1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

S^{3}LAM: Structured Scene SLAM

Mathieu Gonzalez Eric Marchand Amine Kacete Jerome Royan

IROS, 2022
Not Open Source

May 13th, 2023

Outline

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

Brief Introduction of Semantic SLAM

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

Map Forms:

- Sparse, semi-dense (purely geometric, lacks semantic information)
- Semantic (provide supprots to applications use the map)

CNN(object detection and segmentation) assists semantic mapping, which makes **objects as high level landmarks**:

- Some methods requires a specialised object pose estimation algorithm;
- Some works represent objects in a generic way using quadrics or 3D bounding boxes and use a generic object detector.

Introduction to S^{3}LAM

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

S^{3}LAM:

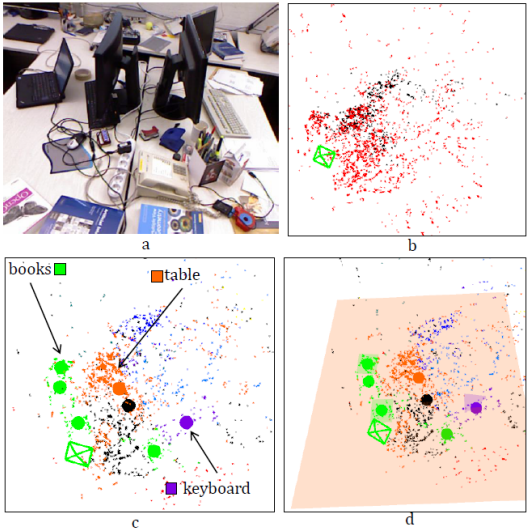
- A monocular SLAM system based on ORB-SLAM2;
- Use panoptic segmentation CNN(Detectron2 [1]) to segment generic objects;
- Objects are seen as clusters of triangulated 3D points with semantic information;
- Points clusters provide constraints to the map.

[1] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," https://github.com/facebookresearch/detectron2, 2019.

Results of S^{3}LAM

1. Introduction
2. Methodology
   2.1. Clusters Creation
   2.2. Map Optimization from Structure Estimation
3. Experiments
   3.1. Implementation Details
   3.2. Impact of the Planes Constraints on Pose Error
   3.3. Qualitative Analysis of S^{3}LAM
4. Expand

- b: map built by ORB-SLAM2
- c: map of clusters where each cluster centroid is represented with a sphere
- d: map of clusters with estimated planes



Outline

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

S^{3}LAM: A Cluster based SLAM

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

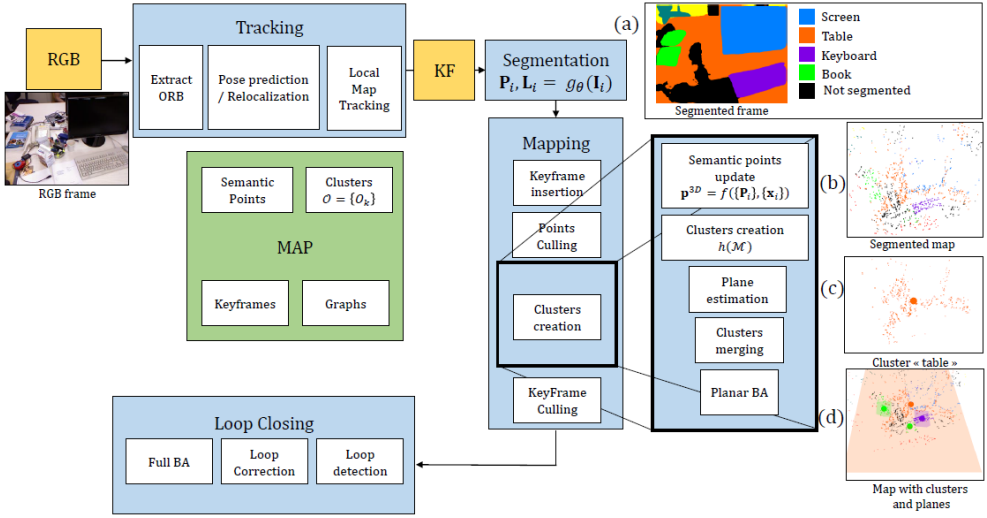
- Map is represented as a set of point clouds, grouped according to the object instance;
- Use the prior knowledge of objects to enrich SLAM;
- Goal: estimate the pose ${}^{c_i}\mathbf{T}_w \in SE(3)$ of a monocular camera between world frame $\mathcal{F}_w$ and camera frame $\mathcal{F}_{c_i}$ at time i.

Pipeline:

1. CNN segments the keyframes;
2. Use the output of CNN to compute the probability distribution of map points;
3. Create clusters of points based on the semantic class and fit planes for planar classes;
4. Apply BA constrained by planes.

Pipeline of S^{3}LAM

1. Introduction
2. Methodology
   2.1. Clusters Creation
   2.2. Map Optimization from Structure Estimation
3. Experiments
   3.1. Implementation Details
   3.2. Impact of the Planes Constraints on Pose Error
   3.3. Qualitative Analysis of S^{3}LAM
4. Expand



Introduction to Panoptic Segmentation

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

Panoptic segmentation:

- Combination of semantic segmentation and instance segmentation:
  - Semantic segmentation: each pixel is classified in a given class;
  - Instance segmentation: multiple objects of the same class are segmented separately.
- Takes longer to obtain, but **allows to know which keypoints belong to detected objects**.
- Panoptic segmentation separates multiple instances of a single class, **allowing to treat each object separately**.
- Panoptic segmentation networks run at 10 to 20 fps.

Problem Faced by Panoptic Segmentation

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

Define function $g_\theta(\mathbf{I}_i) \rightarrow \mathbf{P}_i, \mathbf{L}_i$ as panoptic segmentation:

- $\mathbf{I}_i$ is the input RGB image at time i;
- θ is the network parameter;
- $\mathbf{P}_i \in [0,1]^{W \times H \times C}$ is probability map, W and H represent the width and height of RGB image, C is the number of semantic classes, $\mathbf{P}_i(u, v, c)$ corresponds to the pixel (u, v) belongs to class c;
- $\mathbf{L}_i \in \mathbb{N}^{W \times H}$ is the instance map in wich each object is segmented and given unique ID.

Problem: The instance map is not temporally consistent, meaning that we cannot simply rely on the object IDs to create the map.

Solution to the Problem of Panoptic Segmentation

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

For each detected instance in $\mathcal{L}^i$ we compute the IoU with all instances of the same class in $\mathcal{L}^{i-1}$ and $\mathcal{L}^{i-2}$ and take its maximum to track the ID.

- An instance is well tracked if the IoU is above threshold $\tau_{i-1} = 0.65$ for $\mathcal{L}^{i-1}$ and $\tau_{i-2} = 0.4$ for $\mathcal{L}^{i-2}$.

Problem: when a cluster that has left the camera field of view reappears in the image, the segmentation network creates a new instance, which leads to the creation of a new cluster.

Solution to the Problem of Panoptic Segmentation

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

Problem: when a cluster that has left the camera field of view reappears in the image, the segmentation network creates a new instance, which leads to the creation of a new cluster.

Solution 1: define a merging function to fuse the clusters together when the distance between their centroid is lower than τ_{merge} and more than 80% of clusters points descriptors match.

Solution 2: Reproject cluster points in 2D and assigning them the id of the segment in which they fall.

Clusters Creation

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

Define fusion function $f(\{\mathbf{P}_i\}, \{\mathbf{x}_i\}) \rightarrow \mathbf{p}^{3D}$ of multiple observations:

- $\{\mathbf{P}_i\}$ is a set of probability maps at different times;
- $\{\mathbf{x}_i\}$ is a set of keypoints corresponding to one 3D point ${}^w\mathbf{X}$;
- $\mathbf{p}^{3D} = (p_1, ..., p_c)$ is probability distribution of ${}^w\mathbf{X}$.

The function can be written using Bayes rules:

$$p_c = \mathbb{P}(c|\{\mathbf{P}_i\}) = \frac{1}{Z}\mathbb{P}(c|\{\mathbf{P}_{i-1}\})\mathbf{P}_i(u, v, c) \tag{1}$$

c is the class label of ${}^w\mathbf{X}$, Z is normalization factor.

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

Clusters Creation

$$f(\{\mathbf{P}_i\}, \{\mathbf{x}_i\}) \rightarrow \mathbf{p}^{3D}$$

$$p_c = \mathbb{P}(c|\{\mathbf{P}_i\}) = \frac{1}{Z}\mathbb{P}(c|\{\mathbf{P}_{i-1}\})\mathbf{P}_i(u, v, c)$$

f allows to obtain a semantic map $\mathcal{M} = \{({}^w\mathbf{X}, \mathbf{p}^{3D}, c^*, l)_j\}$:

- each point ${}^w\mathbf{X}$ has a probability distribution $\mathbf{p}^{3D}$, ID l extracted from instance map $\mathbf{L}_i$ and semantic class $c^* = \arg\max_c \mathbf{p}^{3D}$.

Clusters Creation

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

Define clustering function $h(\mathcal{M}) \rightarrow \mathcal{O}$:

- Semantic map $\mathcal{M} = \{({}^w\mathbf{X}, \mathbf{p}^{3D}, c^*, l)_j\}$;
- $\mathcal{O}$ is a partion of $\mathcal{M}$ in K clusters $\mathcal{O} = \{O_k | k \in [1, K]\}$;
- Function groups points according to their semantic class and instance.

Each cluster can be defined as $O_k = \{\{{}^w\mathbf{X}\}, c_k\}$:

- $\{{}^w\mathbf{X}\}$ is the position of a set of points belonging to the cluster;
- c_k is cluster class.

The centroid of each cluster is computed with the mean of 3D points and continuously updated, hence it moves when new parts of the objects are seen.

Structure Estimation

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

Most man-made objects can be approximated with a more or less complex geometrical model, from a simple plane or box to the exact 3D model of the object.

- This work proposes to model some objects using planes. Not only do we model large surfaces such as tables, walls or floor but also model small objects like keyboards and books.

Planes are represented using 4D vector $\pi = (a, b, c, d)^\top$ and $\|\pi\|_2 = 1$, planar points $\mathbf{X}$ in homogeneous coordinates staisfy:

$$\pi^\top \mathbf{X} = 0 \tag{2}$$

Structure Estimation

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

- Contrary to most planar SLAM systems, this work does not need to use multiple specific CNNs or depth to estimate plane parameters.
- Fit a plane using the triangulated 3D points of this cluster in the world coordinates system. This is done using SVD inside a RANSAC loop (PCL library has corresponding functions).
- This work fit planes for small specific objects which is not limited to very large planes.
- A plane is accepted if it is supported by enough inliers, which depends on the cluster class.

Map Optimization

Include constraint as a regularizer:

$$ {}^{c}\hat{\mathbf{T}}_w, {}^{w}\hat{\mathbf{X}} = \arg\min_{{}^{c}\mathbf{T}_w, {}^{w}\mathbf{X}} \sum_{i,j} \rho(\|\mathbf{x}_{i,j} - \text{proj}({}^{c_i}\mathbf{T}_w, {}^{w}\mathbf{X})\|_{\Sigma_{i,j}}) + \sum_{k} \sum_{j \in O_k} \rho(\|\pi_k^\top {}^{w}\mathbf{X}_j\|_\sigma) \quad (3) $$

- $\|\pi_k^\top {}^{w}\mathbf{X}_j\|$ is the 3D distance between 3D point ${}^{w}\mathbf{X}_j$ and the plane π_k which corresponds to the cluster O_k;
- $\sigma(= 100)$ corresponds to its uncertainty; ρ is Huber loss function.

Computing (3) needs its jacobian:

$$ \frac{\partial \pi_k^\top {}^{w}\mathbf{X}_j}{\partial {}^{w}\mathbf{X}_j} = \pi_k^\top \quad (4) $$

Outline

1. Introduction
2. Methodology
   2.1. Clusters Creation
   2.2. Map Optimization from Structure Estimation
3. Experiments
   3.1. Implementation Details
   3.2. Impact of the Planes Constraints on Pose Error
   3.3. Qualitative Analysis of S^{3}LAM
4. Expand

1. Introduction
2. Methodology
   2.1. Clusters Creation
   2.2. Map Optimization from Structure Estimation
3. Experiments
   3.1. Implementation Details
   3.2. Impact of the Planes Constraints on Pose Error
   3.3. Qualitative Analysis of S^{3}LAM
4. Expand

Implementation Details

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

Panoptic Segmentation: Panoptic Feature Pyramid Networks based on ResNet 101

PC: Intel Xeon @3.7GHz, 16GB RAM, Nvidia RTX2070

Datasets: TUM RGB-D, KITTI

Metrics: ATE and RMSE

1. Introduction
2. Methodology
   2.1. Clusters Creation
   2.2. Map Optimization from Structure Estimation
3. Experiments
   3.1. Implementation Details
   3.2. Impact of the Planes Constraints on Pose Error
   3.3. Qualitative Analysis of S^{3}LAM
4. Expand

TABLE I

COMPARISON OF THE ATE (MM) OF OUR OUR APPROACH AGAINST STATE OF THE ART ON THE TUM DATASET.

Sequence	[31] (RGB-D)	ORB-SLAM 2 [1]	[14] (RGB) w. planes	[14] (RGB) w. quadrics	[22]	S^{3}LAM
fr1_xyz	9.6	9.2	10.3	10.0	-	**8.8**
fr1_floor	13.8	18.1	16.9	-	-	**14.7**
fr1_desk	15.3	13.9	12.9	**12.6**	-	13.2
fr2_xyz	3.3	2.4	**2.2**	**2.2**	-	2.4
fr2_desk	12.0	8.0	7.3	**7.1**	-	7.8
fr3_nost_text_near	-	20.3	-	-	-	**15.3**
fr3_nost_text_near (merged books)	-	20.3	-	-	-	**13.5**
fr3_nost_text_near (loop)	10.9	14.5	-	-	11.4	**10.9**
fr3_str_text_near	-	14.0	-	-	**10.6**	11.2
fr3_str_text_far	-	10.6	-	-	**8.8**	9.2
fr1 mean	12.9	13.9	13.4	-	-	**12.2**
fr2 mean	7.7	5.2	4.8	**4.7**	-	5.1
fr3 mean	-	13.0	-	-	**10.3**	10.4

We compare our approach to: the base system of S^{3}LAM, ORB-SLAM2 and both the monocular and RGB-D approaches of Hosseinzadeh et al [31], [14] that use both quadrics and planes.

We also compare against the plane based approach of [22] that report experiments on a few strongly planar sequences from the TUM dataset

Qualitative Analysis of S^{3}LAM

1. Introduction
2. Methodology
   2.1. Clusters Creation
   2.2. Map Optimization from Structure Estimation
3. Experiments
   3.1. Implementation Details
   3.2. Impact of the Planes Constraints on Pose Error
   3.3. Qualitative Analysis of S^{3}LAM
4. Expand

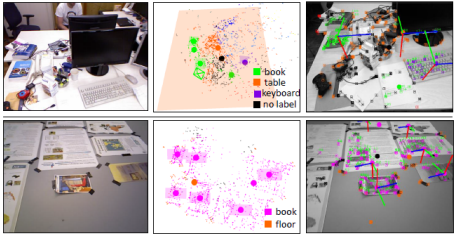


TABLE II

COMPARISON OF THE ATE (CM) OF OUR APPROACH AGAINST ORB-SLAM2 ON THE KITTI RAW DATASET.

sequence	ORB-SLAM 2 [1]	Ours
0926-0011	17.7	**15.5**
0926-0013	18.0	**7.5**
0926-0014	76.2	**64.5**
0926-0056	49.8	**49.3**
mean	40.4	**34.2**

TABLE III

MAXIMUM, MINIMUM AND AVERAGE VALUES OF ANGLES BETWEEN PLANE NORMALS. THE CLOSER TO 0 THE BETTER.

Sequence	max. angle	min. angle	med. angle
fr1_desk	3.6°	2.4°	2.9°
fr3_nost_text_near	1.4°	0.8°	0.8°
fr3_nost_text_near (merged)	0.0°	0.0°	0.0°

Outline

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

Expand

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

1. The usage of panoptic segmentation, which gives the instance ID while instance segmentation and object detection do not.
2. I can try introducing object constraints in my graduation project:
   - Add object constraints to traditional odometry and keypoints observation constraints.
   - How to represent object constraints: plane feature constraint or IoU or others?
   - How to match objects with IDs or other feature information?
   - Maybe I can add object semantic detection box to static map created by SLAM system.

1. Introduction
2. Methodology
2.1. Clusters Creation
2.2. Map Optimization from Structure Estimation
3. Experiments
3.1. Implementation Details
3.2. Impact of the Planes Constraints on Pose Error
3.3. Qualitative Analysis of S^{3}LAM
4. Expand

Thank you!