Real-time capable stereo-vision based 3D object detection

Student

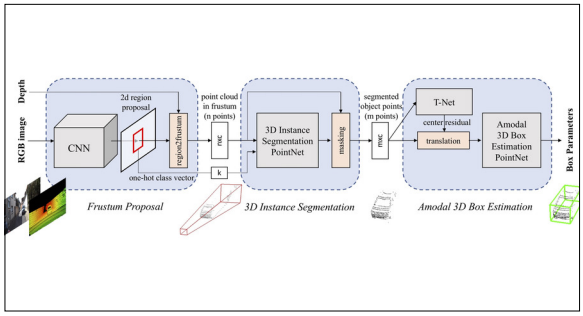
Philip Zellweger

Introduction: The ability to perceive objects in real-time in the environment is becoming increasingly crucial for robots and mechatronics system solutions. For example, in the field of BIM (Building Information Modelling), online object detection in combination with a mapping algorithm for the environment can realize an automatic process monitoring of the built structure. Stereo-vision-based systems offer a low-cost solution for the virtual reconstruction of the environment. In combination with deep neural networks, the generated 3D data is processed. Objects in the virtual environment are classified and localized.
This thesis develops a concept that enables online 3D object detection. The Object detection is focused on indoor scenes. Based on state of the art approaches, a network architecture is selected that fulfils the requirements of real-time capability and high detection accuracy at the same time.

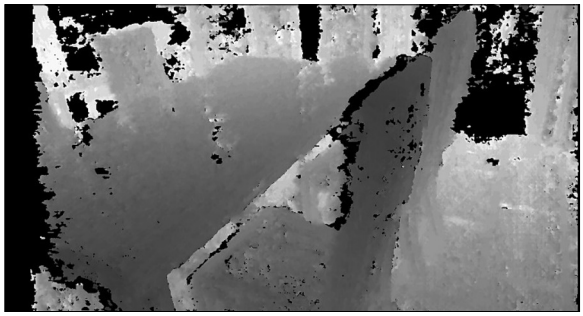
Approach / Technology: To perform 3D object detection, the Frustum PointNets architecture extended with the YoloV3 architecture as 2D detector is used. The combination is chosen because of the high opportunity for a fast inference performance. The two-stage 2D to 3D detection approach offers to split the network into a primary and secondary detector. This will be used to improve processing performance over the whole system, where only 3D points inside the detected 2D bounding boxes will be reconstructed. Related to this work, the YoloV3 architecture is implemented in the first step.

Conclusion: The presented system architecture offers real-time and accurate 3D object detection. The concept based on FrustumPointNets for 3D object detection must be realized and evaluated in a next step. An end-to-end evaluation of the 2D/3D detector architecture is needed. After training on SUN-RGBD dataset, the network will be transfer-learned on own captured data to improve prediction performance. Step by step, new objects must be trained to saturate the network capacity thoroughly.

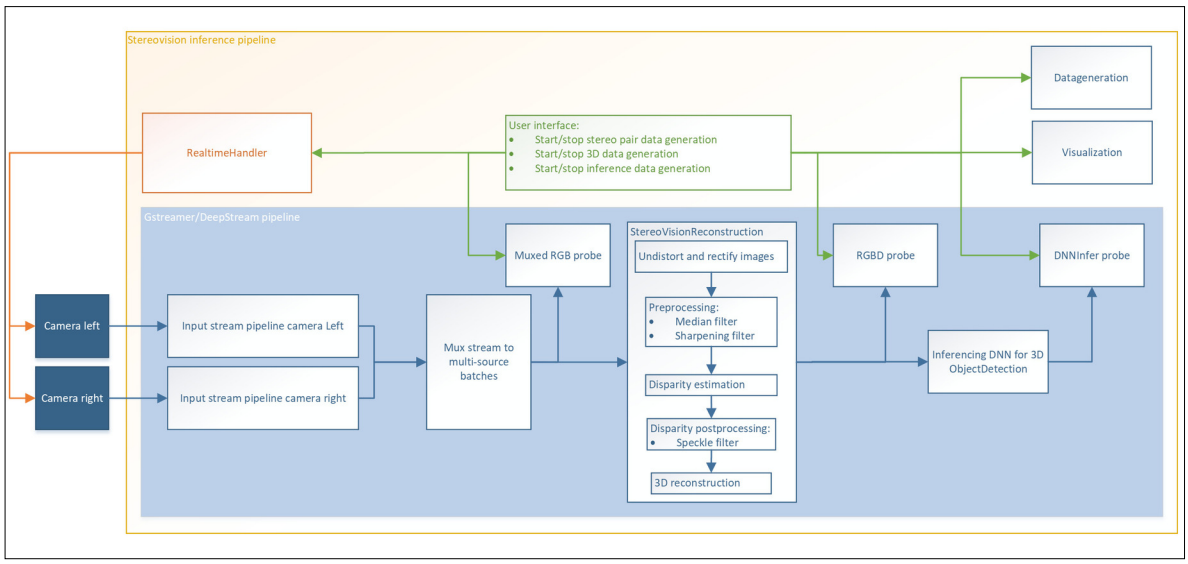
Frustum PointNets for 3D Object Detection from RGB-D Data
https://ieeexplore.ieee.org/document/8578200



Depth map of own captured indoor scene
Own presentment



System architecture real-time inference pipeline
Own presentment



Examiner
Prof. Dr. Dejan Šeatović

Subject Area
Mechatronics and Automation, Software and Systems, Data Science