ERCIM "ALAIN BENSOUSSAN" FELLOWSHIP PROGRAMME

Scientific Report

First name / Family name	Durga Prasad Bavirisetti
Nationality	Indian
Name of the *Host Organisation*	Norwegian University of Science and Technology (NTNU)
First Name / family name of the *Scientific Coordinator*	Frank Lindseth
Period of the fellowship	01/02/2022 to 31/01/2023

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

As a postdoctoral researcher, my primary objective is to investigate vision-based methods for perceiving the environment around autonomous vehicles. Throughout my tenure, I have examined both 2D and 3D vision problems, including object detection and segmentation tasks in image and LiDAR point cloud domains. My main activities during this postdoc are as follows:

1. **Data acquisition:**

The acquisition of images and point clouds for autonomous vehicles is a crucial step in ensuring their safe and efficient operation. One method for acquiring this data is by using a monocular camera and LiDAR sensor attached to a vehicle.

The NTNU Autonomous Perception Laboratory's main vehicle [1], as shown in Fig.1 is used for data acquisition. It is a Kia e-Niro that is equipped with a variety of sensors including eight cameras, three LiDARs, two radars, and one GNSS system. The vehicle is also loaded with NVIDIA DRIVE software, which allows for the capture of high-resolution RGB images using a 60-degree front camera and point clouds using a 360-degree 120-Channel Ouster LiDAR sensor.

To process the LiDAR point cloud data, the Ouster python software development toolkit (SDK) [2] is used. This toolkit allows the data to be saved in a proper format for further processing, ensuring that it is ready for use in the autonomous vehicle's navigation and decision-making systems.

Fig. 1: Test vehicle of NTNU's Autonomous Perception (NAP) Lab.

2. **Data preprocessing:**

To enhance image recognition capabilities, LiDAR images were created from LiDAR point clouds using the Ouster Python SDK. This process generates images [3] in four layers: range, signal, Near-IR, and reflectivity. To improve the quality of these images and eliminate noise, various preprocessing methods were applied to both LiDAR and camera images, such as auto exposure and beam uniform corrections. This helped in ensuring that the images were of the highest quality and free of any interference that could negatively impact the image recognition tasks.

3. **Data labeling:**

Labeling images is an essential step in the object detection and segmentation tasks for autonomous vehicles. The preprocessed data, including both LiDAR and camera images, needs to be labeled according to the required tasks, such as the detection of static and dynamic objects. In order to accomplish this, we have used computer vision labeling tools such as CVAT [4] and V7 [5]. These tools are considered state-of-the-art in the field and provide a wide range of features that help to speed up the labeling process.

One of the key features of these tools is AI-assisted labeling. This allows the tool to automatically suggest labels based on the image, which greatly speeds up the process. By using these tools, we were able to quickly and accurately label the images for object detection and segmentation tasks.

4. **2D and 3D perception:** *Testing and developing vision algorithms to detect static and dynamic objects.*
Once the data is labeled according to the project's needs, the next step in the vision pipeline for autonomous vehicles is to test or develop algorithms for detecting both static objects [6] such as roads, traffic signs, and poles, and dynamic objects [7] like pedestrians, cyclists, cars, and trucks. During my tenure, we evaluated different vision-based object detection and segmentation algorithms for recognizing objects in both images and LiDAR point clouds. One of the major developments was QT-UNet, a new Vision Transformer [8] UNet-based architecture for semantic segmentation of road objects on monocular camera images, which also demonstrated effective performance in the field of medical image computing. Furthermore, we have developed a multi-task vision transformer technique to perform both segmentation and depth estimation using a single image for autonomous driving purposes.

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

The articles under preparation or review:

1. H˚aversen A.H, **Bavirisetti D.P***, Kiss G., and Lindseth F. “QT-UNet: A Self-supervised self-querying all-Transformer U-Net for 2D and 3D segmentation”, PLOS ONE. (Submitted)
2. **Bavirisetti D.P**, Martinsen, H.R, Kiss G., and Lindseth F. “A Multi-task Vision Transformer for Segmentation and Monocular Depth Estimation for Autonomous Driving”. (Under preparation)

Other publications during this period:

1. Li, Y., Liu, G., **Bavirisetti, D.P**., Gu, X. and Zhou, X., 2023. Infrared-Visible Image Fusion Method Based on Sparse and Prior Joint Saliency Detection and LatLRR-FPDE. *Digital Signal Processing*, p.103910.
2. Xu, R., Liu, G., Xie, Y., **Bavirisetti, D.P.** Qian, Y. and Xing, M., 2022. Multiscale feature pyramid network based on activity level weight selection for infrared and visible image fusion. JOSA A, 39(12), pp.2193-2204.
3. Tang, H., Liu, G., Tang, L., **Bavirisetti, D.P**. and Wang, J., 2022. MdedFusion: A multi-level detail enhancement decomposition method for infrared and visible image fusion. Infrared Physics & Technology, 127, p.104435.
4. Peng, Y., Liu, G., Xu, X., **Bavirisetti, D.P.**, Gu, X. and Zhang, X., 2022. MFDetection: A highly generalized object detection network unified with multilevel heterogeneous image fusion. Optik, 266, p.169599.

III – ATTENDED SEMINARS, WORKSHOPS, AND CONFERENCES

1. Seminar series by Prof. Wolfgang Förstner on photogrammetric computer vision and semantic perception from 31.08.2022 to 09.09.2022 at NTNU, Gloshaugen.
2. Weekly seminar series by NorwayAI at NTNU.
3. Biweekly seminars by computing (COMP) group in the department of computer science at NTNU.

IV – RESEARCH EXCHANGE PROGRAMME

During my research exchange program (REP), I collaborated with Professor Hung Son Nguyen at the University of Warsaw, Poland. I presented a research talk [9] titled "2D and 3D Vision for Autonomous driving" in the Department of Mathematics at the University of Warsaw and explored the potential for working together on various optimization techniques and neural networks in the field of autonomous driving.

References:

[1] https://www.ntnu.edu/idi/naplab

[2] https://ouster.com/developers/ouster-sdk/

[3] https://ouster.com/blog/object-detection-and-tracking-using-deep-learning-and-ouster-python-sdk/

[4] https://www.cvat.ai/

[5] https://www.v7labs.com/

[6] Li, Yanfen, et al. "A deep learning-based hybrid framework for object detection and recognition in autonomous driving." IEEE Access 8 (2020): 194228-194239.

[7] Lee, Hojoon, et al. "Moving object detection and tracking based on the interaction of static obstacle map and geometric model-free approach for urban autonomous driving." IEEE Transactions on Intelligent Transportation Systems 22.6 (2020): 3275-3284.

[8] Khan, Salman, et al. "Transformers in vision: A survey." ACM computing surveys (CSUR) 54.10s (2022): 1-41.

[9] https://mimuw.edu.pl/aktualnosci/seminaria/2d-and-3d-vision-autonomous-driving-ad