The Superpixel Superhighway: A Comparison of Machine Learning Algorithms for Road Classification

Jacob Lambert
Institute for Aerospace Studies
University of Toronto
jacob.lambert@mail.utoronto.ca

Rikky Duivenvoorden
Institute for Aerospace Studies
University of Toronto
rikky.duivenvoorden@mail.utoronto.ca

Abstract

In this paper, we show how to solve robotics using machine learning.

1 Introduction

1.1 Related Work

2 Dataset

2.1 KITTI

The dataset used in these experiments is provided by the Karlsruhe Institute of Technology (KIT) and Toyota Technological Institute (TTI) in Chicago[1]. This vision benchmarking suite has become a standard for comparing the performance of algorithms for tracking, navigation, estimating scene flow and object segmentation and classification.

The dataset for road classification is seperated in three different environments: urban unmarked (uu), urban marked (um) and urban multiple marked lanes (umm), each shown in figure 1. Each set contains roughly a hundred RGB images of size 375×1242 pixels, taken from a monocular camera mounted on top of a car.

Figure 1: The three environments in the KITTI dataset: uu (left), um (middle) and umm (right)

The urban unmarked environment is the most challenging of the three as the typical images are of narrow roads in dense city environments. There are several obstructions, some static such as shadows, parked vehicles or street signs, and others dynamic like pedestrians, cyclists and moving vehicles. There are also several intersection parking lots or driveways which are visually very similar to roads, but they are not included in the ground truth image. Figure 2 shows the ground truth image for the left most image in figure 1. The urban marked dataset features similar scenes but with road markers, delimiting lanes and often the side of the road, making the segmentation and classification task slightly easier. urban multiple marked lanes contains images of larger roads such as boulevards or highways. The scenes in this dataset are typically much less cluttered than the former two and the roads have clear delimiters.

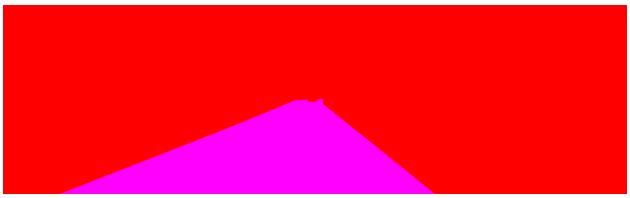

Figure 2: Ground truth image for the left-most image in figure 1 provided in the KITTI dataset. Pink pixels are classified as road and red pixels as non-road.

2.2 Superpixel Segmentation

While the classification could be performed for each pixel, the approach taken in this work is to first segment each image into a set of superpixels. There are several advantages to this, the first and foremost is that it heavily reduces training time; there are approximately 46 million pixels per dataset. If we can segment the image into a set of similar looking patches, we heavily reduce the number of training examples without losing much information. Furthermore, the information provided by individual pixels is limited by the RGB color intensities and the pixel position. A set of pixels allows us to compute a variety of features, as dicussed in section 2.3. Superpixel segmentation essentially allows us reduce the number of examples while increasing the dimensionality of each examples.

Here we use the Simple Linear Iterative Clustering (SLIC) algorithm[2] to segment the each image in approximately 750 superpixels. The algorithm creates superpixels by clustering pixels of similar colors. Intuitively, the approach is adequate for segmenting roads as we expect them to be uniform in color but figure 3 shows that SLIC performs poorly in a variety of scenarios. Notable examples are driveways which SLICO fails to segment in the top image of figure 3. In the middle image, SLIC manages to segment the road from the grassy region on the left, but fails to properly differentiate road from sidewalk on the right. In the bottom image, SLIC segments all shadows in the scene regardless of the actual scene content. The images in each respective datasets are similar and as such there were many cases of poor segmentation which inevitably led to several superpixels representing both road and not-road. Classification on those superpixels is especially challenging, as shown in section 4

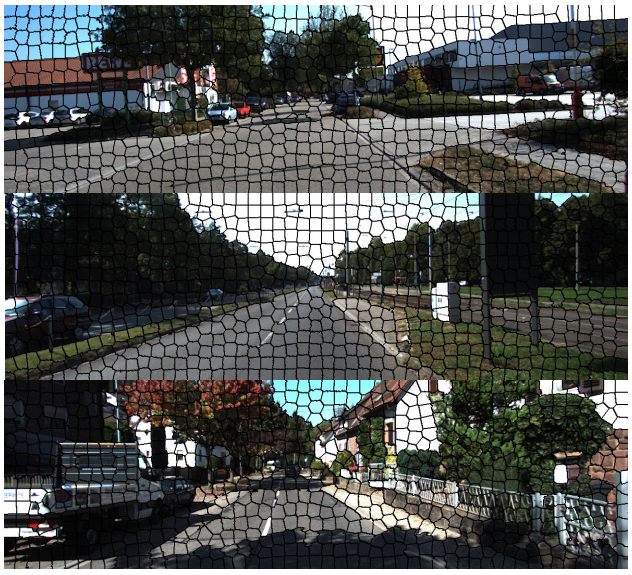

Figure 3: Performance of the SLIC algorithm on the uu dataset (top), um dataset (middle) and umm dataset .

As the segmentation is imperfect, some superpixels will inevitably contain positive and negative examples. To solve this issue, a new ground truth image was created (shown in figure **??**), where each superpixel was classified based on the majority of pixels in its cluster.

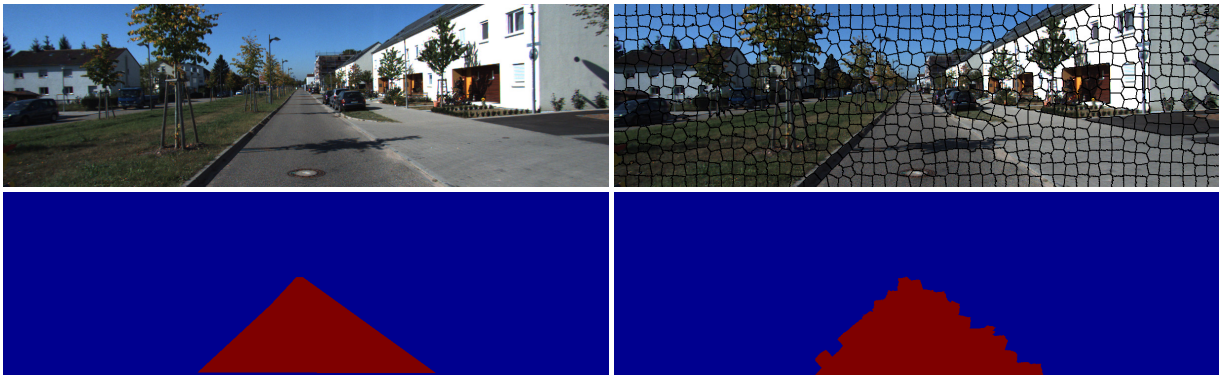

Figure 4: The image segmentation pipeline is shown here. The starting image (top left) is first segmented into superpixels by SLIC (top right) then the label (bottom left) is adjusted to be uniform within superpixels (bottom right).

2.3 Feature Selection

The types of feature computed for each superpixel is a critical step in the classification pipeline as they need to convey enough information about the superpixel to allow classification. In this experiment chose a total of 14 simple and intuitive features. An example $\mathbf{x}^{(i)}$ is stored as

$$\mathbf{x}^{(i)} = \left[\mathbf{r}^{(i)} \;\; \boldsymbol{\mu}^{(i)}_{RGB} \;\; \boldsymbol{\sigma}^{(i)}_{RGB} \;\; \boldsymbol{\mu}^{(i)}_{\nabla RGB} \;\; \boldsymbol{\sigma}^{(i)}_{\nabla RGB}\right]. \tag{1}$$

$\mathbf{r}^{(i)} = \left[x^{(i)} \;\; y^{(i)}\right]$ is the centroid of the superpixel. For the ith superpixel composed of N pixels each with location $\mathbf{r}^{(i)}_n$, the centroid is given by

$$\mathbf{r}^{(i)} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{r}^{(i)}_n. \tag{2}$$

Then, $\boldsymbol{\mu}^{(i)}_{RGB} = \left[\mu^{(i)}_R \;\; \mu^{(i)}_G \;\; \mu^{(i)}_B\right]$ and $\boldsymbol{\sigma}^{(i)}_{RGB} = \left[\sigma^{(i)}_R \;\; \sigma^{(i)}_G \;\; \sigma^{(i)}_B\right]$ are the respective pixel intensity mean and variance for Red, Blue and Green color channels. Finally, we compute the gradient magnitude of the image in each color channel, which produces an image shown in figure 5. As computing the gradient picks out high frequency color variations, we expect it to pick out edges or high irregular regions like vegetation. Roads which are uniform should have low gradient magnitude and even lower variation. $\boldsymbol{\mu}^{(i)}_{\nabla RGB}$ and $\boldsymbol{\sigma}^{(i)}_{\nabla RGB}$ are then the mean color and variation for each color channel in the gradient image.

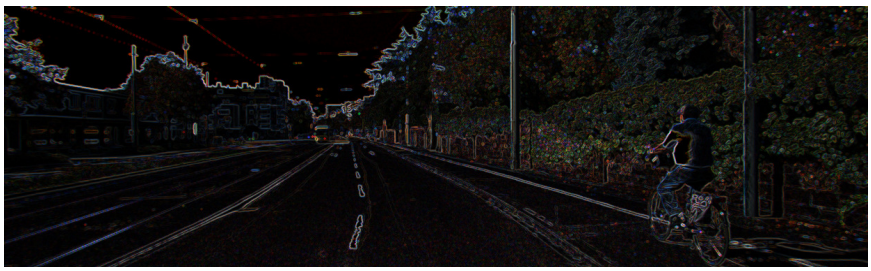

Figure 5: Gradient magnitude of one of the images.

3 Algorithms

3.1 *k* Nearest Neighbors

3.2 Support Vector Machines

3.3 Neural Network

4 Results

5 Discussion

References

[1] A. Geiger, P. Lenz & R. Urtason, (2012) Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. *Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI.

[2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, & S. Ssstrunk, (2012) SLIC Superpixels Compared to State-of-the-art Superpixel Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **34**(11):2274-2282.