Image & 4D radar fusion for autonomous driving

Background:

Robust detection, localization and tracking of objects is essential for autonomous driving. Computer vision has largely driven development in recent years based on camera sensors, but 3D localization from images is still challenging [1]. Lidar point clouds provide accurate localization in 3D [2] by measuring distance and even 3D object's shapes, but Lidar data is less semantic, rather sparse and its range is typically limited to 150m. Radars achieve larger ranges up to 300m adding object velocity through Doppler Effect. However, the returns are even sparser than Lidar, and less precise in terms of localization [3]. Cost and limited resolution of range sensors still keep them as promising complementary devices to video processing, evolving forcibly towards fusion strategies [4][5] that may consider both the 3D localization capabilities of range sensors and the higher spatial resolution of image data.

[1] X. Chen, K. Kundu, Z. Zhang, H. Ma, S. Fidler, and R. Urtasun, "Monocular 3D Object Detection for Autonomous Driving," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 2147–2156. doi: 10.1109/CVPR.2016.236

[2] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast Encoders for Object Detection From Point Clouds," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 12689–12697. doi: 10.1109/CVPR.2019.01298

[3] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, "Three Decades of Driver Assistance Systems: Review and Future Perspectives," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 4, pp. 6–22, 2014, doi: 10.1109/MITS.2014.2336271

[4] V. John and S. Mita, "RVNet: Deep Sensor Fusion of Monocular Camera and Radar for Image-Based Obstacle Detection in Challenging Environments," in *Image and Video Technology*, Cham, 2019, pp. 351–364. doi: 10.1007/978-3-030-34879-3_27

[5] S. Vora, A. H. Lang, B. Helou, and O. Beijbom, "PointPainting: Sequential Fusion for 3D Object Detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 4603–4611. doi: 10.1109/CVPR42600.2020.00466

Research question:

How can fusion of range and image data can better help the robust detection, localization and tracking of objects for autonomous driving?

Hypothesis:

Late fusion (merging after decisions) may swiftly help as complement to camera sensors (as when getting range for object instances detected/segmented/classified in images), whereas more thorough contributions can be achieved from early fusion when neural architectures jointly exploit merged sensor data (image and range).

Research method:

Both fusion strategies require registered and synchronized data from range and image sensors. A particular extreme case is the fusion data produced by cameras and low resolution imaging radars [4][6]. The challenge is to design an interpolation or depth image generation method able to generate a high resolution depth image from the radar data *considering* the higher resolution of the registered image data, in order to assign correct range to objects detected in the images, or to perform an easier and most robust detection of such objects.

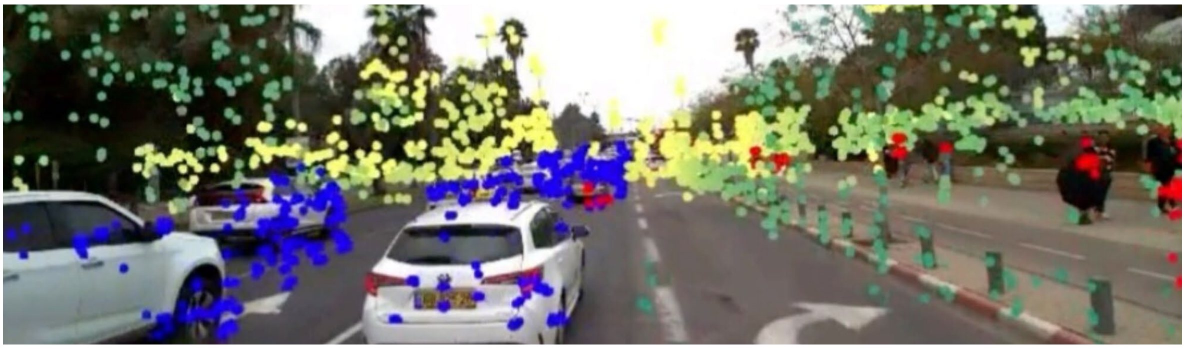

Figure 1. 4D Imaging Radar from Mobileye (from [6])

The aim of the proposed project would be to start exploring how to obtain such high resolution depth data interpolation considering registered image data, by preserving the contours of objects found in the image while assigning correct range to the detected objects.

[6] "Imaging radar gets a second look for AVs," *i-Micronews*, Jan. 20, 2022. https://www.i-micronews.com/imaging-radar-gets-a-second-look-for-avs/

Resources:

The nuScenes dataset [7] will be exploited for experimentation. Testing with state of the art metrics should assess the validity of explored solutions proposals. An interpolation method on depth data [8] and a 2-branch DL architecture with GCL [9] are proposed as starting points to evolve towards the needs of the current project. Previous projects exploring neural architectures on pointclouds in automation [10] will contribute experience in state of the art pointcloud networks and 3D data annotation.

[7] H. Caesar *et al.*, "nuScenes: A Multimodal Dataset for Autonomous Driving," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 11618–11628. doi: 10.1109/CVPR42600.2020.01164

[8] I. Caminal, J. R. Casas, and S. Royo, "Slam-Based 3D Outdoor Reconstructions From Lidar Data," in *2018 Int. Conf. 3D on Immersion (IC3D)*, Dec. 2018, doi: 10.1109/IC3D.2018.8657869

[9] M. Hu *et al*, "PENet: Towards Precise and Efficient Image Guided Depth Completion," in *2021 IEEE Int. Conf. on Robotics Automation (ICRA)*, May 2021, doi: 10.1109/ICRA48506.2021.9561035

[10] Ò. Lorente, J. R. Casas, S. Royo, and I. Caminal, "Pedestrian Detection in 3D Point Clouds using Deep Neural Networks," *arXiv:2105.01151 [cs]*, May 2021, http://arxiv.org/abs/2105.01151

Student background:

This proposal is targeted as a Master Thesis project. The candidate should have experience in Computer Vision, 3D data processing (depth images, pointclouds), Deep Learning, Calibration/registration technologies, C++/python.

Supervisors:

Josep R. Casas, UPC, josep.ramon.casas@upc.edu
José A. Lázaro, UPC, jose.antonio.lazaro@upc.edu