A Study on Semantic Segmentation for Autonomous Vehicles

Vinícius Almeida dos Santos
viniciusas@edu.univali.br
Universidade do Vale do Itajaí
Itajaí, Santa Catarina, Brasil

Thiago Felski Pereira
felski@univali.br
Universidade do Vale do Itajaí
Itajaí, Santa Catarina, Brasil

Rodrigo Lyra
rodrily@gmail.com
Universidade do Vale do Itajaí
Itajaí, Santa Catarina, Brasil

ABSTRACT

Autonomous vehicles are already a reality, and there are still several challenges to overcome. One important challenge for the adoption of these vehicles is perceiving its surroundings. This necessity of perception can be fulfilled by digital cameras. When working with digital image processing, the quality will be limited by real-time constraints. As several works indicate, this real-time constraint for autonomous vehicles is at most 100ms per frame. Also, by improving the processing time, the chances of accidents involving autonomous vehicles may be decreased. This paper analyses the advantages and drawbacks of semantic segmentation and also presents a study to implement perception for autonomous vehicles by accelerating a semantic segmentation algorithm, also used by other works on the field. To accelerate the algorithm, spacial parallelism will be used.

KEYWORDS

Digital Image Processing, Semantic Segmentation

1 INTRODUCTION

An vehicle may be considered autonomous once its capable of moving and taking decisions in an intelligent manner and without the need of a pilot [1]. Even so, there are several levels of automating vehicles, as stated by [2] informative (Society of Automotive Engineers). There are several aspects to fulfill in order to achieve a fully autonomous vehicle, which is the fifth class of the SAE [2] informative.

The aspects to consider a vehicle as autonomous are: perception of surroundings, route planning, and vehicle control. Perception, as a dependency for the others, is already a big challenge. This paper is concerned on the use of images from cameras for perception.

A large amount of information can be processed from processing digital images. However, a lot of processing may be required to do so. This may bring several complications considering that autonomous vehicle are subject to a realtime constraint.

2 REALTIME

From the perception to the decisions of an autonomous vehicle, it must respond in time for reacting to potentially dangerous situations. Therefore, it is considered an critical realtime system, where any violation of the time constraint may prove itself fatal [3].

In the case of autonomous vehicles, the control system must respond within the maximum time of 100 milliseconds, which represents 10 FPS. No articles were found on the literature clearly stating the realtime constraint. However, several works consider this time (100 ms) enough to attend realtime [4–9].

Representing the need of realtime, per example, the vehicle must not encounter problems when making turns, changing lane, activating breaks to avoid accidents, breaking for pedestrian, etc. All examples involves from perception to the vehicle control, which means that the perception task must run in less than the constraint limit.

3 SEMANTIC SEGMENTATION

Semantic segmentation is a task made with the objective of segmenting relevant objects from an image [10]. An example in autonomous vehicles, would be segmenting an image in pedestrians, road and vehicles. Its different from usual image segmentation, in which segments the image by similarity or discontinuation of image characteristics. On this semantic segmentation, each pixel in the image will have a label to identify.

With the resulting image, there are several data that is not available when using bounding box based methods. On bounding box methods, the only information available is a rectangle identifying the location of objects. A problem with this approach is that roads, for example, that can occupy most part of an image, is not necessarily well represented by a rectangle.

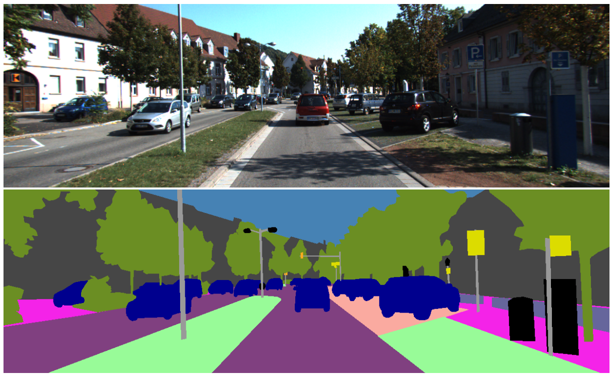

Figure 1: Sample and ground truth image [11] from KITTI dataset [12]

In the Figure 1, there are two images. The first image indicates a sample image from the dataset, whereas the second image indicates the exact labels for the first image [11]. This wide image is generated by using fisheye lens on the camera, where the image suffers from a fixable distortion. The dataset were first made by Geiger et al. [12], and were adapted by Alhaija et al. [11] to enable its use for semantic segmentation.

When applying semantic segmentation on an image, specific objects are searched. According to Uijlings et al. [13], one advantage of methods based on regions is the possibility of better accuracy for detection of objects like grass, sky, and water, which do not have a a specific format like other objects, such as cars, traffic signs, pedestrians, etc.

One algorithm to implement semantic segmentation, some options are the approaches of Girshick et al. [14] and Ren et al. [15]. These approaches use proposed regions provided by selective search [13]. There are other approaches like the one of Caesar et al. [10], which seeks to discover the class of each pixel in the image, which also obtains great results.

On the approaches of Girshick et al. [14] and Ren et al. [15], the worst bottleneck is the selective search execution. Therefore, to accelerate the execution of these algorithm, this paper seeks to accelerate selective search execution time by using spacial parallelism.

3.1 Selective Search

Selective search [13] is an algorithm used by Girshick et al. [14], which can be used for semantic segmentation. Selective search use a graph based segmentation algorithm by Felzenszwalb et al. [16], and by using several diversification strategies, generate initial regions to start the algorithm. After that, the algorithm uses several similarity metrics between the regions, merging them hierarchically, which increase the amount of proposed regions. The hierarchical grouping runs until a big interest region is merged.

3.2 Spacial Parallelism

To accomplish the realtime constraint, several approaches can be used. One of these is using spacial parallelism, which separates algorithms in parts that can be executed simultaneously. The acceleration depends on processors and coprocessors that support several executions [17], like multicore CPUs or GPUs. Spacial parallelism, is widely used in image processing and computer vision applications.

4 DEVELOPMENT

To apply spatial parallelism for the selective search [13] algorithm, initially, the graph segmentation [16] were analyzed. Unfortunately, this algorithm could not be parallelized because each iteration depends on the previous one, which inhibits the simultaneous execution. The approach taken is parallelize the selective search algorithm directly.

The selective search algorithm [13] can be parallelized at the level of its several diversification strategies. Since each diversification strategy runs without dependency from each other, each iteration can run in a different CPU core. However, because the large complexity of running the graph segmentation [16], its not possible for the iterations to run on the GPU.

5 TESTS

The selective search algorithm [13] were implemented in C++, with OpenMP for parallelization. The only diversification strategy used is the variation of the k parameter of the graph segmentation [16] from 50 to 300 with a step of 50, as described on Uijlings et al. [13] paper. The input image used were reduced to 600x500. The computer used for the tests have an intel i7 processor with 4 cores.

Running in 686 milliseconds with 4 threads, the speedup is reasonable, as represented on the Figure 2. Considering that running single core, the spent time was 1219 milliseconds.

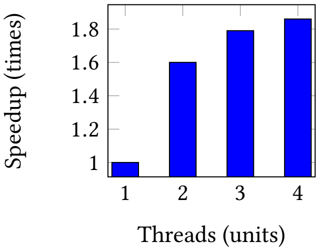

Figure 2: Selective search speedup

6 FINAL REMARKS

The method R-CNN from Girshick et al. [14], using [13], have good results on the literature for classification, recognition, and semantic segmentation. Speeding up its execution is essential to enable it to problems with realtime constraints. On tests, we did not achieved realtime constraints, however, by reducing the input image, its easily achievable. For a future work, the next step is evaluating the reduced quality of the solution compared to execution time. Other possible future works are: Use of different segmentation methods, in order do achieve a better time; Experimenting other diversification strategies;

ACKNOWLEDGMENTS

This research was sponsored by Artigo 170 from Uniedu research scholarship.

REFERENCES

[1] Ingemar J. Cox and Gordon T. Wilfong, editors. *Autonomous Robot Vehicles*. Springer-Verlag, Berlin, Heidelberg, 1990. ISBN 0-387-97240-4.

[2] SAE. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. Standard, SAE - Society of Automotive Engineers, Janeiro 2014.

[3] Fan Liu, Ajit Narayanan, and Quan Bai. Real-time systems, 2000.

[4] Margrit Betke, Esin Haritaoglu, and Larry S. Davis. Real-time multiple vehicle detection and tracking from a moving vehicle. *Machine Vision and Applications*, 12(2):69–83, Aug 2000. ISSN 1432-1769. doi: 10.1007/s001380050126. URL https://doi.org/10.1007/s001380050126.

[5] Jaesik Choi. Realtime on-road vehicle detection with optical flows and haar-like feature detectors. 2012.

[6] H. Lyu, H. Fu, X. Hu, and L. Liu. Esnet: Edge-based segmentation network for real-time semantic segmentation in traffic scenes. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1855–1859, Sep. 2019. doi: 10.1109/ICIP.2019.8803132.

[7] Adam Paszke, Abhishek Chaurasia, Sangpil Kim, and Eugenio Culurciello. Enet: A deep neural network architecture for real-time semantic segmentation. *CoRR*, abs/1606.02147, 2016. URL http://arxiv.org/abs/1606.02147.

[8] Michael Treml, Jose Arjona-Medina, Thomas Unterthiner, Rupesh Durgesh, Felix Friedmann, Peter Schuberth, Andreas Mayr, Martin Heusel, Markus Hofmarcher, Michael Widrich, Bernhard Nessler, and Sepp Hochreiter. Speeding up semantic segmentation for autonomous driving. 12 2016.

[9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. URL http://arxiv.org/abs/1409.1556.

[10] Holger Caesar, Jasper Uijlings, and Vittorio Ferrari. Region-based semantic segmentation with end-to-end training. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 381–397, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46448-0.

[11] Hassan Alhaija, Siva Mustikovela, Lars Mescheder, Andreas Geiger, and Carsten Rother. Augmented reality meets computer vision: Efficient data generation for urban driving scenes. *International Journal of Computer Vision (IJCV)*, 2018.

[12] A Geiger, P Lenz, C Stiller, and R Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013. doi: 10.1177/0278364913491297. URL https://doi.org/10.1177/0278364913491297.

[13] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders. Selective search for object recognition. *International Journal of Computer Vision*, 104(2):154–171, Sep 2013. ISSN 1573-1405. doi: 10.1007/s11263-013-0620-5. URL https://doi.org/10.1007/s11263-013-0620-5.

[14] Ross B. Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *CoRR*, abs/1311.2524, 2014.

[15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 91–99. Curran Associates, Inc., 2015. URL http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf.

[16] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International journal of computer vision*, 59(2):167–181, 2004.

[17] Donald G Bailey. *Design for embedded image processing on FPGAs*. John Wiley & Sons, 2011.