Workshop 2023
ePicture This
an initiative by PENTA label projects
MANTIS and IMAGINATION with AENEAS support

ABSTRACT PRESENTATION

Title: Multi-sensor data fusion for VRU detection
Presenter name: Ljubomir Jovanov
Company name / Institute: imec-Ghent University
Project name: NextPerception
Funding group: ECSEL
Abstract can be published on website: ☒ **YES** ☐ **NO**

To further advance the road safety, the European Commission has set the goal of zero traffic deaths on EU roads by 2050. An important role in achieving these goals belongs to the intelligent infrastructure. As a part of NextPerception project, we are developing algorithms for sensor fusion, aimed at improving the accuracy of detection of vulnerable road users (VRU). We further use this rich representation of the surroundings to detect, track and predict the behaviour of VRU.

Primarily we rely on visible light cameras, radar, thermal camera and lidar. While visible light cameras may enable a sufficient level of detection accuracy in daylight conditions, challenging illumination conditions require additional information to keep the level of detection performance at the acceptable level. For this reason, we introduce additional sensors like radar and thermal cameras in the system. Automotive radars provide better robustness in challenging atmospheric conditions like rain, fog, smog, as well as in poor lighting conditions. The main drawback of radar sensors is their inherently low angular resolution.

None of the above-mentioned sensors is ideal nor it can operate reliably in all conditions. Therefore, sensor fusion algorithms must operate in such a way that all information is used in an optimal way. Moreover, sensor fusion algorithms must be able to detect situations when some of the sensors produce faulty outputs, due to the external conditions or sensor failure, and to assign this data lower reliability.

Within the NextPerception project, we have first developed a radar-video fusion method for VRU detection. This method relies on two ResNet-18 based feature extractors for both radar and visible light data. In the second stage, the proposed method performs semantic segmentation of the RGB images from the camera and depth map estimation based on monocular data. In the final stage, we employ two feature extraction networks, for RGB images and depth maps and finally perform object detection based on the fused features. Besides detection, the proposed method generates depth maps, which have distance measurements associated to each pixel in the depth maps, which are also useful for ADAS applications, like assisted parking.

We also develop a new method for fusion of visible and thermal video streams for VRU detection. In the first stage, we rely on two feature extractors to acquire salient features from RGB and thermal images. Feature maps generated in the first step are then fed into cross modality transformer, which performs association of features from different modalities. Next, we feed these features into region proposal network, to obtain object locations. Finally, Multimodal Fusion Transformer performs the association of generated regions of interest from different modalities and determines the locations of the detected objects. The main strength of the proposed method is a novel fusion approach, where we treat all detections from different modalities as a potential source of information and associate them according to the semantic criteria.