Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Real-Time Classification of Road Type and Condition in Passenger Vehicles

Tim Beilfuss * **Karl-Philipp Kortmann** * **Mark Wielitzka** *
Christian Hansen * **Tobias Ortmaier** *

* *Institute of Mechatronic Systems, Leibniz University Hannover, GER (e-mail: beilfuss.tim@gmail.com, {kortmann, wielitzka, hansen ortmaier}@imes.uni-hannover.de)*

Abstract: Modern vehicles are equipped with numerous sensors and hence offer an increasing degree of environmental perception. In this work, a method is presented that is able to classify different road types and their conditions based on standard vehicle sensors. Therefore, training and validation data on two routes in urban traffic and on federal highways was gathered using a Volkswagen Golf GTE Plug-In Hybrid. The method uses features based on both frequency and time domain extended with a physical vehicle sub-model. For the classification a decision tree model is trained offline and implemented for online use on target hardware commonly used in modern vehicles. A Bayesian and Markov based filter is used to smooth the output and increase the accuracy of the classification.
Since the method is based on sensors that are available in modern vehicles, there is no need for additional hardware, reducing the effort required for implementation. Results show promising classification performance, especially for classifying cobblestone. The three classes of good, medium and bad asphalt labeled relatively precise despite very similar characteristics. Possible applications of the approach could be to adapt vehicles suspension and driving dynamics, to parameterize driver assistance systems, or to update road maps according to their current condition.

Keywords: Real-time systems, classification, machine learning, Markov models, decision trees, vehicle dynamics, inertial measurement units, sensor fusion

1. INTRODUCTION

Knowledge about type and condition of the currently used road surface is important for both vehicle operation and road maintenance. The latter needs information about the condition to effectively maintain the road network. A vehicle can adjust its suspension, driving dynamics and assistance systems according to the road to increase the passengers' comfort and safety.

In the literature, a variety of methods applicable for the identification of various environment conditions have been proposed. Early works have been accomplished by Iagnemma and Dubowsky (2002) and Sadhukhan and Moore (2003) on unmanned vehicles and rovers. They used internal sensors to compute the wheel slip over several terrains with an online algorithm. The authors refrained from using visual and auditory data to be free from environmental conditions such as weather and light. Weiss et al. (2007) used and compared machine learning algorithms. They classified terrain types using an accelerometer mounted on a small vehicle while Ward and Iagnemma (2009) used a suspension mounted accelerometer in a passenger vehicle. Their input features are vibration-based and post processed by a principal component analysis. A support vector machine (SVM) classified terrain types in real time. Another approach is presented by Wang et al. (2011) using a similar setup and adding a camera. A Fast Fourier Transform (FFT) based feature generation was followed by an artificial neural network. Prażnowski and Mamala (2016) suggested mounting a three-axis accelerometer to the windshield. Using lateral as well as vertical vibrations they classified three individual asphalt roads. Using two accelerometer units mounted on the rigid vehicle above the rear wheels, Du et al. (2016) estimated the International Roughness Index (IRI), a measure used for evaluating roads based on the surface profile. The IRI calculation based on accelerometers was shown by Du et al. (2014), while Prasad et al. (2013) showed the relationship between IRI and visible surface distresses. Road roughness was also estimated by González et al. (2008) using accelerometers and a half-car model.A Bayes filter was implemented by Komma et al. (2009) to consider the classification history when classifying terrain.
Besides using accelerometer data, auditory approaches are proposed by multiple authors, with Zhao et al. (2013) estimating IRI using statistical methods and wet road surface detection with an SVM by Alonso et al. (2014). Abdić et al. (2016) applied a long short term memory recurrent neural network (LSTM) to estimate the wetness of the road surface from audio data.

In contrast, this paper focuses on a classification based on sensors that are already integrated in modern passenger vehicles from factory and are mostly unaffected by environmental conditions such as weather and light. Additionally, the classification is proven to run online with the computational resources already available in modern passenger

vehicles. The machine learning algorithm is trained with data collected during regular test drives on public roads. In this way we aim for a robust and reliable method that provides useful environmental information with an easy implementation and transfer to other modern passenger vehicle by adapting the presented training and feature extraction methodology.
The document is organized as follows: The method of data measurement and collection is described in section 2. Section 3 describes the realization of the proposed method, including feature generation, machine learning model with post processing, implementation and validation. Results are discussed in section 4. The final section 5 presents the conclusion and proposals for future work.

2. METHODOLOGY AND DATA COLLECTION

This section specifies how the data used for offline training is accumulated. The proposed methodology is displayed in Fig. 1.

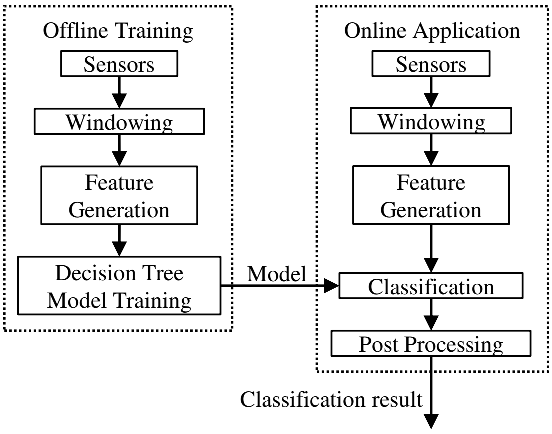


Fig. 1. Overview of the proposed method for offline training and and online application.

During offline training, the sensor data gets windowed into frames of 512 samples each, resulting in a 5 second time frame at 100 Hz sampling rate. Features are generated based on these frames using multiple methods described in subsection 3.1. The decision tree algorithm is trained with the features and labels of the recorded data set. The same windowing and feature generation is applied to the online classification. The previously trained decision tree model is used, and a post processing is applied to improve the performance. While this work focuses on the data of the vehicle integrated sensors, a second data set is created which includes additional sensors. A conclusive comparison will evaluate the potential of additional sensors.
Subsection 2.1 describes the equipment used for acquiring the data necessary for training and validation. Information about the sensors is provided in subsection 2.2 before subsection 2.3 states the method of generating the data set.

2.1 Equipment

To collect data for training and validation in section 3 a Volkswagen Golf VII GTE Plug-In Hybrid is used as test vehicle. Besides using only internal sensors, an additional Inertial Measurement Unit (IMU), a GeneSys ADMA, has been installed to create the second data set (Fig. 3). Because of the IMU's high quality sensors, its measurements outperform standard sensors used in modern vehicles with a higher sample rate and resolution, as shown in Fig. 2. The IMU is mounted to the seat rail of the co-driver seat to maximize the rigidity. According to the rules of rigid body kinematics, all acceleration and rotation values are transformed to the unloaded vehicle's center of gravity (CoG). Data are known of the car's Controller Area Network (CAN) bus and IMU is collected by an ETAS ES910 Prototyping and Interface Module as shown in Fig. 3. This module is also used for the later inference (online classification) testing, as its hardware structure and computing power is similar to a state of the art vehicle's electronic control unit (ECU). A Global Positioning System (GPS) is used for accurate location information of the car, needed for correct labeling of the data as described in subsection 2.3.

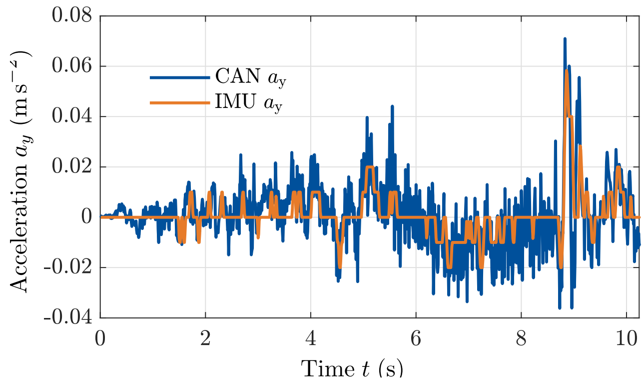


Fig. 2. Comparison of lateral accelerometer signals of CAN and IMU.

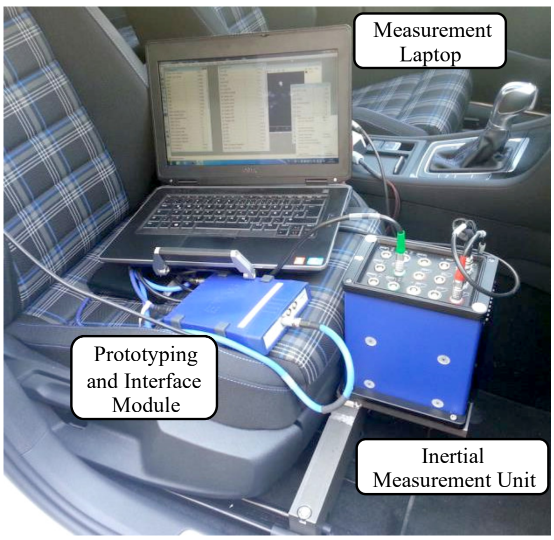


Fig. 3. Setup of the hardware used for data collection and online application.

2.2 Sensors

Two data sets are created. The first (Vehicle-only) includes only vehicle integrated sensors. The second (Vehicle-extended) includes the IMU as well as the vehicle sensors. Including the additional sensors will allow for an estimation of how these sensors might benefit the classification performance. Table 1 shows a compilation of the sensors of the two data sets.

Table 1. Overview of the used sensors for both data sets. ○: Vehicle-only, ●: Vehicle-extended. The wheel speed sensor data is converted into velocity by the vehicle's electronic control unit.

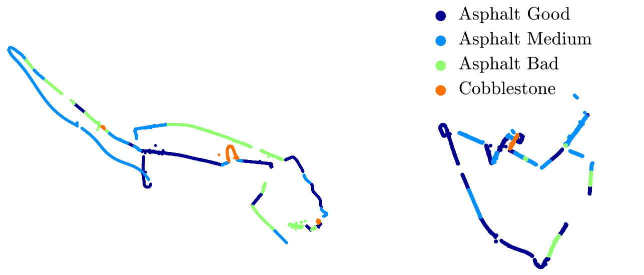
	Data Set	Unit
Long. Acc. CAN	○●	$\mathrm{m\,s^{-1}}$
Lat. Acc. CAN	○●	$\mathrm{m\,s^{-1}}$
Vert. Acc. IMU	●	$\mathrm{m\,s^{-1}}$
Wheel speed sensor CAN	○●	$\mathrm{km\,h^{-1}}$
Roll Sensor IMU	●	$\mathrm{rad\,s^{-1}}$
Pitch Sensor IMU	●	$\mathrm{rad\,s^{-1}}$

While several methods previously presented used a vertical accelerometer, the test vehicle does not include the sensor by default. Since the excitation of each wheel depends on its contacted road surface, increasing road roughness leads to greater differences in the wheel excitations. The resulting impact on spectral density of rolling and pitch rate is advantageously highly correlated to those of the lateral and longitudinal acceleration and therefore also partly reflected in the vehicle's CAN data.
The data was recorded with a sample rate of $f_s = 100\,\mathrm{Hz}$. As the IMU has a sample rate of 1 kHz, the data is down to 100 Hz, to achieve comparable data sets. Additionally their resolution is limited to steps of $0.01\,\mathrm{m\,s^{-2}}$ while the IMU's measurement resolution amounts to $0.001\,\mathrm{m\,s^{-2}}$. The limited sampling rate of 100 Hz may omit the detection of minor differences between road surface textures, as Johnsson and Odelius (2012) found correlations in much higher frequencies.

2.3 Test Method and Labeling of the Data

The car was driven several times on two different routes. The first route is used to collect data sets for the offline training, while the second was used for the validation. The training route has a length of 16.2 km and the validation route is 8.4 km long. Adding up all test drives a total of over 204 km was driven for training and validation purpose (Fig. 4). The labels classifying the condition of the asphalt roads have been chosen by visual inspection of the road based on the working paper *Arbeitsausschuss Systematik der Strassenerhaltung (2005)* used by road maintenance in Germany. The asphalt roads were divided into segments of roughly 150 m length to reduce the effort of the manual labeling. Afterwards they were classified as good, medium and bad road segments depending on how much surface area is damaged and the amount of repairs done to the road segment (Fig. 5):

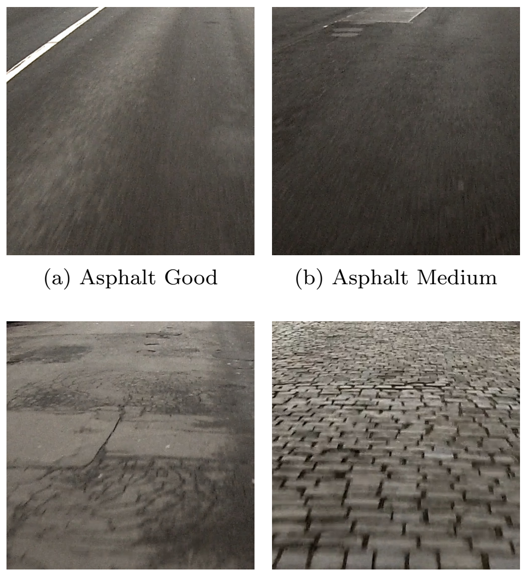
- Asphalt Good: 1 % to 5 % damaged surface
- Asphalt Medium: 6 % to 15 % damaged surface
- Asphalt Bad: 16 % to 50 % damaged surface



(a) Training route (b) Validation route

Fig. 4. Labeled maps of the routes driven. The gaps in the routes originate from obstructed GPS signals and unlabeled sections when no definite label could be chosen.

There were no roads with more than 50 % damaged surface. Fig. 6 shows proportions of the road classes for the training and validation data set. Examples for the classes are displayed in Fig. 5. The car was driven at lower speeds



(a) Asphalt Good (b) Asphalt Medium

(c) Asphalt Bad (d) Cobblestone

Fig. 5. Exemplary photos of the different road classes.

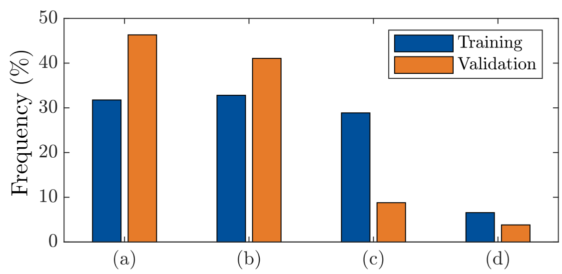


Fig. 6. Label distribution of the training and validation data set in percentage. (a): Asphalt Good; (b): Asphalt Medium; (c): Asphalt Bad; (d): Cobblestone

($30\,\mathrm{km\,h^{-1}}$ to $50\,\mathrm{km\,h^{-1}}$) in urban traffic as well as higher speeds on federal highways ($50\,\mathrm{km\,h^{-1}}$ to $100\,\mathrm{km\,h^{-1}}$). Several driving styles (like economical, normal, and aggressive), different test drivers as well as driving modes

including combustion, hybrid, and electric have been applied during data acquisition. Road condition anomalies like train tracks have been removed from the data set to prevent them from influencing the training.
Due to the method of the visual inspection, the labels reflect the visually perceived condition rather than the actual road roughness. Although this may result in anomalies such as labeling a patchy, but otherwise smooth road beeing as low quality and vice versa the method delivers satisfying labels. Variations within the 150 m road segments are not reflected in the labeling which omits small variations in road condition for the sake of feasability of labeling. Nonetheless the labels provide a comprehensible rating for the road conditions along the two routes.

3. CLASSIFICATION USING VEHICLE SENSORS

This section presents the methods used for offline training and online application. It begins with the feature generation in subsection 3.1. We then explain the decision tree model in subsection 3.2 followed by the post processing described in subsection 3.3. The section is completed by the real time implementation in subsection 3.4.

3.1 Feature Generation

The features are generated using a sliding window of the sensor's data input. The prediction at time t is done using the vector $[x_{t-d}, ..., x_{t+d}]$ with d being the half width of the window. This converts the sequential supervised learning problem into a supervised learning problem, cf. Dietterich (2002). A window size of 512 recorded samples has been chosen for each sensor. The number is selected to maintain a balance between having enough information in the window e. g. for the FFT and limiting the length to only base the features on the most recently traversed road surface. This trade-off is known as the *Kuepfmueller's uncertainty principle*, stating that it is impossible to minimize settling time and bandwidth at the same time.
The 512 sample window chosen for feature generation and model training would result in a new classification every 5 s. We chose to add an overlap of the windows to reduce the time between classifications. As a result, the feature generation provides more training data for the machine learning algorithm and more frequent classification results for the online application. An overlap of 362 samples resulted in the best classification result with unseen validation data. This equals to a new classification every 1.5 s. Increasing the overlap resulted in overfitting the training data while decreasing led to an underfitting. A possible reason for overfitting is the increased amount of classifications while the traveled distance stays the same. The underfitting results from too little training data. To ensure a minimum of excitation and to avoid multiple classification at standstill, a minimum speed of 5 kilometers per hour is specified.
Table 2 summarizes the following features by type and data source.

Table 2. Overview of the generated features. ○: Vehicle-Only; ●: Vehicle-Extended

	Accelero-meter	Wheel Speed Sensor	Rotational Speed Sensor
Frequency Based	○●	-	●
Time Based	-	○●	●
Model Based	●	-	-

Frequency Domain Features A hamming window is applied to the 512 sample window of the lateral and longitudinal acceleration as well as roll and pitch sensors, which are then transformed into frequency domain by a 512-point FFT. Previous works have chosen 128- to 1024-point FFTs, see Weiss et al. (2007) and Sadhukhan (2004). The hamming window improves the FFT's performance by reducing the distorting effect resulting from the signal's edges at the beginning and end of the window (spectral leakage).
The Nyquist theorem limits the theoretically usable bandwidth from 0 Hz to 50 Hz. A feature selection of the optimal bandwidth parameters in spectral density regarding the presented classification task has been performed and accordingly the mean power spectral density between 15 Hz to 21 Hz is selected and calculated as feature for each sensor's signal.

Time Domain Features Due to the connection between road surface condition and vibration amplitude, it is reasonable to extract additional statistical features from the time series.
A variation of the empirical standard deviation σ with sliding mean value has been calculated as feature of the statistical dispersion for the wheel speed and rotational speed sensors (roll and pitch)

$$\sigma_{feat} = \sqrt{\frac{1}{512}\sum_{i=1}^{512}(x_i - \bar{x}_i)^2}\,, \qquad (1)$$

where the signals mean $\bar{x}_i$ is generated using a sliding mean window of 50 samples. This allows the disturbing influence of non-stationary driving characteristics (such as cornering or speed variation) to be taken into account within a time window.

Model Based Features The IMU's vertical accelerometer is filtered with a band-pass filter and then transformed with an inverse quarter car model (Fig. 7) from $\ddot{x}_1$ to $\ddot{x}_2$ to estimate the vertical movement of the wheel. While the parameters m_1 and c_1 has been pre-determined by information of the manufacturer, the other parameters are identified offline via an experimental vibration testing. By this we reduce the vehicle suspension's influence over the measurement, since a suspension or wheel mounted acceleration provides a measurement reflecting the roads deflection more accurate. Due to the road profile being unknown the model must be simplified. The transformation is done using the following equation:

$$\ddot{x}_2 = \frac{1}{m_2} - d_1(\dot{x}_2 - \dot{x}_1) - c_1(x_2 - x_1). \qquad (2)$$

Since that the validity of this method is limited to the quarter car assumption, a stationary state in vibration, the assumption of linearity of the dynamic system, and constant parameters (which vary in reality with e. g. the vehicle's loading condition), the resulting value have to be seen more as additional features than a precise state variable.

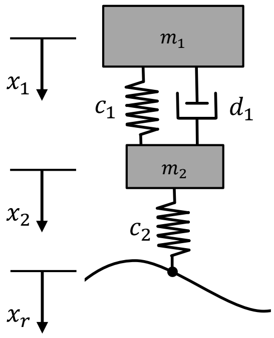


Fig. 7. Quarter car model, x_1: sprung mass position, x_2: unsprung mass position, x_r: road profile, m_1: sprung mass, m_2: unsprung mass, c_1: spring stiffness, c_2: tire stiffness, d_1: suspension damping

3.2 Decision Tree Model

A decision tree model has been chosen for multiple reasons. The method requires little computing costs and benefits from its simple structure, allowing for an easy comprehension of the trained model and reducing the effort required for programming the predictor on the target hardware. Preliminary investigations showed slightly better classification results for more advanced machine learning methods but as this work mainly focuses the applicability to state of the art vehicles and their (ECU) the selection was limited by the available computing power.

Decision trees predict classes based on a series of nodes where the data is split according to trained thresholds. Each node corresponds to a feature the threshold is applied to, with each of the end nodes being one of the classes. The model was trained using the *Gini's diversity index* and an optimized minimum node size with 10-fold cross-validation. A total of 15 features, respectively 46 for the vehicle-extended data set, were made available to the training algorithm. The classification tree algorithm chooses between the features, selecting features to minimize the impurity of the nodes. That way it may choose a single feature for multiple nodes applying different thresholds or not use a feature at all. The trained model also returns probabilities for each classification result.

Fig. 8 shows the splitting process of the decision tree model for the extended data set. The colored bars symbolize the tree splits while the width of the grey stripes is proportional to the amount of data each split cuts. Names of the splits refer to features selected by the training algorithm. Following are descriptions for the exemplary features shown in Fig. 8:

- Lateral 15-21 Hz: Frequency domain feature of the lateral accelerometer between 15 Hz to 21 Hz.
- RL Tire: Time based feature of the real left tire wheel speed sensor.
- Roll: Time based feature of the roll rate.
- Vertical 10-21 Hz: Model based feature of the vertical accelerometer in frequency between 10-21 Hz.

3.3 Post Processing

Following the classification, a Bayesian updating is performed. The Bayesian updating takes the classification score value (obtained by training accuracy or confidence for the specific decision tree's end node) as input. This increases the classification accuracy because previous classifications are taken into consideration. It also smooths the classification result which is useful because the road changes are even for high-speed driving of rather low frequency in comparison to the classification rate 2/3 Hz as explained in subsection 3.4.

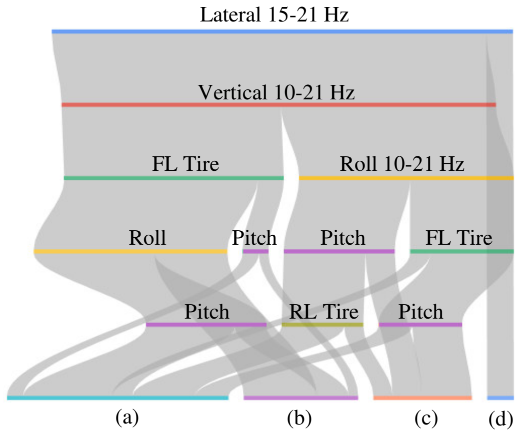


Fig. 8. Classification tree for the Vehicle-Extended data set. (a): Asphalt Medium; (b): Asphalt Good; (c): Asphalt Bad; (d): Cobblestone

The old state probabilities P_{Old} are multiplied by the new probabilities $P_{\text{Prediction}}$ from the decision tree classifier:

$$P_{\text{Bayes}} = \frac{P_{\text{Old}}\, P_{\text{Prediction}}}{\sum P_{\text{Old}}\, P_{\text{Prediction}}} \tag{3}$$

$$P_{\text{New}} = M_{\text{Markov}}\, P_{\text{Bayes}} \tag{4}$$

$$M_{\text{Markov}} = \begin{bmatrix} 0.9657 & 0.0230 & 0.0102 & 0.0011 \\ 0.0168 & 0.9672 & 0.0118 & 0.0042 \\ 0.0159 & 0.0114 & 0.9699 & 0.0028 \\ 0.0188 & 0.0134 & 0.0175 & 0.9503 \end{bmatrix}. \tag{5}$$

After a normalization, the updated probabilities P_{Bayes} are multiplied with a Markov transition matrix. The Markov transition matrix is estimated by the transition of true labels in the training data set. It describes the empirical transition probabilities from one road type to itself or another. The resulting new state P_{New} is saved for the next time step. The transition matrix itself is of first order, only considering the latest state. However, the prior Bayesian updating includes all previous states.

3.4 Real Time Implementation

The method was implemented in MATLAB Simulink, allowing C-code generation for the Prototyping and Interface Module. The calculation time required for each call every 1.5 s on a PC with an Intel Core i7-8700k and 16 GB of RAM clocked at 3200 MHz is 74 ms. Offline training for the entire data set including feature generation requires 179 s. To ensure the online classification when running the algorithm on the Prototyping and Interface Module, it was tested with a Hardware-in-the-loop (HIL) test setup and finally in the actual vehicle without any problems regarding online inference within the demanded classification time of 1.5 s.

4. RESULTS

This section presents the results of the previously described method. The confusion matrices are displayed in Fig. 9 and 10 and include the post processing described in subsection 3.3. The confusion matrices for the training process are displayed in appendix A. The performance is measured with the Matthews correlation coefficient (MCC) for multi-class cases introduced by Gorodkin (2004). It ranges from a minimum value between -1 to 1. A coefficient of 0 equals to a random classification, below 0 being worse than random while 1 resembles a perfect classification result.
The accuracy when differentiating between asphalt and cobblestone is high (82.9 % accuracy for the validation data set after post processing). The classification of cobblestone can be done using only one threshold value. The lateral accelerometer feature suffices to distinct the cobblestone with the mean excitation at 15 Hz to 21 Hz, as shown in Fig. 8. The inaccuracies when labeling cobblestone mostly result from the windowing of the data and the Bayesian smoothing. The results for the detection of cobblestone are similar to comparable works.
Differentiating asphalt road conditions is more challenging. The classifications of the training data set show 56.7 % to 71.9 % accuracy after post processing when labeling the training data set using the vehicle-extended data set. Generally, the computed features of the asphalt roads are very similar to each other. It seems like the measuring frequency might be too low to pick up the differences in texture. Though only a small portion of the roads driven were bad enough to cause a lot of suspension movement. Those road segments were classified with higher accuracy. As expected the average performance of the validation data set is slightly lower. However, the validation results for good asphalt are better than the training results. This and the low validation accuracy when classifying bad asphalt can be explained by the differences in label distribution between the training and validation data sets. The amount of bad asphalt in the validation data set is low (cf. Fig. 6). This makes false positive classifications have a high impact on the accuracy score. The aforementioned inaccuracies in labeling of the data complicate the training and validation. The minor differences and large overlaps in resulting vibrations of the driven asphalt roads complicate the labeling. More steps of the classification tree are needed which increases the error rate further.
Another outcome of the experimental evaluation is, that adding more sensors increases the accuracy. The IMU's higher quality sensors allow for more accurate data features. Especially the vertical accelerometer allows for a wider variety of features, such as the inverted quarter car model. In addition to the improved classification accuracy, the classification confidence of the classification tree model is improved as well, resulting in the increased performance of the Bayesian updating. Vehicle speed did not have a noticeable impact on classification accuracy, since it was inherently included in the trained model.
The post processing improves the MCC by 0.09 for the training data and 0.06 for the validation data and can therefore be considered to be a useful extension. The additional smoothing due to the Bayesian updating increases the method's confidence because the classification changes are less frequent.

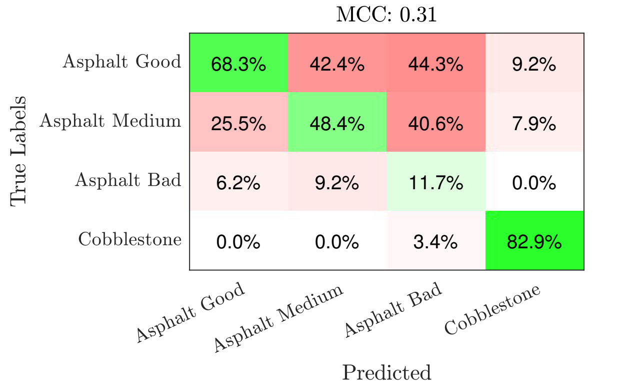


Fig. 9. Confusion Matrix of the Validation, Vehicle-Only.

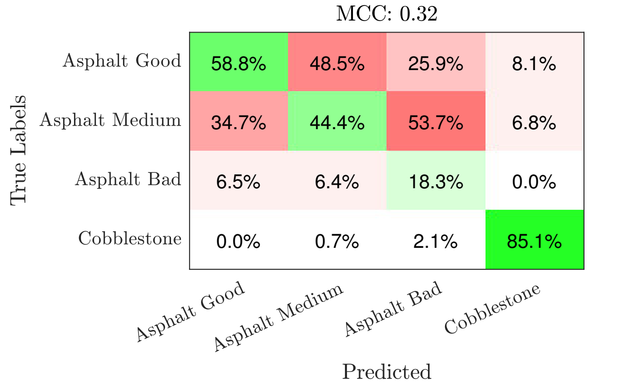


Fig. 10. Confusion Matrix of the Validation, Vehicle-Extended.

5. CONCLUSION

We presented a new approach to classify roads based on default vehicle sensors and computational resources using real world training data. This makes the method more applicable to real world scenarios compared to previous works done on unmanned vehicles or cars with suspension mounted sensors. The training data was gathered without artificial restrictions or any augmentation.
The method's ability to classify with good accuracy without depending on visual or audio data makes it robust. To ensure universal applicability on vehicles, a large variety of vehicles needs to be included in the training data, as well as more comprehensive parameter variations (e.g. on climatic, geographical and loading conditions), which has not been feasible or focused in this work. However, since the used internal sensor equipment is representative for state-of-the-art vehicles, the hardware requirements for a technical transferability of the presented methods to any modern car are given without additional hardware costs. While the IMU's additional and more precise sensors improved the accuracy of the training data set noticeably, the difference of the validation data set was minor. This is partly because the vehicles lateral and longitudinal accelerometers are able to pick up a lot of acceleration that is similar to the vertical acceleration.
Another labeling method, based on auditory or visual

sensor data, which are strongly sensitive to environmental conditions, might improve the method's performance. A higher measuring frequency might improve the method's ability to detect smaller surface irregularities. More road classes and types, such as pavement and dirt, could also be included when more data is available. A neural network approach seems also promising, especially with more powerful hardware finding its way into newer cars. For future applications it is conceivable that the vehicle can send the information to a central server to create a road map based on the results of the classification.

REFERENCES

Abdić, I., Fridman, L., Marchi, E., Brown, D.E., Angell, W., Reimer, B., and Schuller, B. (2016). Detecting road surface wetness from audio: A deep learning approach.

Alonso, J., López, J.M., Pavón, I., Recuero, M., Asensio, C., Arcas, G., and Bravo, A. (2014). On-board wet road surface identification using tyre/road noise and support vector machines. *Applied Acoustics*, 76, 407–415. doi: 10.1016/j.apacoust.2013.09.011.

Arbeitsausschuss Systematik der Strassenerhaltung (2005). Schadenskatalog fuer die messtechnische und visuelle zustandserfassung.

Dietterich, T. (2002). Machine learning for sequential data: A review. *Structural, Syntactic, and Statistical Pattern Recognition.*

Du, Y., Liu, C., Wu, D., and Jiang, S. (2014). Measurement of international roughness index by using z -axis accelerometers and gps. *Mathematical Problems in Engineering*, 1–10.

Du, Y., Liu, C., Wu, D., and Li, S. (2016). Application of vehicle mounted accelerometers to measure pavement roughness. *International Journal of Distributed Sensor Networks*, 12(6), 8413146. doi:10.1155/2016/8413146.

González, A., O'brien, E.J., Li, Y.Y., and Cashell, K. (2008). The use of vehicle acceleration measurements to estimate road roughness. *Vehicle System Dynamics*, 46(6), 483–499. doi:10.1080/00423110701485050.

Gorodkin, J. (2004). Comparing two k-category assignments by a k-category correlation coefficient. *Computational Biology and Chemistry*, 28(5), 367 – 374. doi: https://doi.org/10.1016/j.compbiolchem.2004.09.006.

Iagnemma, K. and Dubowsky, S. (2002). Terrain estimation for high-speed rough-terrain autonomous vehicle navigation. *Proc. SPIE 4715, Unmanned Ground Vehicle Technology IV*, 4715, 256–266. doi: 10.1117/12.474457.

Johnsson, R. and Odelius, J. (2012). Methods for road texture estimation using vehicle measurements. *Proc. of the 25th Int. Conf on Noise*, 1573–1582.

Komma, P., Weiss, C., and Zell, A. (2009). Adaptive bayesian filtering for vibration-based terrain classification. *2009 IEEE International Conference on Robotics and Automation.*

Prasad, J.R., Kanuganti, S., Bhanegaonkar, P.N., Sarkar, A.K., and Arkatkar, S. (2013). Development of relationship between roughness (iri) and visible surface distresses: A study on pmgsy roads. *Procedia - Social and Behavioral Sciences*, 104, 322–331. doi: 10.1016/j.sbspro.2013.11.125.

Prażnowski, K. and Mamala, J. (2016). Classification of the road surface condition on the basis of vibrations of the sprung mass in a passenger car. *IOP Conference Series: Materials Science and Engineering*, 148, 012022. doi:10.1088/1757-899X/148/1/012022.

Sadhukhan, D. (2004). *Autonomous ground vehicle terrain classification using internal sensors.* Master's thesis, Florida State University, Tallahassee, Florida.

Sadhukhan, D. and Moore, C.A. (2003). On-line terrain estimation using internal sensors. *Proceedings of Florida Conference on Recent Advances in Robotics.*

Wang, S., Kodagoda, S., Wang, Z., and Dissanayake, G. (2011). Multiple sensor based terrain classification. In *Australasian Conference on Robotics and Automation.* ACRA.

Ward, C.C. and Iagnemma, K. (2009). Speed-independent vibration-based terrain classification for passenger vehicles. *Vehicle System Dynamics*, 47(9), 1095–1113. doi: 10.1080/00423110802450193.

Weiss, C., Fechner, N., Stark, M., and Zell, A. (2007). Comparison of different approaches to vibration-based terrain classification. In *European Conference on Mobile Robots.* ECMR.

Zhao, Y., McDaniel, J.G., and Wang, M.L. (2013). Iri estimation using probabilistic analysis of acoustic measurements. *Materials Performance and Characterization*, 2(1), 339–359. doi:10.1520/MPC20130018.

Appendix A. CONFUSION MATRICES OF THE TRAINING

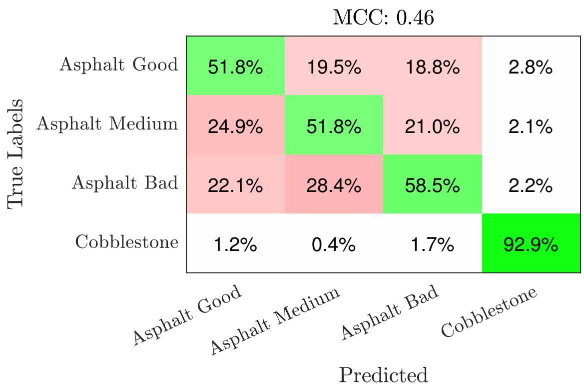


Fig. A.1. Confusion Matrix of the Training, Vehicle-Only.

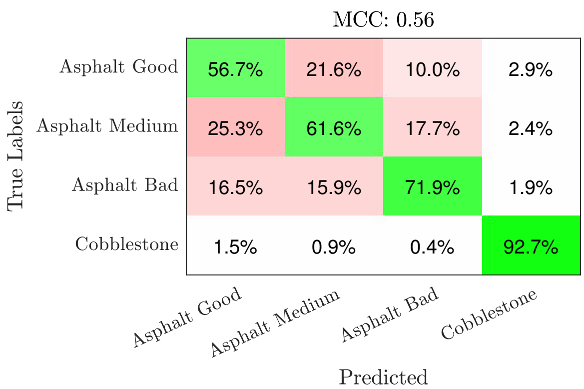


Fig. A.2. Confusion Matrix of the Training, Vehicle-Extended.