Deep Learning based Visual SLAM System for Dynamic Environment
(BTech Project)

Shubham Shankar , 190101107
Multimedia Lab (Under Dr. Arijit Sur)

SLAM : Simultaneous Localization and Mapping

- SLAM (Simultaneous Localization And Mapping) is the process of simultaneously constructing the map of an unknown environment and tracking the position of the sensor with respect to the map.
- Based on the type of the sensor used SLAM systems can be classified into multiple categories :
  - LIDAR (Light Imaging Detection and Ranging) SLAM
  - RADAR (Radio Detection and Ranging) SLAM
  - **Visual SLAM**

SLAM is used in UAVs, robots and drones

Some of the applications of SLAM include :

- Autonomous vehicles.
- Search and rescue in high-risk or difficult-to-navigate environments.
- Unmanned Aerial Vehicles (UAV).

SLAM algorithms can be classified into feature based and direct algorithms

Depending on the method used for representing an image SLAM algorithms can be classified into two categories :-

- **Feature-based methods** :- Extract certain key points from the given image frame. These keypoints are used for localization and mapping.
- **Direct methods** :- Image intensities are directly used to estimate the location and construct the map. These methods do not perform any abstraction and thus tend to be computationally expensive.

Literature Survey

- ORB-SLAM2 is an open source feature based visual SLAM algorithm which extracts ORB features from input images that are later used for tracking and mapping.
- DS-SLAM is based ORB-SLAM2. It uses semantic segmentation along with geometry based methods to remove features associated with moving objects.
- Geometry based method are used to determine the initial set of dynamic feature points.
- Semantic segmentation is used to determine the contours associated with moving objects.
- If the contour of an object contains a certain number of the dynamic feature points then the objects is treated as a moving object and all feature points associated with it are removed.

Limitations of previous work

- Traditional methods like ORB-SLAM2 **does not filter dynamic key points** out, which leads to poor performance issues in tracking and mapping phases in dynamic environment.
- DS-SLAM uses **semantic segmentation** to identify dynamic objects and eliminate dynamic key points. But it frequently fails to identify dynamic key points located on the **object's contour**.
- **Semantic Segmentation** based approach would also fail in cases where the motion of a **static object** is caused because of the motion of a **dynamic object**.

Problem Statement

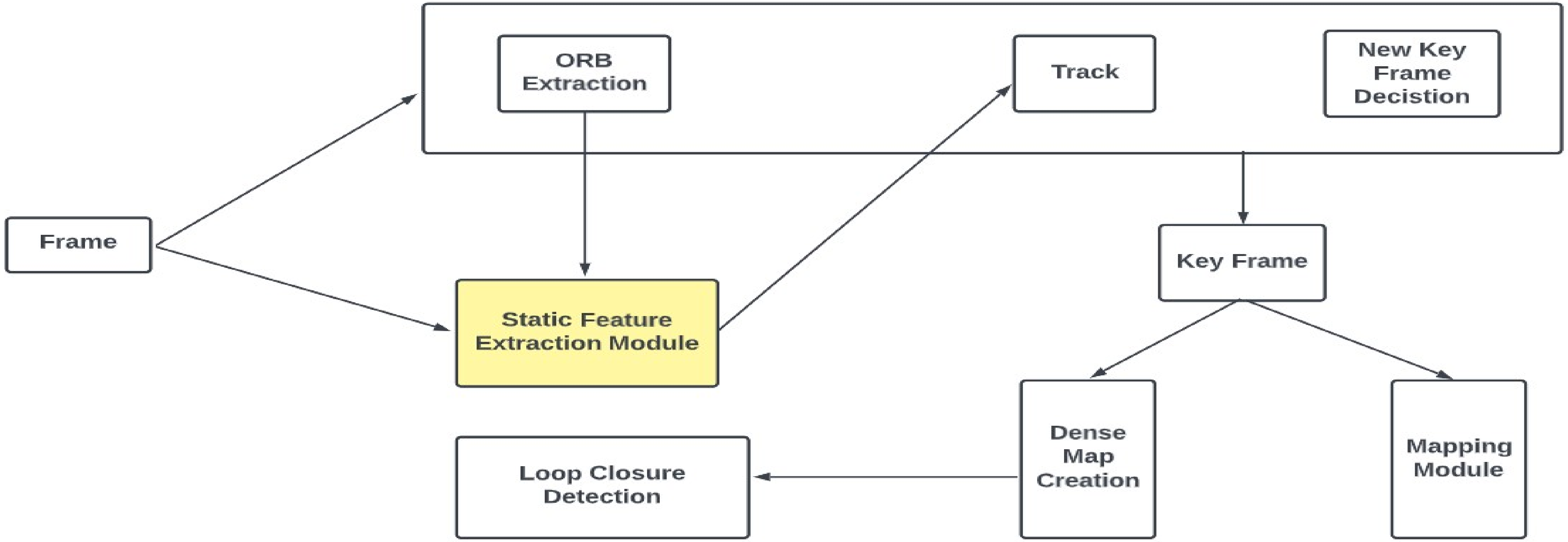
Given a sequence of image frames develop an algorithm which can accurately identify static feature points in the current image frame.

Contributions

A robust **static feature extraction** module on top of the ORB-SLAM2 algorithm which is capable of identifying dynamic key points under **different conditions** based on the following two **novel approaches** is proposed

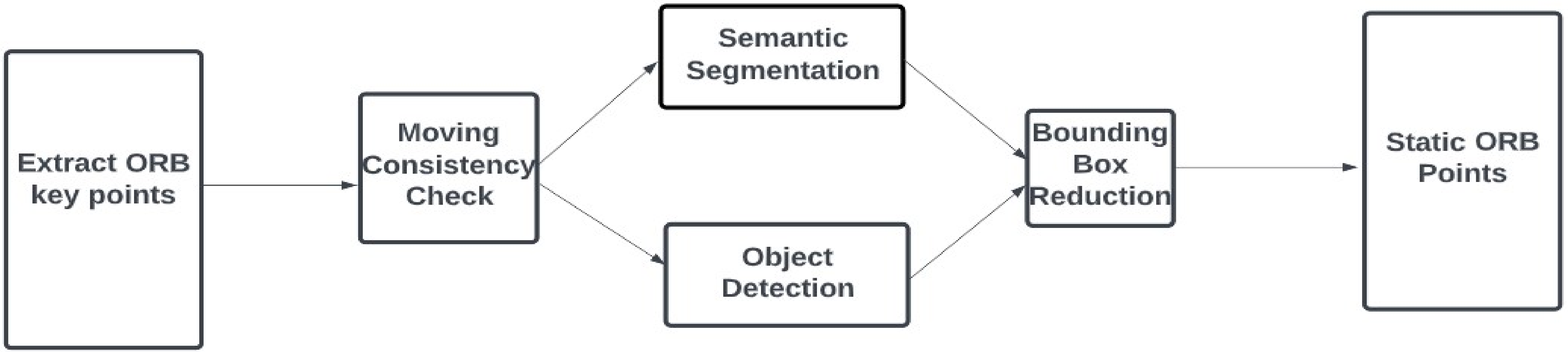
1. **Bounding Box Reduction** technique to accurately identify contours using **pre-trained semantic segmentation** and **object detection** models of moving objects and remove dynamic key points within the contours
2. **Optical Flow Estimation** technique that utilizes **pre-trained optical flow generation** models to identify dynamic keypoints using the displacements of pixels.

Architecture



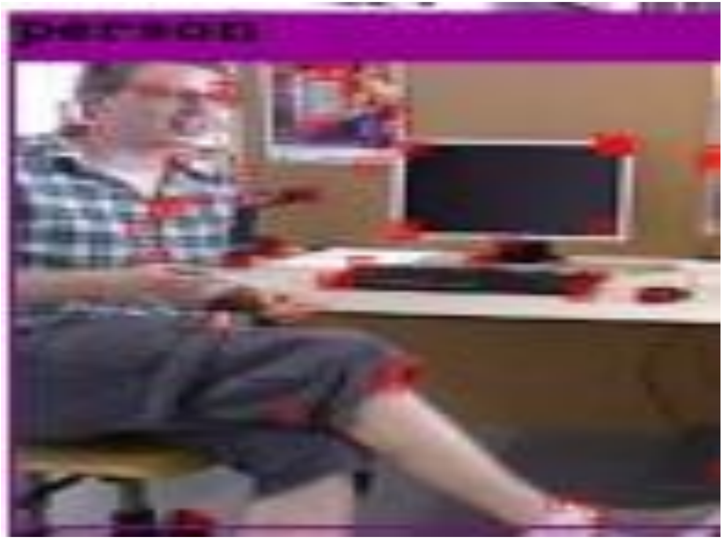
Bounding Box Reduction

Bounding Box Reduction Overview



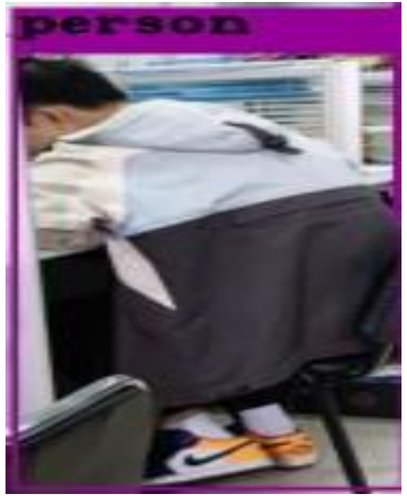
Discarding all key points in the object's contour is not optimal

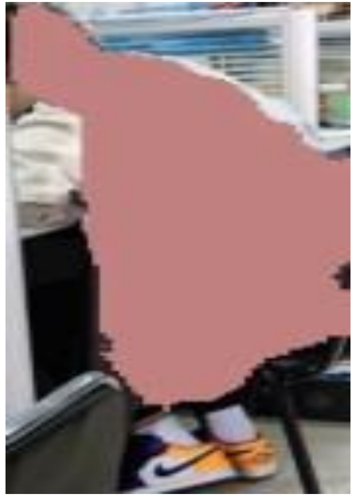
- Simply discarding all feature points in an objects **rectangular contour will not be optimal** because this contour may also contain static feature points.
- If the number of **feature points is too low** the tracking phase will fail.
- YOLOv3 has been used for object detection.



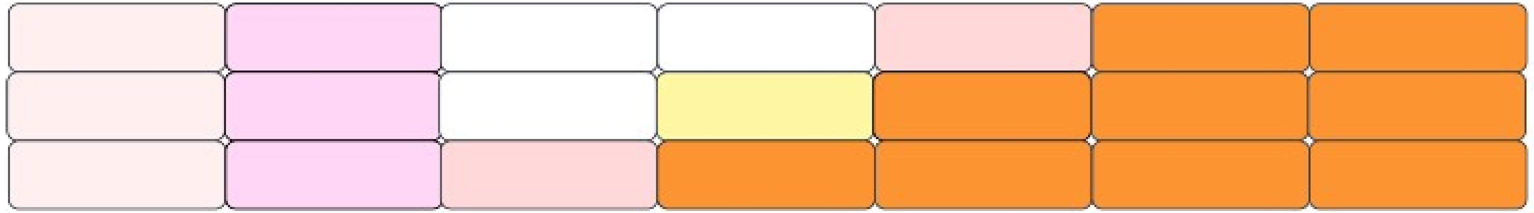
Use bounding box reduction to obtain a more accurate contour.

- Any pixel whose **manhattan distance** to the closest pixel which belongs to the same class as the original object is less than a particular threshold (*thr)* is treated as a dynamic pixel.
- PSPNet is used for generating the semantic segmentation results.

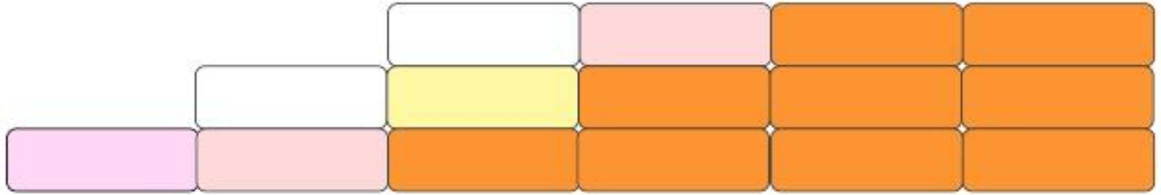




Bounding Box Reduction Working



Object Detection Contour

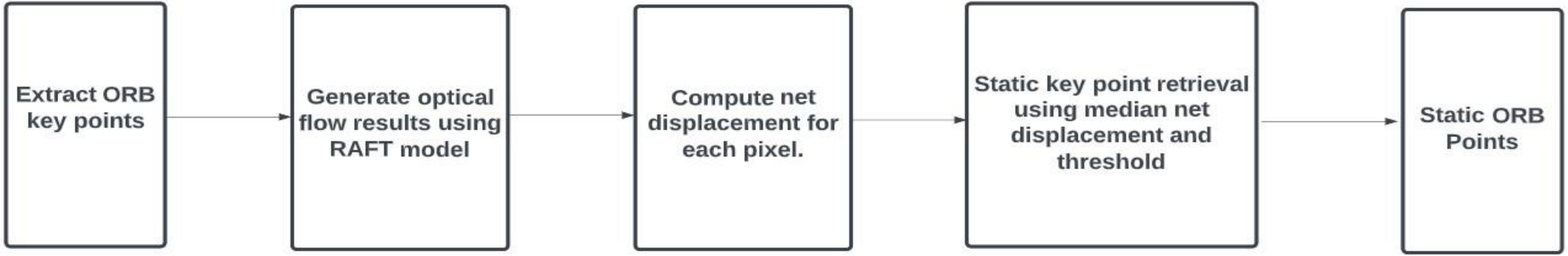


Reduced Bounding Box

Bounding Box Reduction with threshold set to 2 pixels

Optical Flow Estimation

Optical Flow Estimation Overview


Extract ORB key points
Generate optical flow results using RAFT model
Compute net displacement for each pixel.
Static key point retrieval using median net displacement and threshold
Static ORB Points

Optical Flow

- RAFT model uses the current frame and the previous image frame to determine the displacement of each pixel in the image frame.
- Compute the median displacement of all the pixels.
- Every keypoint in a potentially dynamic object detected using object detection, whose displacement does not lie in the range [median - delta, median + delta] is treated as a dynamic keypoint and discarded.
- Since the camera used to capture the image frames is moving the static keypoints will have some displacement which will be nearly equal to the camera's displacement but the dynamic keypoints corresponding to moving objects will have a significantly different displacement and hence will be outliers.

Results

Evaluation Metrics

The following metrics are used for comparing results

1. **Absolute Trajectory Error** : Absolute Trajectory Error represents the global consistency of the predicted trajectory.
2. **Relative Pose Error** : Relative Pose Error is an indicator of the local accuracy of the predicted trajectory over a fixed time interval. It comprises of the translational and rotational components.

Comparison

- Our models result is compared with ORB-SLAM2 and DS-SLAM on the TUM RGB-D dataset.
- Our proposed approaches and DS-SLAM have been implemented on top of the ORB-SLAM2 algorithm which is non deterministic. As a result the proposed approaches and DS-SLAM are non deterministic.
- Issue of non determinism has not been addressed by DS-SLAM. Results on some runs are significantly worse than the results claimed in the paper.
- I have addressed this issue by including the average and the best results of the proposed approaches over 100 runs.

ATE (metric absolute trajectory error)

Dataset	ORB-SLAM2	DS-SLAM	Our Model (Bounding Box Reduction)		Our Model (Optical Flow Estimation)	
	best rmse	best rmse	best rmse	avg rmse	best rmse	avg rmse
walking_xyz	0.752	0.024	0.014	0.015	**0.013**	0.014
walking_rpy	0.870	0.442	0.045	0.109	**0.037**	0.125
walking_static	0.390	0.008	**0.006**	0.007	0.008	0.012

RPE (metric transational drift)

Dataset	ORB-SLAM2	DS-SLAM	Our Model (Bounding Box Reduction)		Our Model (Optical Flow Estimation)	
	best rmse	best rmse	best rmse	avg rmse	best rmse	avg rmse
walking_xyz	0.412	0.033	**0.020**	0.022	0.020	0.021
walking_rpy	0.424	0.150	**0.052**	0.185	0.064	0.223
walking_static	0.216	0.010	**0.009**	0.011	0.012	0.021

RPE (metric rotational drift)

Dataset	ORB-SLAM2	DS-SLAM	Our Model (Bounding Box Reduction)		Our Model (Optical Flow Estimation)	
	best rmse	best rmse	best rmse	avg rmse	best rmse	avg rmse
walking_xyz	7.740	0.820	**0.606**	0.631	0.608	0.620
walking_rpy	8.080	3.000	**1.109**	3.690	2.749	4.379
walking_static	3.890	0.260	0.279	0.317	0.320	0.460

Ablation

ATE (metric absolute trajectory error)

Dataset	Semantic Segmentation		Our Model (Bounding Box Reduction)		Our Model (Optical Flow Estimation)	
	best rmse	avg rmse	best rmse	avg rmse	best rmse	avg rmse
walking_xyz	0.013	0.015	0.014	0.015	**0.013**	0.014
walking_rpy	0.201	0.492	0.045	0.109	**0.037**	0.125
walking_static	0.007	0.013	**0.006**	0.007	0.008	0.012

RPE (metric transational drift)

Dataset	Semantic Segmentation		Our Model (Bounding Box Reduction)		Our Model (Optical Flow Estimation)	
	best rmse	avg rmse	best rmse	avg rmse	best rmse	avg rmse
walking_xyz	0.019	0.022	0.020	0.022	0.020	0.021
walking_rpy	0.292	0.743	**0.052**	0.185	0.064	0.223
walking_static	0.011	0.023	**0.009**	0.011	0.012	0.021

RPE (metric rotational drift)

Dataset	Semantic Segmentation		Our Model (Bounding Box Reduction)		Our Model (Optical Flow Estimation)	
	best rmse	avg rmse	best rmse	avg rmse	best rmse	avg rmse
walking_xyz	0.607	0.634	**0.606**	0.631	0.608	0.620
walking_rpy	5.764	14.934	**1.109**	3.690	2.749	4.379
walking_static	0.303	0.533	**0.279**	0.317	0.320	0.460

Future Work

- Develop an algorithm using a combination of the proposed schemes.
- Relax the assumption of uniform displacement in optical flow estimation.
- Pre training models on synthetically prepared datasets.