Precise semantic segmentation from stereo-images for autonomous navigation

Supervisors : Ezio MALIS, Patrick RIVES, Philippe MARTINET

Centre Inria d'Université Côte d'Azur - ACENTAURI in Sophia Antipolis

Research team

ACENTAURI is a robotic team that studies and develop intelligent, autonomous and mobile robots that can help humans in their day-to-day lives at home, at work or during their displacements. The team focuses on perception, decision and control problems for multi-robot collaboration by proposing an original hybrid model-driven / data driven approach to artificial intelligence and by proposing efficient algorithms. The team focuses on robotic applications in smart territories, smart cities and smart factories. In these applications several collaborating robots will help humans by using multi-sensor information eventually coming from infrastructure. The team demonstrates the effectiveness of the proposed approaches on real robotic systems like cars AGVs and UAVs together with industrial partners. Innovation and the transfer of the research work towards industrial partners are a concern of ACENTAURI.

Motivations and general objectives

As part of the development of the autonomous vehicles in the ACENTAURI team at Inria Sophia Antipolis, we are interested in a module for performing the semantic segmentation of the environment from stereo images.

Autonomous cars need to build a representation of the environment in which they navigate. Geometric reconstruction only may not be sufficient for safe navigation in dynamic environments. In this context, it is important to have a semantic classification at the pixel level of the available video streams in order to identify the navigable space and the moving objects around the vehicle in order to anticipate the evolution of the environment in which the vehicle moves. The use of bounding boxes does not offer sufficient precision to the system and a more precise clipping is then necessary.

Semantic segmentation is a field that has seen important advances in recent years. The apparition of fully convolutional networks [1] has allowed both to reduce the number of parameters of the architectures, thus accelerating the computation while reducing the memory cost, but also to work with images at different resolutions. The following works such as SegNet [2] and U-Net [3] have studied in more depth how to transmit information between the convolution and deconvolution layers allowing the generation of semantic maps at any scale level. Indeed, at the first layers of the network, the use of max-pooling functions results in the loss of part of the localization information. It is therefore necessary to propagate part of the information to allow an accurate reconstruction of the semantic map. For example, [4] introduced ParseNet to add local scene context knowledge. Recurrent networks such as Conv-LSTM layers have been used in [5] to improve video segmentation. Some works also use mechanisms from object detection to apply them to segmentation such as Mask-RCNN [6].

The objective of this master internship, after writing a precise state of the art of semantic segmentation methods applicable to stereo-vision videos, is to develop new architectures to meet the specific constraints in this context. Indeed, we aim at taking full advantage of the stereo images as well as of the previous annotations from the previous frames of the video.

Work-plan

The work will be decomposed with incremental steps as follows:

1. Bibliography on semantic segmentation methods applicable to stereo-vision,

2. Mathematical background in machine learning and computer vision

3. Implement a state of the art algorithm for semantic segmentation

4. Implement a new architecture for stereo-vision sequences

5. Provide comparison between the two algorithms

6. Develop real experimentation with our fully equipped autonomous car

7. Writing master Thesis and potential papers

Skills

The candidate is expected to follow a Master in Control, or Computer Sciences, more generally in robot-ics, as well as solid skills in software development (MATLAB/Simulink, LINUX, ROS, Git, PYTHON, C/C ++), and mathematics. A good level of written/spoken English is also important.

How to apply

Interested candidates must send a detailed CV, transcripts bachelor, M1 and M2, a motivation letter and at least one recommendation letter to ezio.malis@inria.fr

Financial support

Financial support offered to the student: around 580 € per month during 6 months.

References

1 J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),pages 3431–3440, 2015.

2 V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet : A deep convolutional encoder-decoder archi-tecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12) :2481–2495, 2017.

3 Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net : Convolutional networks for biomedical image segmentation. volume 9351, pages 234–241, 10 2015.

4 Wei Liu, Andrew Rabinovich, and Alexander Berg. Parsenet : Looking wider to see better.arXiv pre-print arXiv :1506.04579, 2015.

5 A. Pfeuffer, K. Schulz, and K. Dietmayer. Semantic segmentation of video sequences withconvolu-tional lstms. In2019 IEEE Intelligent Vehicles Symposium (IV), pages 1441–1447,2019.

6 K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In2017 IEEE International Conference on Computer Vision (ICCV), pages 2980–2988, 2017.