Performance Analysis of Symbolic Road Recognition for On-road Driving

M. Foedisch, C. Schlenoff, and R. Madhavan†
National Institute of Standards and Technology (NIST)
Gaithersburg, MD 20899-8230, USA
{mike.foedisch, craig.schlenoff, raj.madhavan}@nist.gov

Abstract — Previous approaches to road sensing, namely road detection were based on segmenting the sensor data, i.e. color camera image, into road and non-road areas. Performance evaluation for such algorithms could be performed in a relatively straightforward fashion by comparing the algorithm's result with ground truth. Ground truth for such an image-based evaluation approach could be limited to a geometrical structure describing the road area in the original image. However, the development of our new high-level road sensing approach, which is a model-based approach to road recognition, makes new demands to performance analysis and subsequent performance evaluation which would include comparison with ground truth. In this paper[1], we will briefly describe the new road recognition approach, show performance analysis results and discuss performance evaluation issues.

Keywords: *autonomous driving, model-based perception, road recognition.*

I. INTRODUCTION

Previous approaches to road sensing, namely road detection ([4], [5], [10]) were based on segmenting the sensor data, i.e. color camera image, into road and non-road areas. Performance evaluation for such algorithms could be performed in a relatively straightforward fashion by comparing the algorithm's result with ground truth (see [7], [10]). Ground truth for such an image-based evaluation approach could be limited to a geometrical structure describing the road area in the original image. However, the development of our new high-level road sensing approach [3], which is a model-based approach to road recognition, makes new demands to performance analysis and subsequent performance evaluation which would include comparison with ground truth.

There are several approaches to performance evaluation which can be classified into the following general categories: *comparative* evaluation compares the algorithms performance with similar other algorithms or a ground truth; for *analytic* evaluation the limits, computational complexity and theoretical optimality of the algorithm may be determined; the *performance* on test data and execution times with different parameters may measured; and finally the *appropriateness to the task* can be analyzed given the context of a particular application with its constraints (please refer to [7] for a more detailed discussion on the subject).

We present in this paper a two-level approach to performance analysis for a new road recognition approach providing symbolic descriptions of the road structure. The first level of performance analysis helps point out potentially problematic areas and real-time issues by analyzing the behavior of the tree search-based recognition approach. On the second level an actual performance evaluation is performed by comparing the symbolic results of the algorithm against (semi-) automatically extracted ground truth.

This paper is organized as follows: In Section II we briefly describe the new symbolic road recognition approach. In Section III we introduce the first level of performance analysis and discuss results. Finally, in Section IV we outline the second level which is actual performance evaluation employing ground truth and discuss issues with the automatic ground truth generation as well as results.

II. SYMBOLIC ROAD RECOGNITION APPROACH

Our previous work on road detection on color images demonstrated the advantages of using background knowledge (in terms of models) in order to improve the recognition results [5]. In the following, we will describe a new approach of a model-guided road recognition process [3] and will discuss the type of extracted features, the representation of models, the recognition process, and the representation of the resulting symbolic road data.

A. Feature Extraction

One important assumption of the new approach concerns the orientation of the vehicle on the road. A normal orientation, where the vehicle is limited to traverse only in lanes for which the legal driving direction agrees with the vehicle's direction, provides a canonical form for the appearance of road on images and may therefore simplify the representation process. All other orientations of the vehicle do not comply with the

[1]Commercial equipment and materials are identified in this paper in order to adequately specify certain procedures. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

†Research Staff Member, Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. Prepared by Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6285, managed by UT-Battelle, LLC, For the U.S. Department of Energy under contract DE-AC05-00OR22725. This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC-05-00OR22725 with the U.S. Department of Energy.

normal orientation. We can allow the limitation of a normal orientation if we assume that the autonomous system will be aware of when it leaves the normal orientation (e.g. due to avoidance of obstacles on the road).

Assuming normal orientation of the vehicle on the road, a simple set of features, which are easily extracted and well-understood, can be derived [2]. The features are based on "slices" of the road perpendicular to the direction of the vehicle. They can be extracted by applying one of several approaches for detecting the road area in images or road edge detecting algorithms (e.g. [1], [2], [5], [4], [8]).

Starting at the bottom image row, the left and right road edge points in each row are determined. A pair of road edge points described in both image and world coordinates (through camera calibration) describes one feature item. The process continues bottom-up row-by-row until the world coordinates of the road edges reach a given maximum distance in front of the vehicle (e.g. more than 55 m). Furthermore, additional data will be associated with a feature item, e.g. information about lane markings.

B. Model Representation

Figure 1 depicts our approach for representing road model primitives. A "slice" of road is described by its width (geometrical component) and lane structure in terms of number of lanes and their legal directions (topological component). This representation of road primitives is compatible with the type of feature data described in Section II-A.

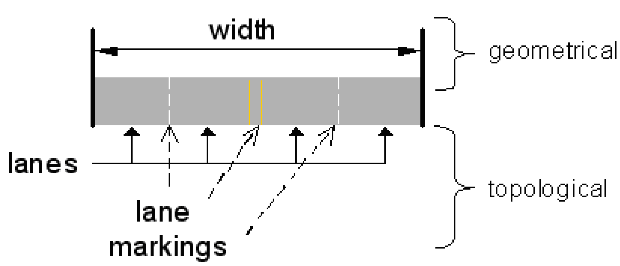


Fig. 1. Geometrical and topological representation of a "slice" of road.

A road type consists of an ordered group of primitive road model items. For such groups additional constraints apply. A road type might require a minimal and/or maximal lateral length or, in the case of road widening and narrowing, a certain monotonic behavior. Other constraints limit the connectivity between (primitive) road types, e.g. a two-lane road segment can connect to a three-lane road segment only through a transitional segment. Primitive road items and road types are organized hierarchically. Additionally, primitive model items are grouped by the type of driving environment, e.g. highway driving, rural road or urban road driving. Appropriate connectors describe transitions from one environment to another (e.g. a highway exit transfers the vehicle from highway driving to rural road driving).

C. Recognition Process

The goal of the recognition process is to find associations between feature items and (primitive) road models and eventually an interpretation of the scene. The application of a tree search algorithm spans potentially all possible associations of feature items and road models [6]. This process, however, is computationally expensive and must therefore be constrained. We define constraints on three different levels, the primitive associations level, the group level, and the symbolic-level interpretation.

On the primitive level, potential associations must comply with unary constraints. For example, in order to associate a feature item to a specific road model, the width of the road has to be similar for both entities. Whenever a feature item is associated with the same model as the previous feature item, group-related constraints apply. Assuming that feature items *1*, *2*, and *3* in Figure 2 are already associated with a model *A*, the association of feature item *4* to model *A* requires compliance of the extended *Group A* to group-related constraints. For example, in the case of a road widening (as part of an intersection) the group should comply with a certain monotonic behavior and the group's length should be within the maximal length of the model. Assuming another situation where feature items *1-4* are already associated to model *A*, associating feature item *5* with model *B* would trigger additional constraints. Starting a new group *B* causes the previous group *A* to be closed. This, for example, requires compliance with the minimum length constraint.

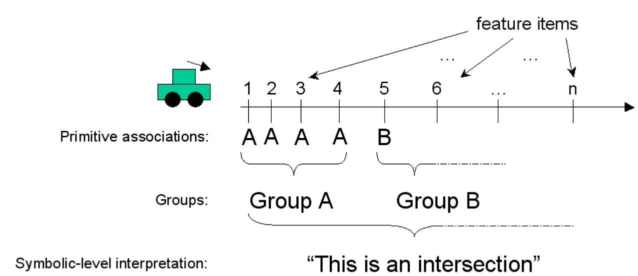


Fig. 2. Recognition process levels: primitive associations level, group level, and symbolic-level interpretation.

Finally, the set of (locally) consistent groups may allow a high-level interpretation of the scene. For example, the occurrence of a regular road segment, a widening segment, a narrowing segment, and another regular road segment (in this order) can be a strong indicator for the existence of an intersection.

Figure 3 depicts an example of a search tree used for the recognition process. On each level of the tree, one single feature item is associated with (potentially) all known models (within the current driving environment, see Section II-B). This potentially huge tree structure (considering all possibilities) will be reduced in numbers of nodes by the above described application of constraints. Branches in the resulting tree that show consistent associations of feature items to models from the root to a leaf of the tree represent surviving

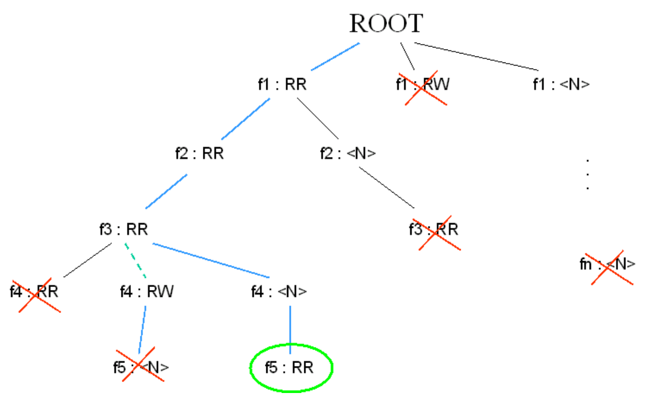


Fig. 3. Sample search tree. The feature items f1 - f5 are associated to the models RR (Regular Road), RW (Road Widening) and $< N >$ (for noise). The branches of the search tree are being pruned whenever the associations are inconsistent on the local, group or global level. Paths reaching from the ROOT to one of the leaves are considered interpretations (e.g. blue path to the green circle).

interpretations.

We use the number of nodes and the number of interpretations as measures for performance analysis described in Section III.

D. Symbolic Representation of Road Structures

Figure 4 shows three examples of the symbolic description of our approach's recognition results. Each node describes one road segment's road type, e.g. node A1 in Figure 4(a) describes a straight road segment of a bi-directional two-lane road. The nodes also contain geometrical information such as the road width and segment length. Due to sensor limitations, however, geometrical measures give only a coarse impression and their interpretation should be considered carefully. The examples in Figure 4(b) and Figure 4(c) show more complex road structures. The occurrence of multiple road segments of several types is represented by a chain of nodes.

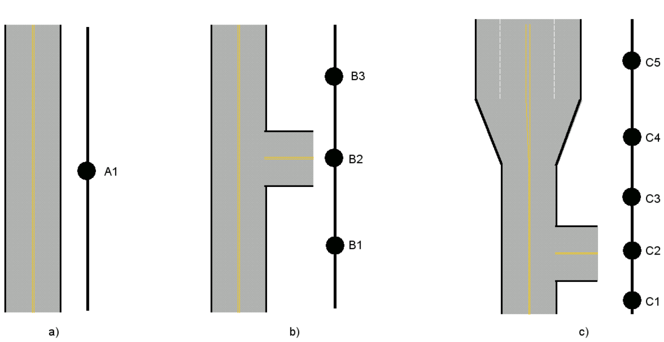


Fig. 4. Examples for symbolic description of recognition results. Each Node is of a certain type, e.g. *Regular Road, two lanes, bi-directional* (A1, B1, B3, C1, and C3), *T-Intersection, from right* (B2 and C2) etc.

III. Performance Analysis

As described in Section II-C, we use a constrained tree search approach for our high-level recognition process. Each execution of tree search can be described by internal parameters describing the resulting search tree structure - the *number of nodes* and the *number of interpretations*. We use both values to gain a first impression of the recognition system's performance.

Figure 5 shows performance analysis results. Figure 5(a) shows the first frame of a test video sequence. In the background the original input image is depicted, in the upper right corner the result of the underlying road detection, in the lower center the most compressed representation of the symbolic results (for the left and right side of the road separately), and on the right side an iconic depiction of the symbolic results can be seen. The graph in Figure 5(b) shows the number of nodes (blue) and the number of surviving interpretations (yellow) for each frame of the test sequence.

From experiments we learned that a typical successful run of our system results in search trees of about a few hundred nodes and about one interpretation. The graph in Figure 5(b), however, shows (in the first half of the sequence) the occurance of a magnitude higher number of nodes (> 2000) as well as sporadic lack of any interpretations. We consider these as clear indicators of problems with the recognition algorithm's performance for the following reasons:

- no interpretations mean lack of results and, therefore, complete failure of the algorithm;
- a high number of nodes is usually (from our experience) connected with failure or at least sub-optimal results;
- a high number of nodes also means a longer processing time which is usually an issue in real-time implementation.

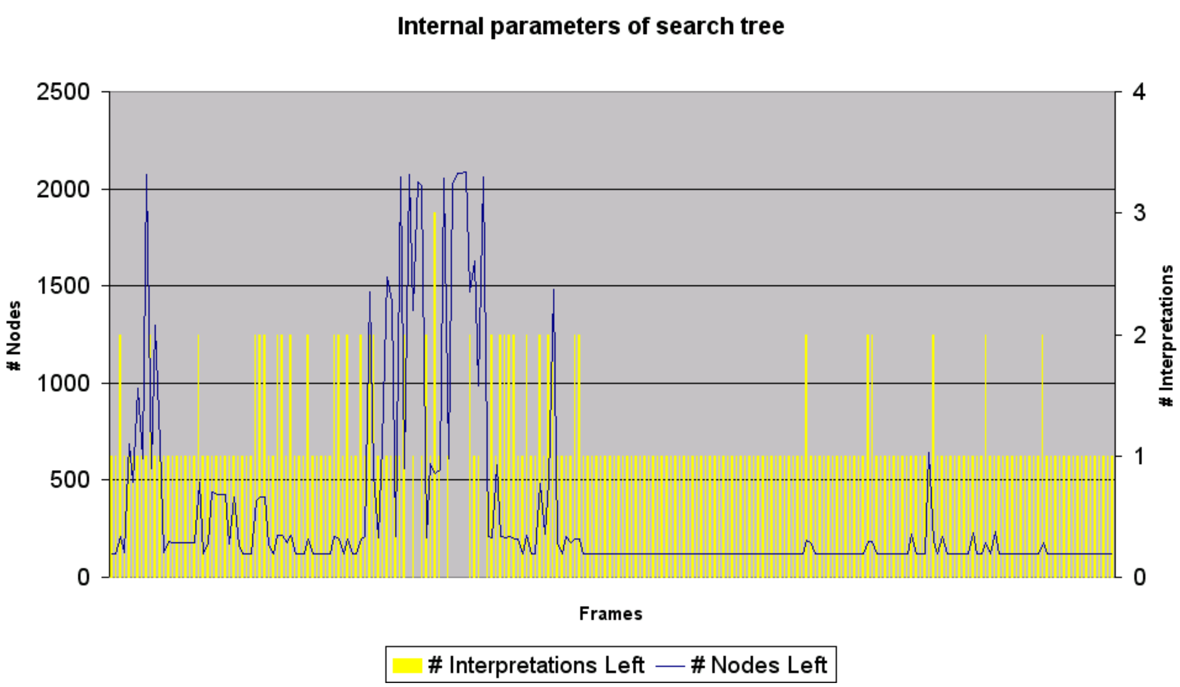
We analyzed the algorithm's performance on the frames that showed no interpretations and/or a high number of nodes and we found out that in these cases most of the problems were due to a calibration issue. Figure 6 shows the performance analysis results for a second run after fixing some the discovered calibration problems. Compared to the results depicted in Figure 5(a), Figure 6(a) shows the correct symbolic result of a bi-directional two-lane road. The graph in Figure 6(b) appears now smoother with just a few problematic frames in the middle of the sequence where a high number of nodes and lack of interpretations point us to areas that need further investigation.

This fairly simple approach to performance analysis can be used to support further development of the algorithm by pointing out video frames that cause problems. An actual performance evaluation beyond mere heuristics, however, requires a more sophisticated approach and is described in the next section.

IV. Performance Evaluation

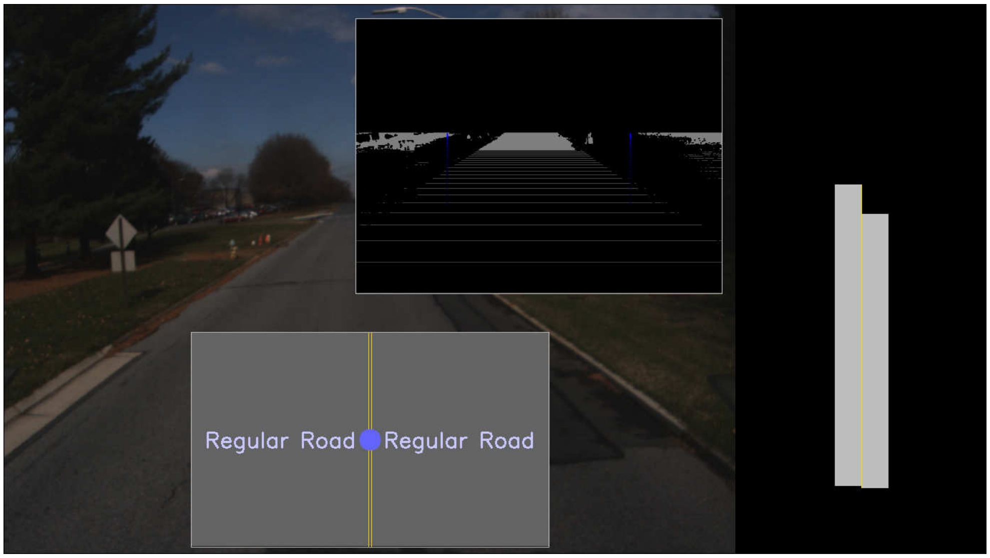
In order to evaluate the performance of an algorithm one needs a reference - the ground truth - to which the algorithm's results can be compared against. Considering our algorithm's results - chains of symbolic nodes - we need a repository of world data from which we can extract comparable structures. We decided to exploit an existing structure - the NIST Road Network Database (RNDB). In the following, we describe

(a)

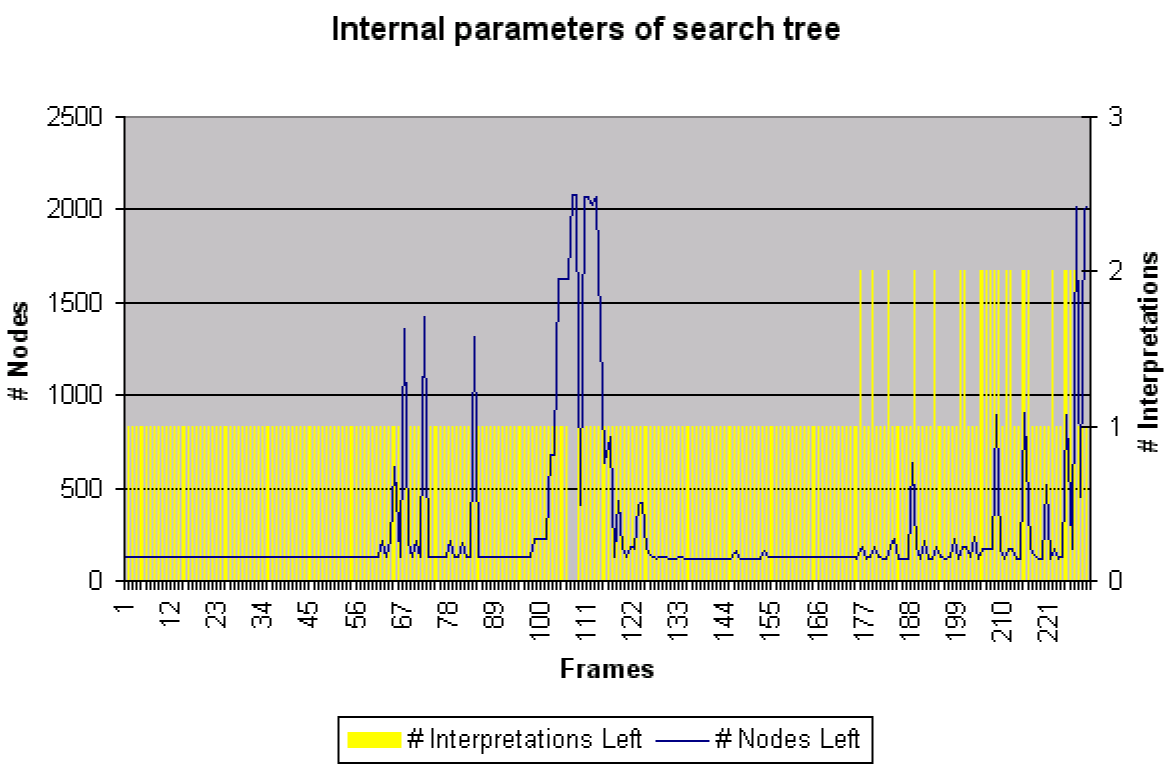


(b)

Fig. 5. (a) First run road recognition result for the first frame of the test sequence. The algorithm erroneously recognized an intersection on the left side of the road. (b) shows the number of nodes (depicted in blue) and the number of interpretations (yellow) for the first run (left road side only).



(a)



(b)

Fig. 6. (a) Second run road recognition result for the first frame of the test sequence. There are no wrongly detected intersections anymore. (b) shows the number of nodes (depicted in blue) and the number of interpretations (yellow) for the second run (left road side only).

briefly the NIST Road Network Database, the extraction of ground truth from this database, and performance evaluation results.

A. *NIST Road Network Database*

In 2004, NIST embarked on an effort to create a Road Network Database (RNDB) structure for the purpose of informing an intelligent vehicle about the structure of the roadway to allow for better path planning and autonomous mobility during on-road driving. This database structure has been represented in a MySQL database [11], documented [9] and populated with detailed instances of roadways on the NIST campus. This section will briefly describe the RNDB and describe how it will be applied to the road recognition approaches described in this paper.

Some of the fundamental components of the Road Network Database are described below:

- *Junctions* - A junction is a generic term referring to two or more paths of transportation that come together or diverge, or a controlled point in a roadway, including lanes splits, forks in the road, merges, and intersections. Junctions are an abstract supertype in the sense that a junction must be one of the types listed above.
- *Intersections* - Intersections are a type of junction in which two or more separate roads come together.
- *Lane Junctions* - A lane junction is a location in a junction in which two or more lanes of traffic overlap.
- *Road* - A road is a stretch of travel lanes in which the name of the travel lanes does not change.
- *Road Segment* - A road segment is a uni-directional stretch of roadway bounded by intersections.
- *Road Element* - A road element is a uni-directional stretch of roadway bounded by any type of junction. Unlike road segments, road elements can be bounded by merging lanes, forks, etc.
- *Lane Cluster* - A lane cluster is a set of uni-directional lanes (with respect to flow of traffic) in which no physical attribute of those lanes change over the span of the lane segment.
- *Lane* - A lane is a single pathway of travel that is bounded by explicit or implicit lane marking. Lanes span the length of a lane cluster of which they are a part.
- *Lane Segment* - A lane segment is the most elemental portion of a road network captured by the database structure. Lane segments can be either straight line or constant curvature arcs. One or more lane segments compose a lane
- *Junction Lane Segments* - A junction lane segment is a constant curvature path through a portion of a lane junction.

For the purpose of road recognition system described in this paper, the two structures that are of most interest are the Road Segment and the Intersection. Figure 7 shows a sample roadway with one of the road segments shaded. There are two intersections shown, represented by black boxes with no lane markings.

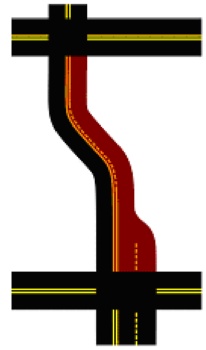


Fig. 7. Sample road network.

The Road Segment database structure contains information such as:

- The road that the road segment is a part of
- The adjacent intersections
- The length of the segment
- The class of road segment (interstate highway, beltway, country road, etc.)

Additional information can also be inferred by looking at other classes that this structure points to, including:

- Beginning and end point of the road segment
- Number of lanes

The database structured has been populated with data from the NIST campus using high-resolution LIDAR scans performed by an external organization. Through post processing, these LIDAR scans were tagged with information about roadways, parking lots, buildings, etc. This information was then converted into the RNDB format and used to populate the database.

Vehicles are localized on this road network using the *Global Positioning System* (GPS) data that is returned from their systems. Although this GPS data is often non-exact, one can still run an algorithm to find the closest road segment to the returned point (this is how off-the-shelf GPS navigation systems work). Since the road segments are defined by their known start and end point, this calculation is relatively trivial.

The Road Segment and Intersection structures in the RNDB correspond nicely to the road and intersection concepts used by the road recognition algorithms. As such, they should provide a nice representation approach for the algorithms.

B. *Ground truth Extraction from RNDB*

After localizing the vehicle's position within the road network, we need to extract the ground truth for the current frame of the video sequence. Figure 8 depicts the approach: From the vehicle's location and orientation a set of symbols describing the road in front of the vehicle is extracted.

Due to the limitations of the sensor, only parts of this symbolic structure are actually within the field of view. Therefore, we need to prune the structure at the maximum look-ahead distance which is known from camera calibration as being 55

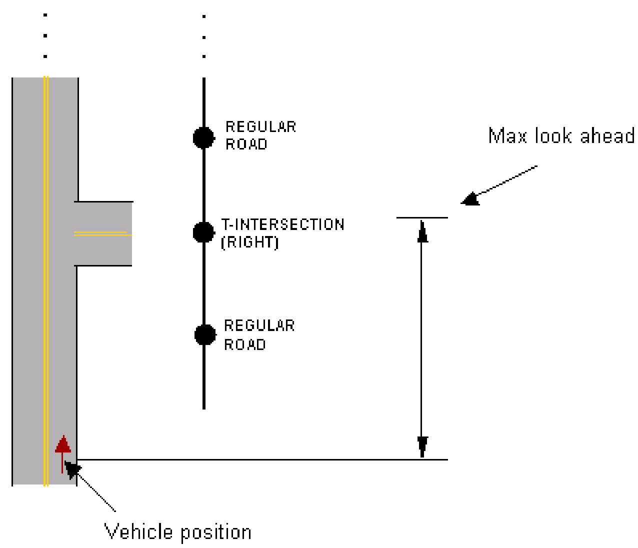


Fig. 8. Simple approach to ground truth extraction.

m. This, however, can only be a coarse estimate of the ground truth because several situations may change the maximum look-ahead distance, e.g.

- whenever the vehicle's orientation is not parallel with the ground, e.g. through tilting due to acceleration, deceleration, or terrain undulations.
- whenever the road's elevation in front of the vehicle differs from the road plane the vehicle sits on.

In the case of the example in Figure 8, the following symbolic road structure could be extracted as ground truth: *(REGULAR ROAD, INTERSECTION)*. The second symbol *(INTERSECTION)*, however, might or might not be part of the actually visible road on sensor data. Such situations require manual correction of the ground truth.

C. Performance Evaluation Results

Figure 9 shows the performance evaluation results for the second run on the video sequence from Section III.

Most of the frames show no classification error at all. The bigger block in the middle of the sequence shows an error of 50 %. This is due to the problem of ground truth generation described in the previous section - the ground truth contains information about an intersection that is actually not yet visible on the sensor data. There are two more peaks in the graph showing a classification error of 100 % (two frames in the middle) and 30 % (two peaks at the end of the sequence). These are good examples for the application of performance evaluation in order to find problematic situations that need further investigation.

In order to compare the performance of the algorithm on video sequences as a whole, we also calculate the minimum, maximum and average classification error throughout the video sequence. This allows for example to compare the performance of different versions (e.g. using different parameters) of the algorithm on the same input data.

The graph in Figure 10 shows the average classification error for the two runs from Section III. The improvement in the second run is reflected by an average error of half the size of the error in the first run.

V. Conclusion

We presented in this paper a two-level approach to performance analysis for a new road recognition approach providing symbolic descriptions of the road structure. The first level of performance analysis helps point out potentially problematic areas and real-time issues by analyzing the behaviour of the tree search based recognition approach. A high number of nodes and lack of interpretations in the resulting search tree are considered as indicators for such problematic areas. On the second level an actual performance evaluation is performed by comparing the symbolic results of the algorithm against a (semi-) automatically extracted ground truth. We pointed out situations where a manual correction of the ground truth is necessary. Both methods of performance analysis proved helpful for the ongoing further development of high-level road recognition for on-road driving. In order to allow comparison of different approaches to road sensing, more efforts are needed to bring together worldwide groups and to agree on common grounds for performance analysis in the future.

References

[1] R. Chapuis, R. Aufrere, and F. Chausse. Accurate Road Following and Reconstruction by Computer Vision. *IEEE Trans. on Intelligent Transportation Systems*, 3, 2002.

[2] D. DeMenthon and L. Davis. Reconstruction of a Road by Local Image Matches and Global 3D Optimization. In *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, 1990.

[3] M. Foedisch, C. Schlenoff, and M. Shneier. Towards an Approach for Knowledge-based Road Detection. In *Proceedings of the ACM Workshop on Research in Knowledge Representation for Autonomous Systems*, pages 1–8, 2005.

[4] M. Foedisch and A. Takeuchi. Adaptive Real-Time Road Detection Using Neural Networks. In *Proc. of the Intl. Conf. on Intelligent Transportation Systems*, 2004.

[5] M. Foedisch and A. Takeuchi. Adaptive Road Detection through Continuous Environment Learning. In *Proc. of the Applied Imagery Pattern Recognition Workshop*, 2004.

[6] W. Grimson. *Object Recognition by Computer: The Role of Geometric Constraints*. The MIT Press, 1990.

[7] T. Hong, A. Takeuchi, M. Foedisch, and M. Shneier. Performance evaluation of road detection and following systems. In *Proc. SPIE Vol. 5609, Mobile Robots XVII; Douglas W. Gage; Ed.*, 2004.

[8] M. Luetzeler and E. Dickmanns. Road Recognition with MarVEye. In *Proc. of the IEEE Intl. Conf. on Intelligent Vehicles*, 1998.

[9] C. Schlenoff, S. Balakirsky, T. Barbera, C. Scrapper, J. Ajot, E. Hui, and M. Paredes. The NIST Road Network Database: Version 1.0. Technical Report Internal Report 7136, NIST, 2004.

[10] C. Tan, T. Hong, M. Foedisch, T. Chang, and M. Shneier. Performance Analysis of a new Road Following Algorithm Based on Color Models. In *Proc. of the SPIE Defense and Security Symp.*, 2006.

[11] M. Widenius and D. Azmark. *MySQL Reference Manager*. O'Reilly & Associates, Incorporated, 2002.

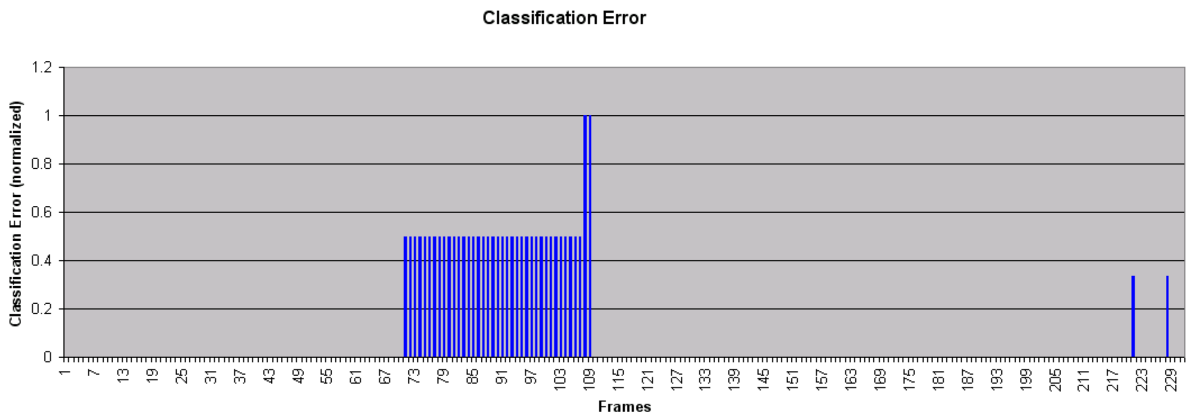


Fig. 9. Performance evaluation results for the second run on the test video sequence.

Fig. 10. Average classification error for the first and second run.