectory
for each l in 1:M do
    if bel(l) < Φ then
        mark L_l as invalid, and remove from L
         afterwards
optimize X_{0:T}, L with pose graph solver according
 to (14)
update X, prepare for next loop
```

初始化阶段，里程计数据 $\mathbf{O}_{1:T}$ 确定所有 landmark 观测 $\mathbf{L}$ 的位置。然后，进行数据关联，如 2.2.1。通过在数据关联和位姿优化之间交替，修正因累积误差导致的错误关联。对于每个新的长方体观测，根据 (1) 计算基于现有簇 $\mathcal{O}_i$ 的预测概率 $pred(i)$，如果最大概率低于预设阈值，则创建一个新簇；否则，该观测将分配给现有的簇。重复此过程，直到处理完所有观测。接下来是姿态优化。优化完成后，更新包括尺度和位置在内的物体外观。具体来说，目标位置 L 用联合优化后的结果更新，目标尺度 S 用 (9) 更新。对象（簇）的最终数量在几次迭代后收敛。

3. Experiments

- 代码：基于 ORB-SLAM2

3.1. Simulated Dataset

- 目标设定：15 个物体，5 个类别，4m 范围内的物体能被检测到
- 数据量：$10 \times 10\text{m}^2$ 房间内采集 800 个时间步
- 噪声：对目标位置和里程信息加入高斯噪声，对目标类别按表-1 加入预测噪声

TABLE I
CONFUSION MATRIX USED IN SIMULATION

Actual Label	Predicted Label				
	class 1	class 2	class 3	class 4	class 5
class 1	0.8	0.06	0.04	0.04	0.01
class 2	0.06	0.78	0.05	0.04	0.02
class 3	0.09	0.03	0.77	0.05	0.01
class 4	0.09	0.03	0.06	0.75	0.02
class 5	0.04	0.03	0.07	0.02	0.79

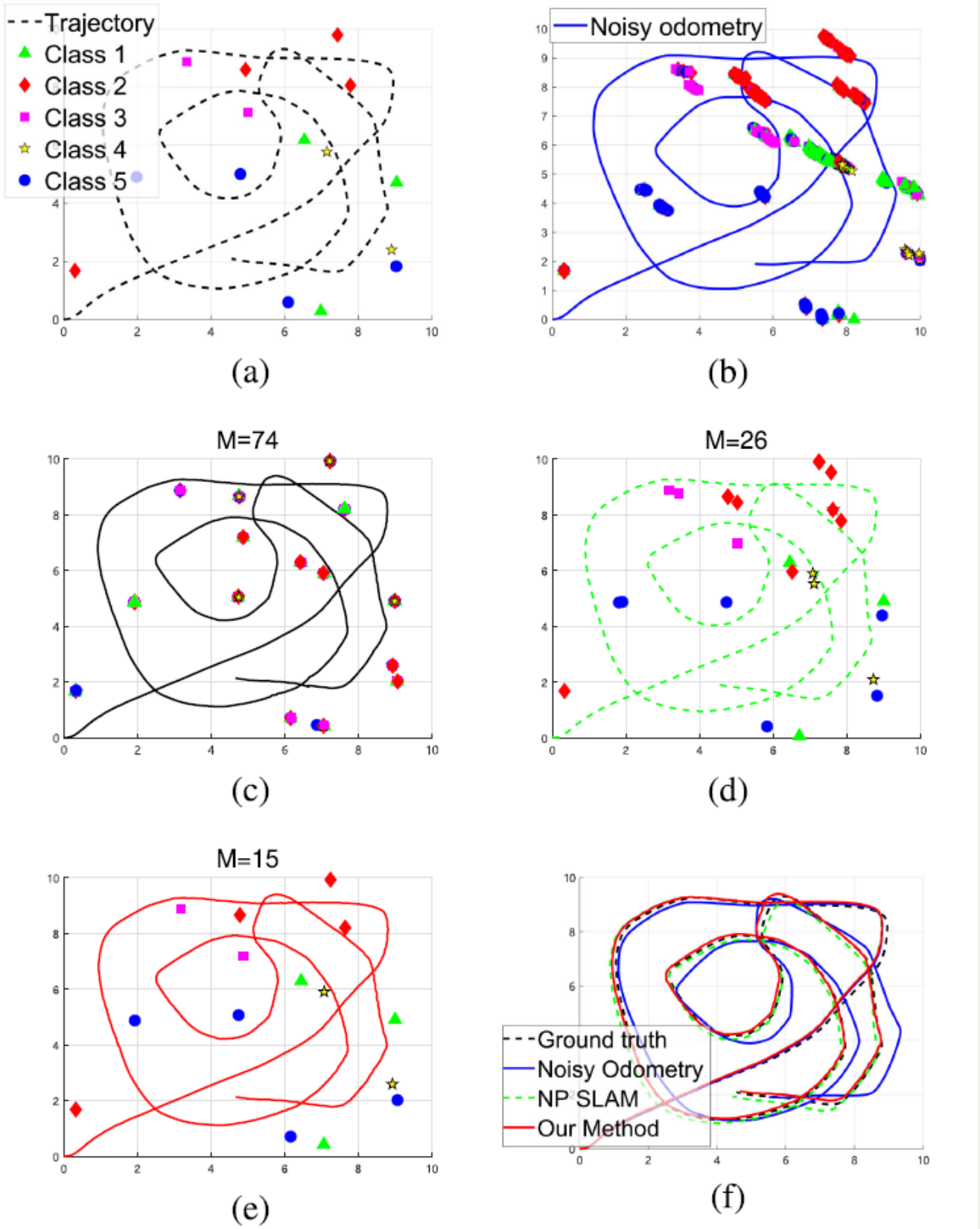

Fig. 5. (a) Ground truth robot trajectory and classes and positions of landmarks. (b) Simulated noisy trajectory, and the open-loop perception of landmarks without data association. (c) Result of maximum likelihood. (d) Result of NP-SLAM [27]. (e) Data association result of our method. (f) Compares trajectories of different methods.

3.2. Turtlebot3 Burger Collected Dataset

如图-8，为目标检测和关联的结果：

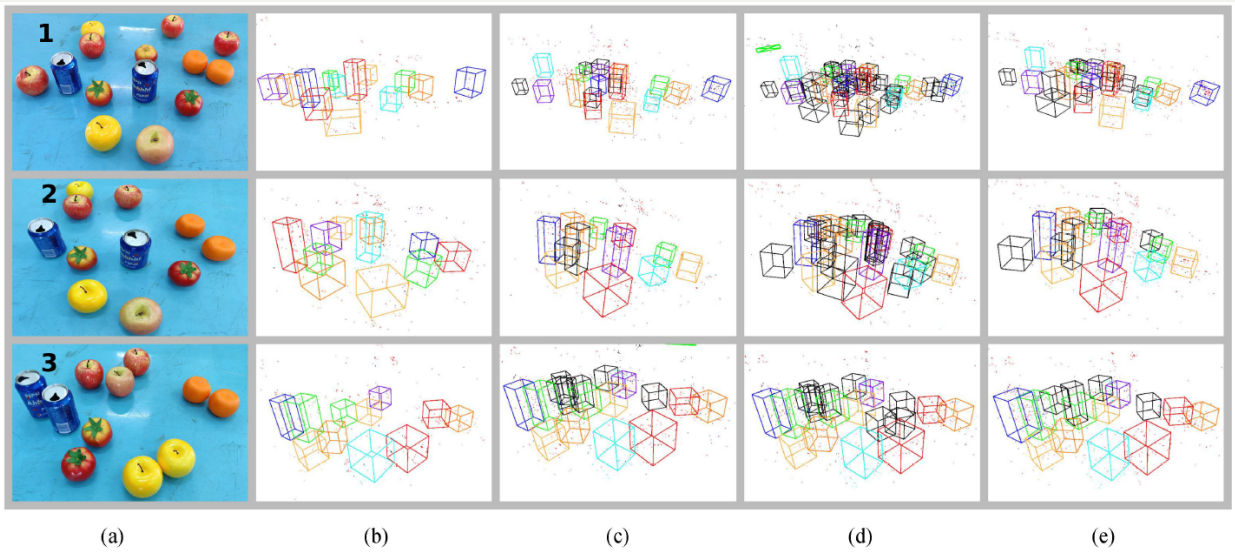

Fig. 8. Cuboid mapping results for different methods. First column (a) shows the picture of the configuration of items. Other columns from left to right, shows the result of Ours in column (b). Intersection over Union (IOU) in column (c). Shared map points counting (SMP) in column (d). Maximum likelihood estimation (MLE) in column (e) for different three configurations of items.

建图（使用 YOLO2）结果：

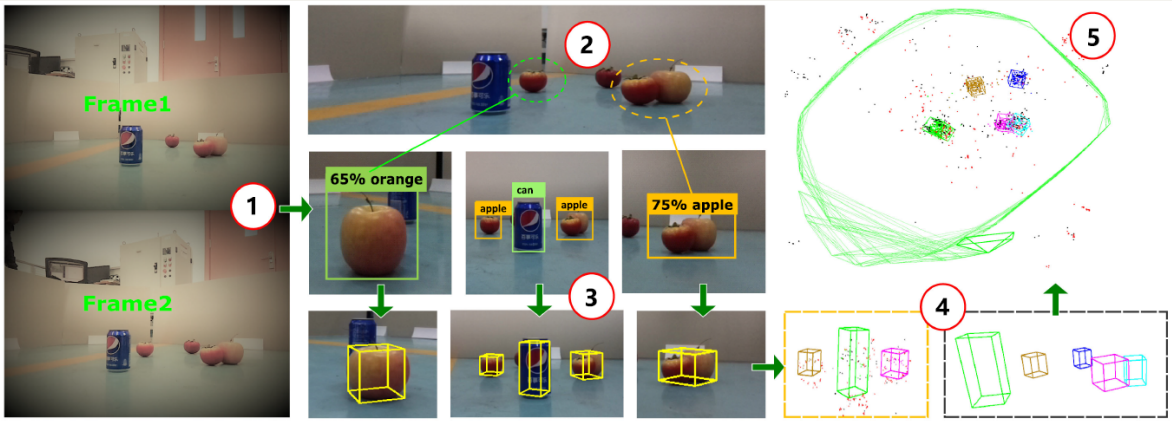

Fig. 2. Key steps of our system. (1) Visual odometry with consecutive frames. (2) Detect 2D bounding boxes with start-of-the-art detectors YOLO2. (3) Detect 3D cuboids using 2D boxes and odometry based in Vanishing Point technique. (4) Perform scale alignment with 3D map points and data association of cuboid landmarks. (5) Consistent mapping with semantic objects as 3D cuboids along with sparse map points.

物体类别识别误差：

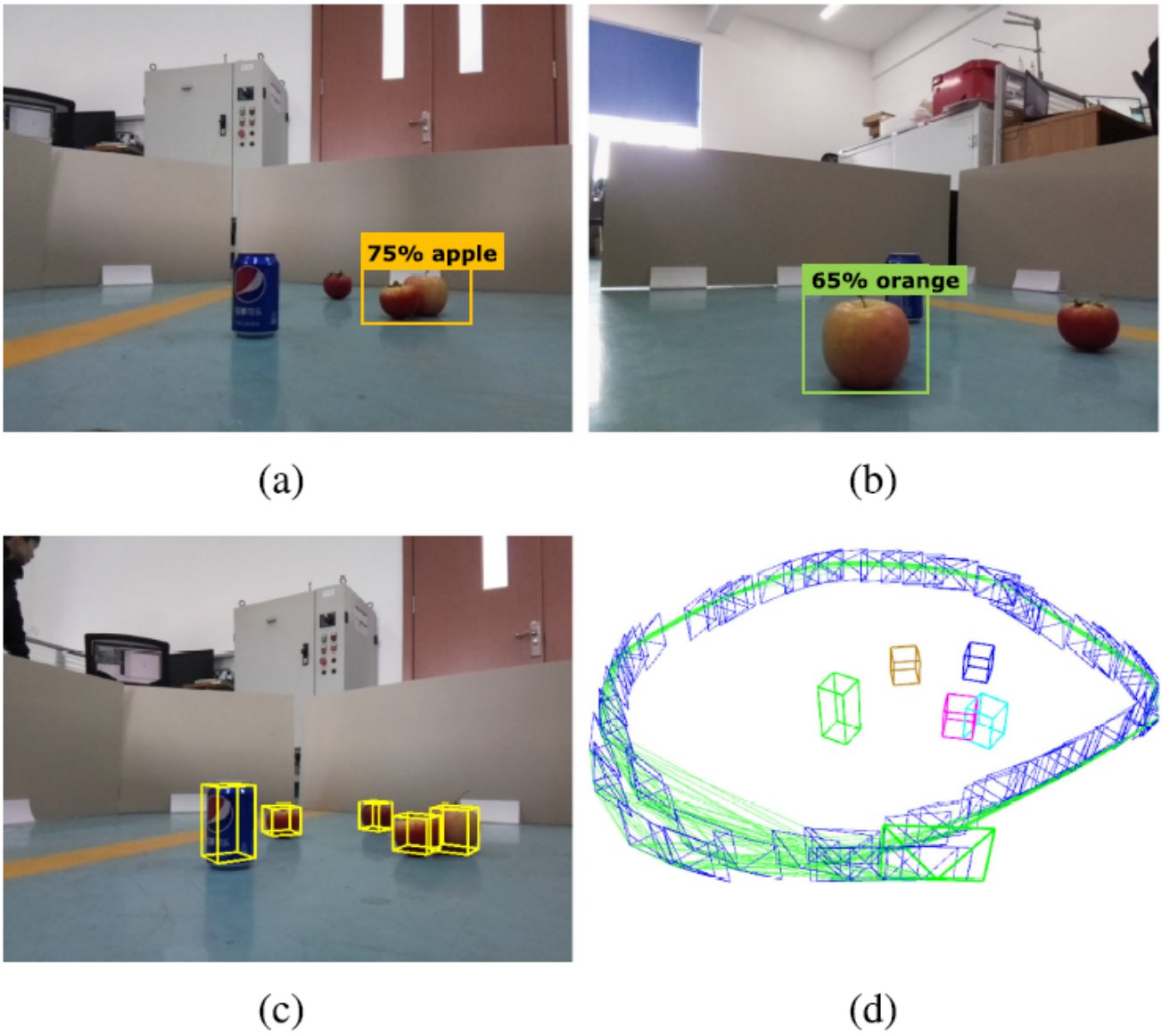

Fig. 1. (a) Two apples are closely placed together, the 2-D object detector failed to separate them. (b) Object detector occasionally produces a false class prediction. For example, it outputs an orange label for an apple. (c) Our data association method can handle the aforementioned situations and yields the right cuboid detections. (d) Optimized camera trajectory along with detected cuboid classes, shapes, and positions.

三种数据融合方式对比（IoU，共享地图点 Shared Map Points - SMP，极大似然估计 MLE）：

TABLE II
COMPARISON OF DIFFERENT METHOD'S DATA ASSOCIATION RESULTS FOR THREE DIFFERENT CONFIGURATION OF ITEMS

Dataset	IOU		SMP		MLE		Ours	Ground Truth
	Before Filter	After Filter	Before Filter	After Filter	Before Filter	After Filter		
Conf. 1	25	19	43	29	33	28	13	14
Conf. 2	16	13	33	23	18	16	11	11
Conf. 3	20	17	26	19	24	15	11	11

阈值设定：

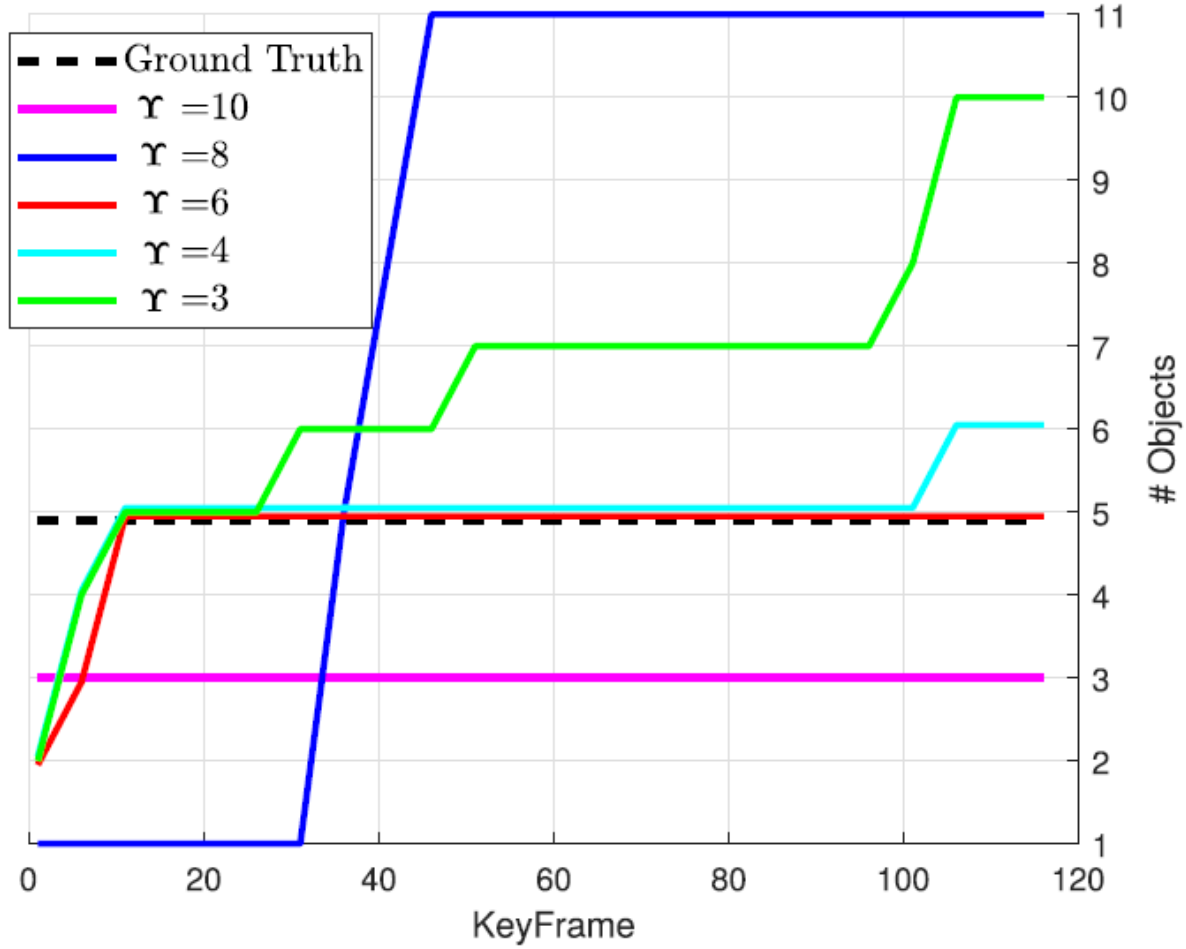

Fig. 6. Data association accuracy analysis with different Υ settings in the Turtlebot3 indoor experiment. Ground truth number of objects is 5 shown in dotted black line. Best association accuracy is achieved when $\Upsilon = 6$.

3.3. KITTI Odometry Dataset

- 阈值：$\Upsilon = 6$

建图结果：

Fig. 7. Mapping result (top view) for the KITTI odometry sequence 07. In addition to the black map points and green KeyFrames, detected cars are rendered as colored rectangles. Seven colors are repeatedly used to tint the cuboid for better visualization. KeyFrame is selected to demonstrate the data association between cuboid observations and existing objects.

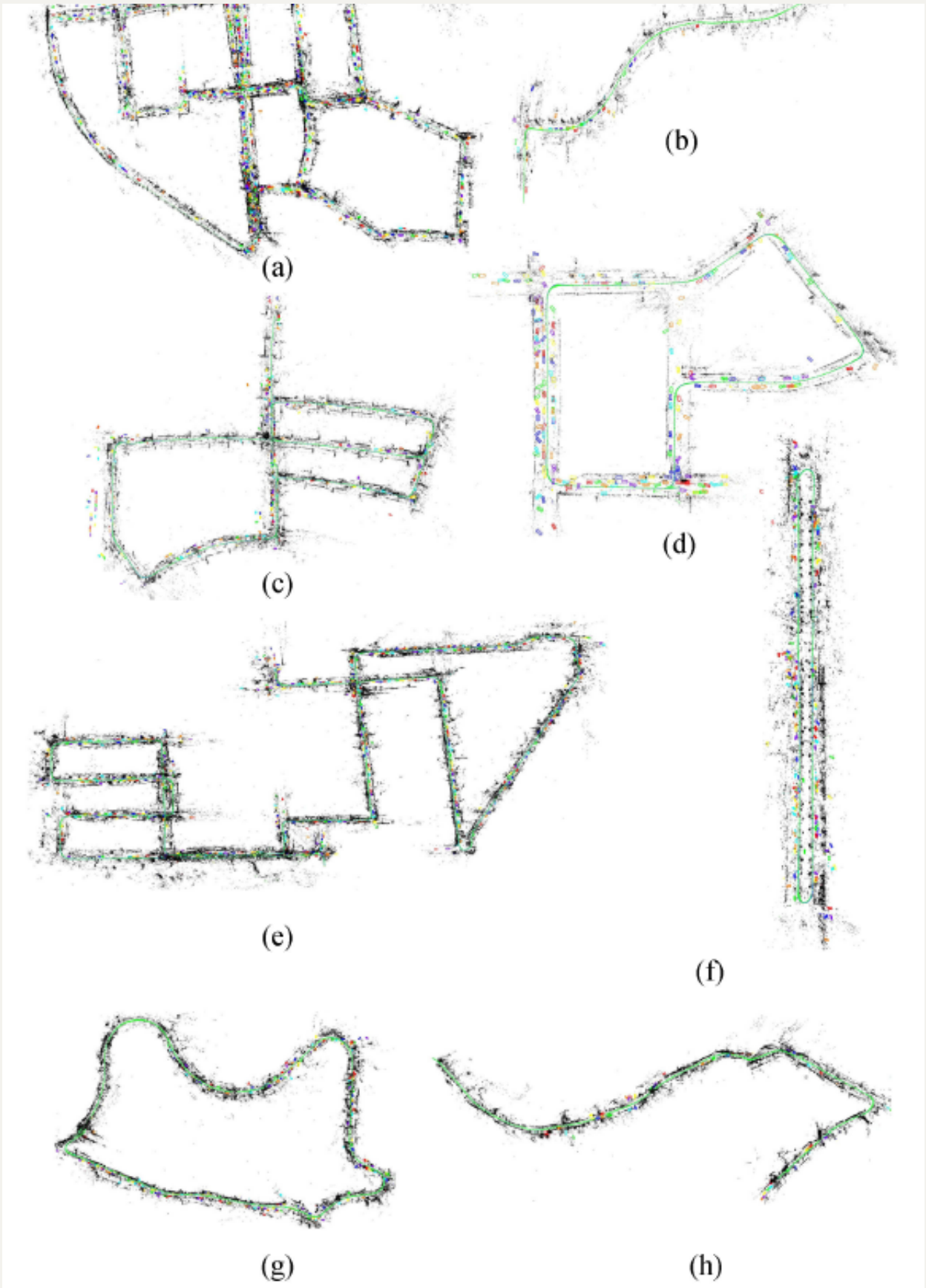

Fig. 10. Cuboid mapping results for KITTI odometry sequences. All the maps are resized to save space. Colored cuboids are so small compared to the full trajectory that they look as if they are dots. (a) 00. (b) 03 .(c) 05. (d) 07. (e) 08. (f) 06. (g) 09. (h) 10.

误差：

TABLE II
COMPARISON OF DIFFERENT METHOD'S DATA ASSOCIATION RESULTS FOR THREE DIFFERENT CONFIGURATION OF ITEMS

Dataset	IOU		SMP		MLE		Ours	Ground Truth
	Before Filter	After Filter	Before Filter	After Filter	Before Filter	After Filter		
Conf. 1	25	19	43	29	33	28	13	14
Conf. 2	16	13	33	23	18	16	11	11
Conf. 3	20	17	26	19	24	15	11	11

3.4. Time Analysis

TABLE V
RUNTIME BREAKDOWN FOR OUR SYSTEM

Component	Tracking	Cuboid Detection	Data Association	Joint Optimization
Time (ms)	33.5	87.9	26.2	284.7

参考文献

- [1] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 7263–7271.
- [2] S.Yang and S. Scherer, "CubeSLAM:Monocular 3D object SLAM," IEEE Trans. Robot., vol. 35, no. 4, pp. 925–938, Aug. 2019.
- [3] S.Yang and S. Scherer, "CubeSLAM:Monocular 3D object SLAM," IEEE Trans. Robot., vol. 35, no. 4, pp. 925–938, Aug. 2019.