Moving object removal for robust visual SLAM in dynamic environment

JIT CHATTERJEE

MOSIG(M1)

<u>*Supervised by*</u>:

Michele Rombaut and

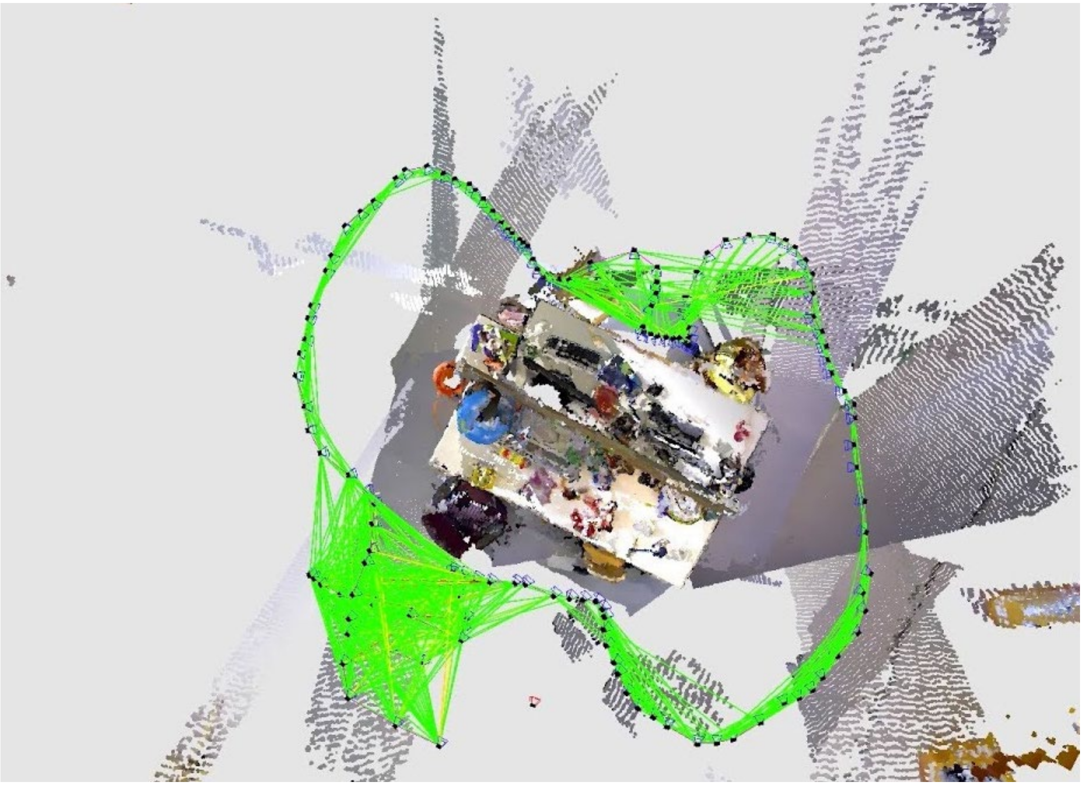
Bruce Canovas

Link: https://drive.google.com/drive/folders/1cpKk4eTDczsLgFjgerGsxdp2bgzIuF3F?usp=sharing

Problem Statement

Removing moving objects for robust visual SLAM in dynamic environment

RGB-D Simultaneous localization and mapping (SLAM)



[1]

Related Works

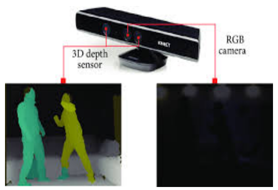
Classical approach

- Optical Flow [2]
- RANSAC [3]

Deep Learning based approach

- DynaSLAM [4]
- Semantic segmentation + Convolutional Neural Networks (CNN) [5]

Proposed Pipeline



RGB-D Frames

Module 1:

Potential Dynamic Object Detection

+

Background-Foreground segmentation

Module 2:

Initial Camera-trajectory Estimation

Module 3:

Dynamic Object Detection

Module 4:

Final Camera-trajectory estimation

+

Sparse 3D reconstruction

Sparse 3D Map

Global Architecture

Input to the System:
RGB-D Frames

Potential dynamic object detection on the RGB frames using YOLOv3-tiny

Background-foreground segmentation of the bounding boxes on the depth frames using adaptive Gaussian Mixture Model Clustering

Removal of feature points from the masks of potential dynamic objects

Initial Camera-trajectory Estimation using RGBD-PTAM

Dynamic object detection using Dense Optical Flow and initial Camera-trajectory estimation

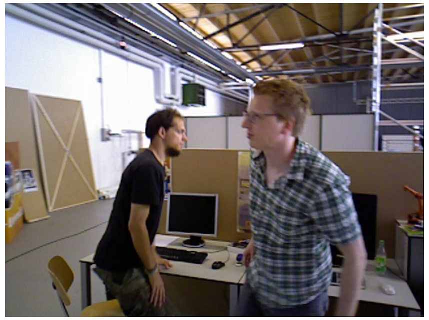
Final Camera-trajectory estimation
+
Sparse 3D map reconstruction using RGBD-PTAM

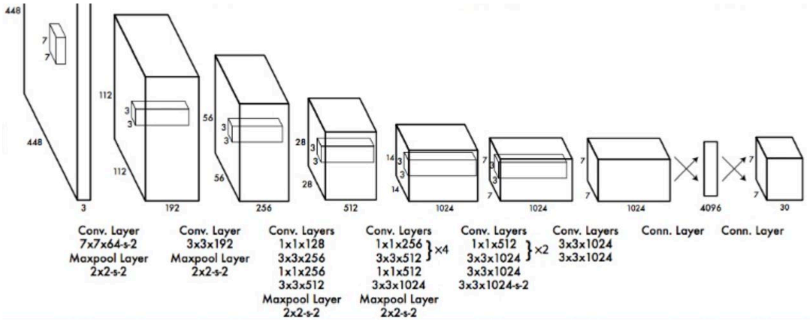
Module 1: #Potential Dynamic Object Detection

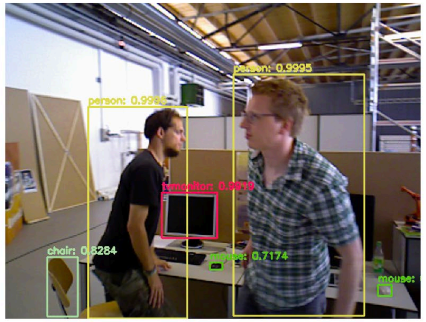
Input RGB Frames

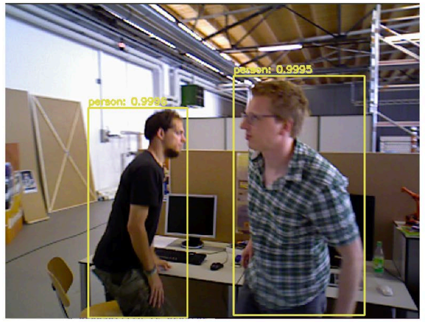
YOLOv3-tiny [6]

Bounding boxes of the potential dynamic object





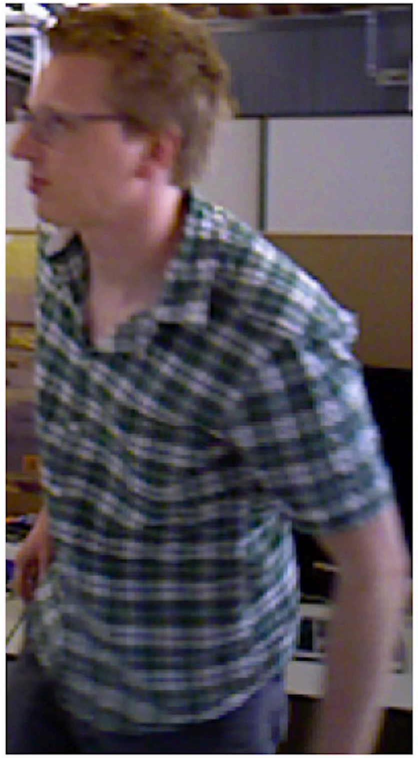




Module 1: #Background-Foreground Segmentation

Bounding boxes of potential dynamic objects → Adaptive Gaussian Mixture Model (GMM) clustering [7] → Masks of potential dynamic objects

Module 1:
#Background-Foreground Segmentation



RGB Image

Segmented image

Depth Image

Segmented image

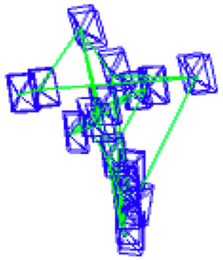
Module 2:
#Initial Camera-trajectory Estimation

Removal of feature points from the masks of potential dynamic objects

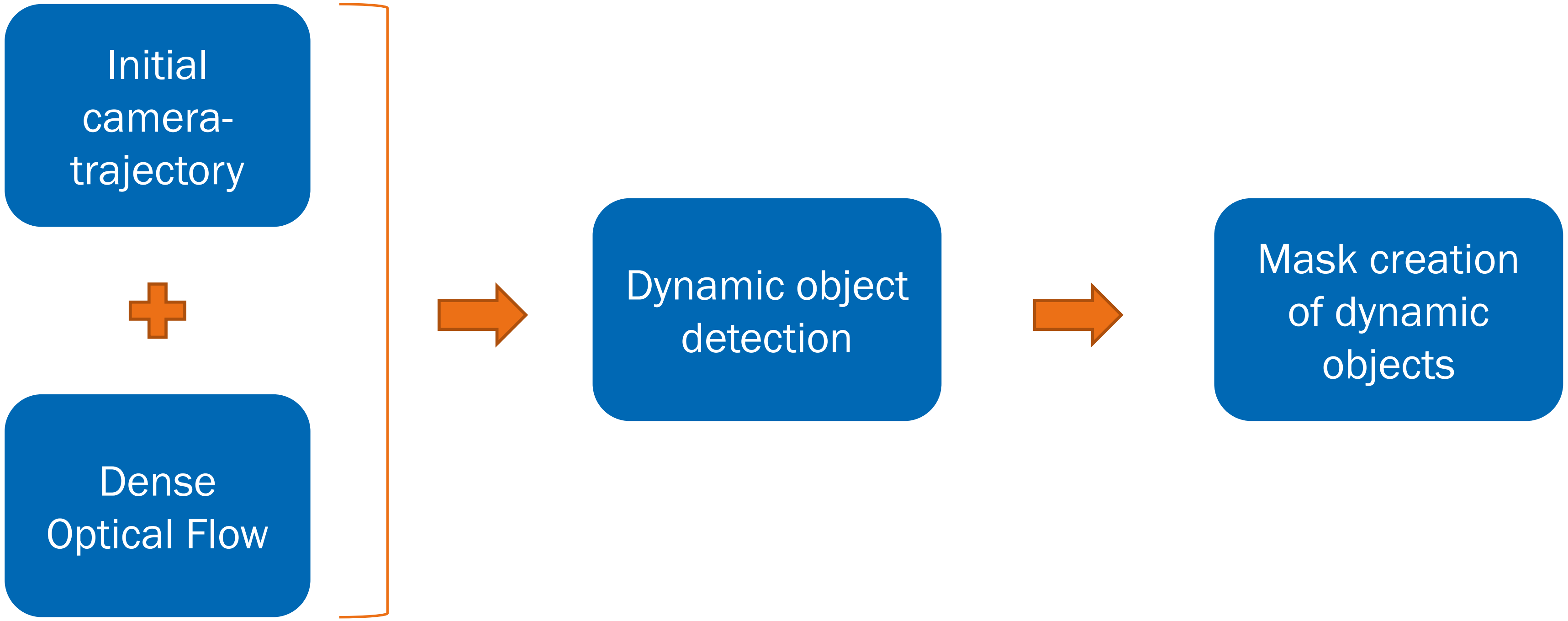
→

Initial camera motion estimation using RGB-D PTAM[8] which uses Good Feature To Track (GFTT) [9]

→



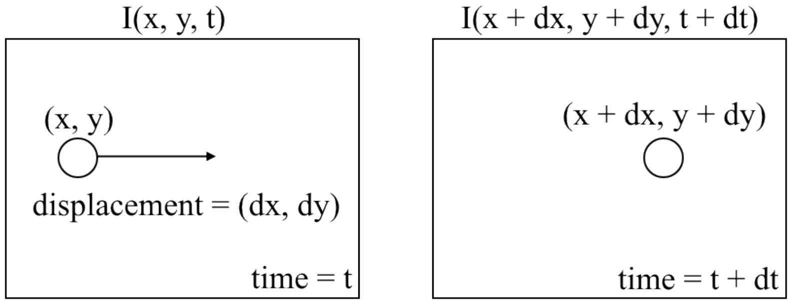
Module 3:
#Dynamic Object Detection using Optical Flow



Module 3: #Optical Flow

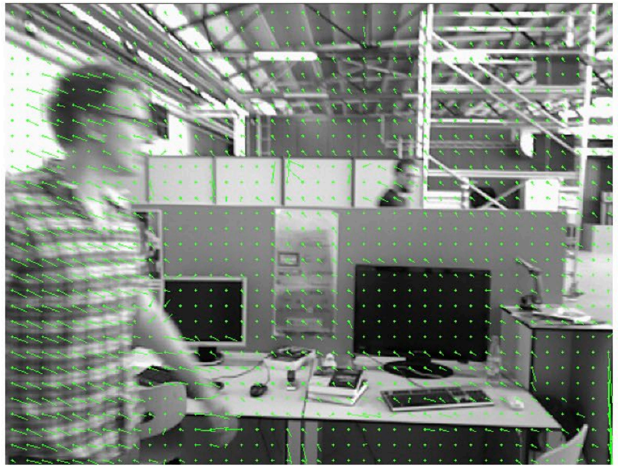
Optical Flow–

Optical flow is the motion of objects between consecutive frames of sequence, caused by the relative movement between the object and camera.



Different kinds of Optical Flow–

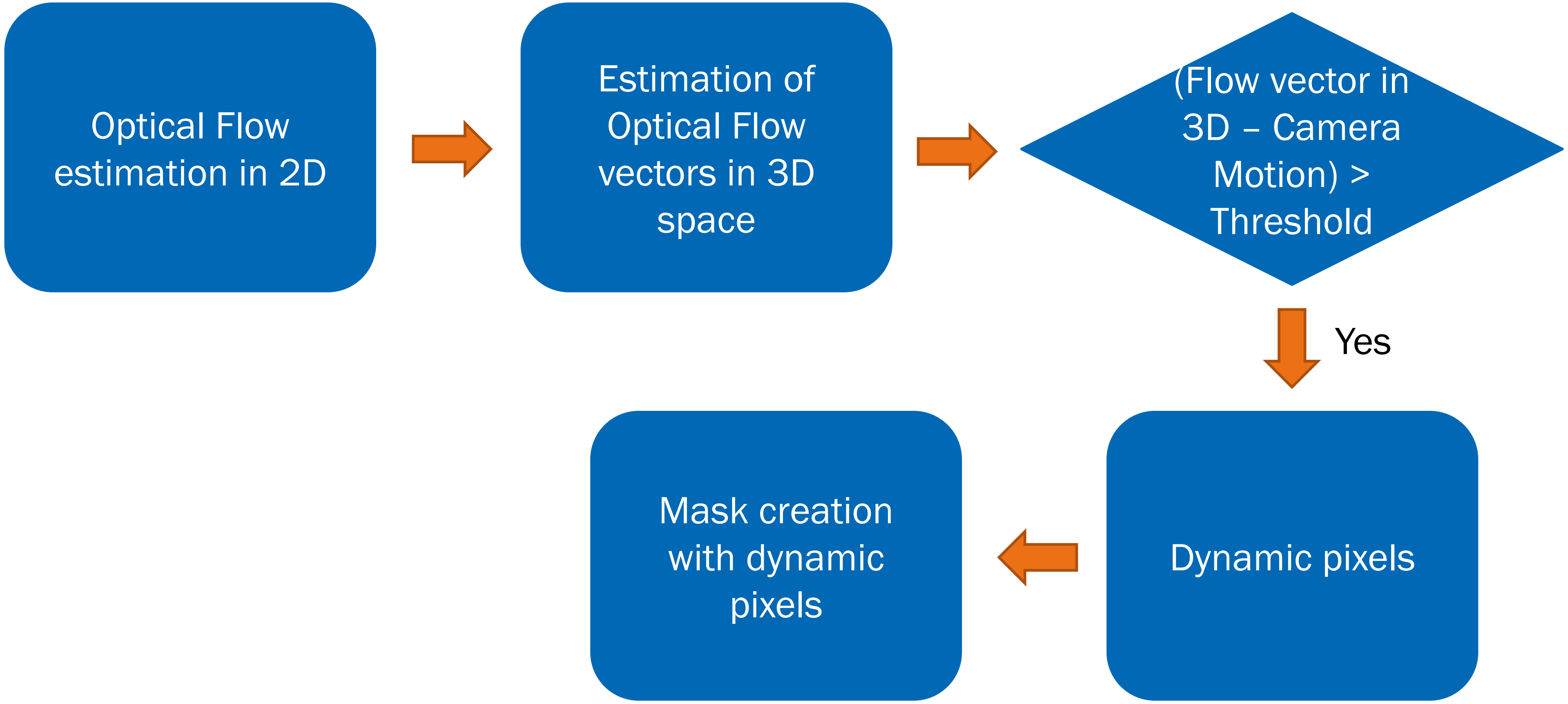
Sparse Optical Flow (Lucas Kande) [10]
Dense Optical Flow (Gunnar-Farneback) [11]



Optical Flow

Dense Optical Flow

Module 3:
#Dynamic pixels detection and mask creation

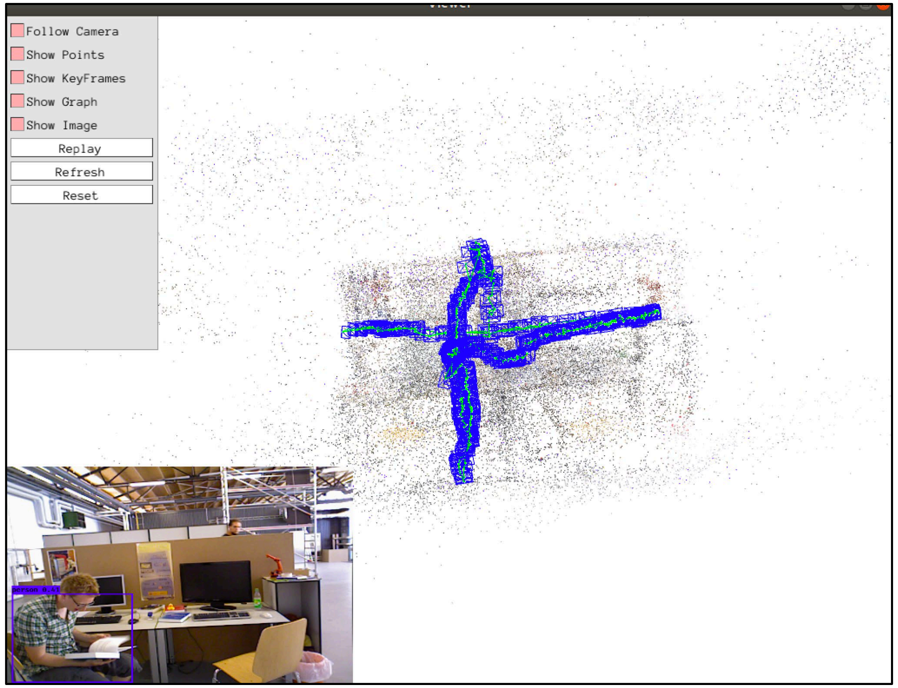


Module 3:
#Dynamic pixels detection and mask creation

Feature points only on the static part

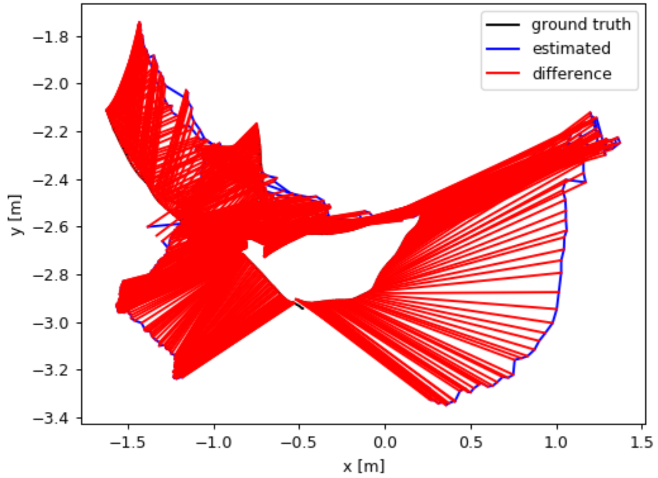
Dynamic object (Black Mask)

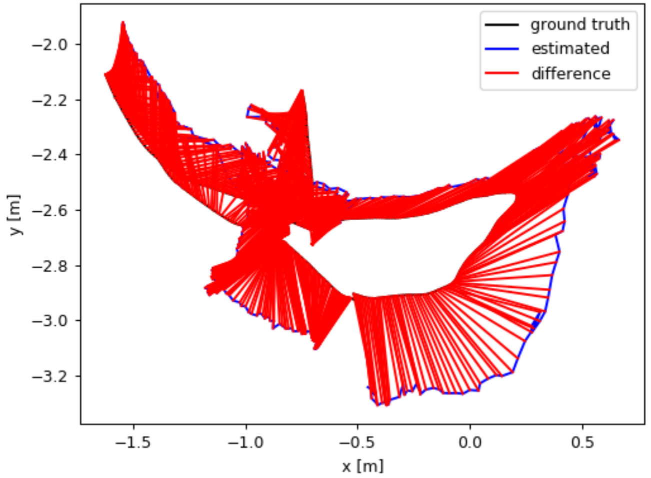
Module 4:
#Final Camera-trajectory estimation and Sparse 3D Map reconstruction

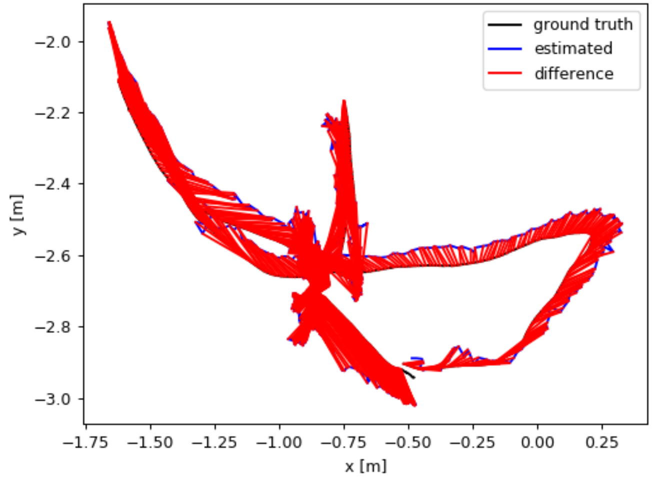


Sparse 3D map generation after eliminating dynamic feature points

Results:
Absolute Trajectory Error (RMSE)

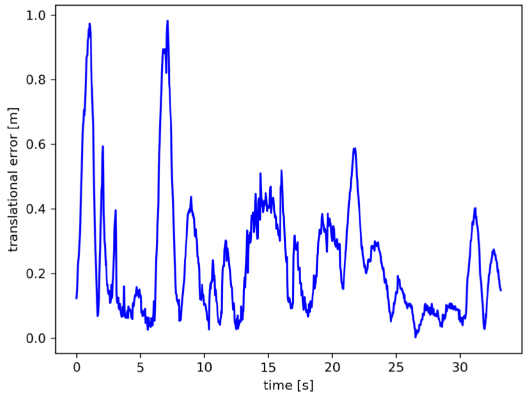


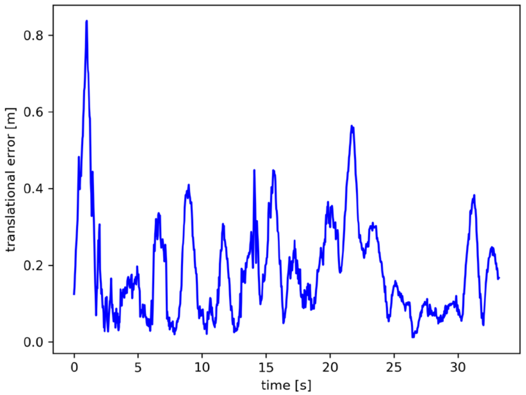


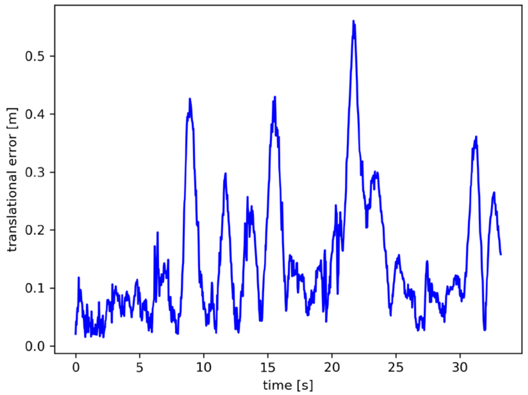


TUM Dataset RGB-D sequences	Normal SLAM	With YOLO	With our method
fr3 walking halfsphere	0.76m	0.33m	0.17m

Results:
Relative Pose Error (RMSE)







TUM Dataset RGB-D sequences	Normal SLAM	With YOLO	With our method
fr3 walking halfsphere	25.85 deg	25.33 deg	25.21 deg

Used Tools

– Tools and Algorithms used

- Python 3
- OpenCV
- YOLOv3-tiny [6]
- Gaussian Mixture Model (GMM) [7]
- Good Features To Track (GFTT) [9]
- Gunnar Farneback optical flow [11]
- RGBD-PTAM SLAM system [8]

Datasets used–

- TUM dataset [12]
- COCO dataset for training YOLOv3-tiny [13]

Conclusions

- Our SLAM system is a combination of conventional and deep learning approach
- Our proposed SLAM system is robust and also light-weight
- Got nice results of TUM dataset

Future Works

- Adaptive threshold in optical flow
- Testing on the robot
- Sparse optical flow approach can speed up the system

Acknowledgments

I really want to thank my supervisors Michele Rombaut and Bruce Canovas, and also want to thank the entire Team COPERNIC, GIPSA Lab for their immense support I received during my internship tenure.

Referrences

[1] Dense Visual SLAM for RGB-D Cameras (C. Kerl, J. Sturm, D. Cremers), In Proc. of the Int. Conf. on Intelligent Robot Systems (IROS), 2013.

[2] Cho J, Jung Y, Kim DS, Lee S, and Jung Y. Moving object detection based on optical flow estimation and a gaussian mixture model for advanced driver assistance systems. Sensors (Basel), 2019.

[3] Shao-Wen Yang and C. Wang. Multiple-model ransac for ego-motion estimation in highly dynamic environments.

[4] F´acil JM. Civera Javier Bescos, Berta and Jos´e Neira. DynaSLAM: Tracking, mapping and inpainting in dynamic environments. IEEE RA-L, 2018.

[5] L. Zhang, L. Wei, P. Shen, W. Wei, G. Zhu, and J. Song. Semantic slam based on object detection and improved octomap. IEEE Access, 6:75545–75559, 2018.

[6] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. https://arxiv.org/abs/1804.02767, 2018.

[7] Thierry Bouwmans, Fida El Baf, and Bertrand Vachon. Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey. Recent Patents on Computer Science, 1(3):219–237, November 2008.

Referrences

[8] Taihu Pire, Thomas Fischer, Gast´on Castro, Pablo De Crist´oforis, Javier Civera, and Julio Jacobo Berlles. S-ptam: Stereo parallel tracking and mapping. Robotics and Autonomous Systems, 93:27–42, 2017.

[9] Jianbo Shi and Tomasi. Good features to track. In 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pages 593–600, 1994.

[10] Bruce Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision (ijcai). volume 81, 04 1981.

[11] Gunnar Farneback. Two-frame motion estimation based on polynomial expansion. In Josef Bigun and Tomas Gustavsson, editors, Image Analysis, pages 363–370, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[12] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of rgb-d slam systems. In 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 573–580, 2012.

[13] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Doll´ar, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. CoRR, abs/1405.0312, 2014.