Preprint of the paper to appear in Proc. IEEE Intelligent Vehicle Symposium, 2021.

Self-Supervised Learning of Camera-based Drivable Surface Roughness

Jan Cech, Tomas Hanis, Adam Konopisky, Tomas Rurtle, Jan Svancar, Tomas Twardzik
Faculty of Electrical Engineering, Czech Technical University in Prague

***Abstract*— A self-supervised method to train a visual predictor of drivable surface roughness in front of a vehicle is proposed. A convolutional neural network taking a single camera image is trained on a dataset labeled automatically by a cross-modal supervision. The dataset is collected by driving a vehicle on various surfaces, while synchronously recording images and accelerometer data. The surface images are labeled by the local roughness measured using the accelerometer signal aligned in time. Our experiments show that the proposed training scheme results in accurate visual predictor. The correlation coefficient between the visually predicted roughness and the true roughness (measured by the accelerometer) is 0.9 on our independent test set of about 1000 images. The proposed method clearly outperforms a baseline method which has the correlation of 0.3 only. The baseline is based on surface texture strength without any training. Moreover, we show a coarse map of local surface roughness, which is implemented by scanning an input image with the trained convolutional network. The proposed method provides automatic and objective road condition assessment, enabling a cheap and reliable alternative to manual data annotation, which is infeasible in a large scale.**

I. Introduction

During any maneuver, vehicle motion is mainly governed by traction forces generated in the wheel to road interface and thus highly depends on road condition. The road surface properties vary significantly and need to be taken into account while driving. Due to its material (tarmac, cobblestones, gravel, forest roads), the roughness of a drivable surface, including various irregularities or anomalies as potholes or bumps, influences significantly traction properties, a braking distance, or vehicle stability in extreme cases. For instance, a vehicle may become uncontrollable or even roll over if it travels at too high speed on a very rough surface. Besides that, tire/vehicle damage is threatening if rough surfaces are not taken into account both on the road and off the road. Skilled human drivers assess the road surface ahead of the car and adjust driving style accordingly or possibly execute an evasive maneuver. Many driver assistance systems and traction control systems have been developed and introduced to support human drivers. However, most traction control systems adapt to road conditions in a reactive way, based on traction system response for given vehicle motion.

In summary, a predictive visual recognition of the surface roughness ahead of the vehicle has potentially a significant impact on safety, maintenance costs, and overall comfort. Recently, methods of image recognition employing deep learning techniques have become very accurate. However, these methods require a large amount of *labeled* data that need to be fed to a convolutional neural network when training. It is even more complicated for surface roughness prediction, since it is unclear how to label the surfaces manually. The surface roughness is rather a continuous quantity. Therefore, labeling would be difficult for a human annotator. Moreover, drivable surfaces have enormous diversity and to label manually representative samples is clearly infeasible.

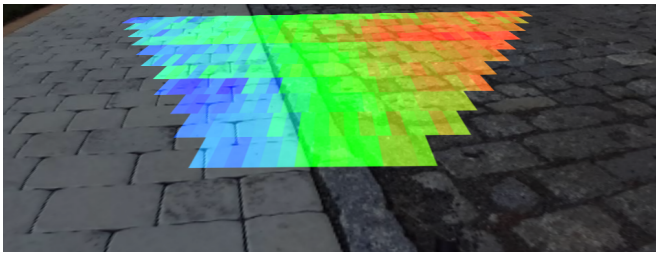

Fig. 1. Color-coded roughness of the surface found by our visual predictor that was trained only by self-supervision from accelerometer data, *without any manual annotation*. Colder colors encodes smoother surfaces (interlocking pavement), while warmer colors rougher surfaces (old cobblestone paving).

Therefore, we propose a self-supervised method for learning a visual predictor of the surface roughness. A large amount of data is collected and labeled automatically using cross-modal data association. The idea is that unlabeled images captured by a camera together with accelerometer signal are acquired while freely driving on various surfaces. The surface roughness is derived from the vertical acceleration of the wheel and this quantity is used to label corresponding images. This way, we collected an extensive labeled dataset and learned an accurate visual surface roughness predictor without any expensive and unreliable manual annotation. An example of a predicted surface roughness map is shown in Fig. 1.

The rest of the paper is structured as follows. Related work is summarized in Sec. II. The proposed method is detailed in Sec. III. Experiments are given in Sec. IV. Finally, Sec. V concludes the paper.

II. Related Work

The roughness of the drivable surface is measured by various techniques and using several modalities. A comprehensive review of the methods is found in e.g., [1]. The taxonomy roughly divides into two groups: (1) effect-based methods, that measure vibrations typically using accelerometers [2], [3], and (2) direct methods, that measure shape of the surface using Laser-scanners [4], [5], Lidars [6], depth cameras [7], [8], stereo cameras [9], or ultrasonic sensors [10], [11].

Recognizing surface roughness from accelerometers was used in the famous Stanley vehicle [12], which won the DARPA Grand Challenge in 2005. The maximum speed of the vehicle was adjusted based on the band-passed filtered signal of the vertical acceleration. Surface properties were estimated similarly even for extra-terrestrial vehicles [13]. An Inertial Measurement Unit (IMU) of a consumer smartphone has become a popular sensor for surface roughness estimation or potholes and road anomalies detection [14], [15], [16]. Recently, deep neural networks were harnessed to recognize road surface type from accelerometer data [17].

Concerning image-based approaches, using a single monocular camera, published methods either detect road damages and anomalies [18], [19], or classify a surface type, rather than predict a continuous quantity related to the roughness of the surface. Typically, they recognize paved/unpaved roads [20], [21], [22], [23] or a small number of road quality [24] or friction [25], [26] levels. A major drawback of all these methods is that they are trained using manually labeled images. The training datasets are small and potentially suffer from inaccuracy of the manual annotation. On the other hand, our proposed method relies on a dataset labeled automatically and objectively by the cross-modal supervision.

The cross-modal supervision principle is well known in machine learning. The general idea is that two modalities are correlated and one of them is used as a supervisory signal for learning a classifier/regressor exploiting the other modality. The principle has been applied in other domains, e.g., Medical imaging [27], Audio-Video analysis [28], [29], in training a monocular image depth estimator [30] or a pseudo-Lidar [31] using supervision by a depth map.

Note that there are methods in the literature that combine both accelerometer data and images [32], [33]. Nevertheless, the methods do not use the acceleration signal to train the vision classifier. The combination of modalities is used to disambiguate and improve the accuracy at the test time.

A similar idea to ours exists in robotics for learning a traversability through the terrain [34]. The problem is different and the methods work with different modalities, e.g., learning a long range terrain classifier from short range sensors [35], or learning a monocular obstacle avoidance from a stereo supervision [36].

For the sake of completeness, we mention recent commercial solutions related to the surface roughness. Mercedes-Benz developed a predictive active suspension system, called 'Magic body control' [37]. The system relies on a forward-looking stereo camera to monitor the road surface. Disturbances, as speed retarders, are communicated to the suspension controller to adjust a real-time response to mitigate effects on the vehicle body. A similar system has been recently introduced by Audi [38]. Cooperative techniques to share information on road anomalies within a fleet have been announced by Landrover [39] or by Mitsubishi [40]. Nevertheless, these proprietary systems are not open, and it is unclear which training strategies were used.

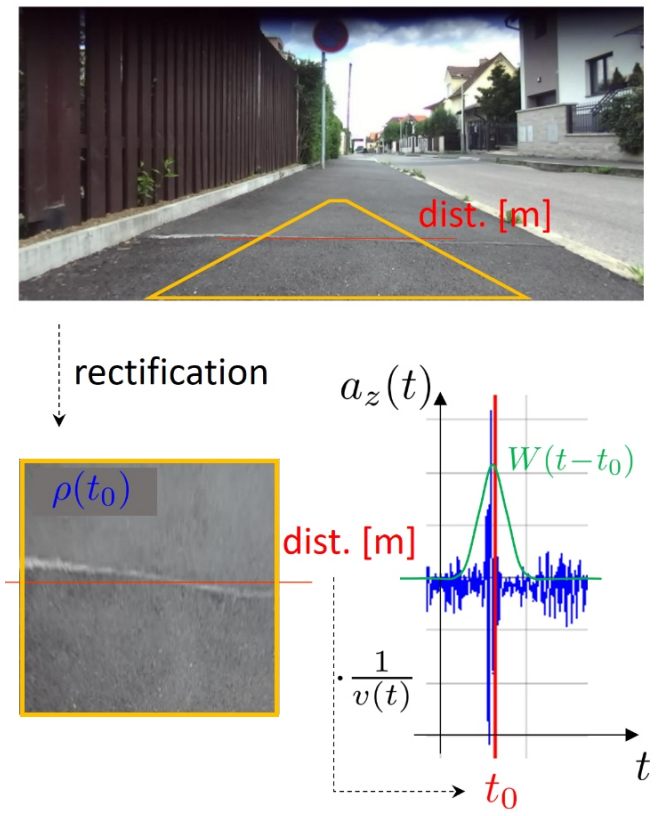

Fig. 2. Overview of the self-supervision process. An input image from the vehicle camera is first rectified and cropped, so the rectified image were capturing a rectangular region of given dimensions in front of the vehicle (in yellow). Red horizontal line delineates the measurement distance (in meters), where the surface roughness is computed. The distance corresponds to measurement time t_0, calculated by vehicle velocity $v(t)$. The surface roughness, a label of the rectified image, is computed by aggregating vertical acceleration $a_z(t)$ over time using Eq. (1).

III. Proposed Method

The goal of our method is to train a visual-only predictor, using a convolutional neural network (CNN), that takes an image captured by a front-looking camera as an input and provides an estimate of the corresponding surface roughness. To train the CNN, a dataset is collected by free driving while recording the camera and accelerometer data. The training images are labeled automatically by associating with the surface roughness computed from the accelerometer signal.

A. Surface roughness

The surface roughness is derived from vibrations recorded by an accelerometer mounted on the front axle of our test vehicle. In particular, the *surface roughness* ρ at measurement time t_0 is defined as

$$\rho(t_0) = \frac{k}{v(t_0)} \sum_t W(t-t_0) a_z(t)^2, \qquad (1)$$

where $a_z(t)$ is vertical acceleration (without the gravity) over time t, $W(t-t_0)$ is a Gaussian-shaped window to aggregate the signal centered at t_0, $v(t_0)$ is the vehicle velocity at time t_0, k is a normalization constant to ensure the surface roughness stays in a reasonable range from 0 to about 1.

Normalization by vehicle velocity $v(t_0)$ is important, since the magnitude of the acceleration signal clearly depends on the vehicle velocity. The dependence is reported to be linear [12], so the effect cancels when divided by the velocity. Quantity ρ in Eq. (1) becomes a local property of the traveled surface itself. Note that the width of the aggregation window should depend on the velocity as well. Nevertheless, we neglected this effect due to the window shape and a stabilized (constant) velocity maintained in the experiments.

The extent of the aggregation window (Gaussian $\sigma = 0.7s$, for velocity 2.5 m/s) was found empirically as a trade-off between a level of filtration and a localization accuracy. Too small window does not filter well high-frequency noise, while a too large window causes artifacts at the surface boundaries.

B. Image labeling

The process is sketched in Fig. 2. A raw image captured by a front-looking camera is first orthographically rectified to a bird's-eye view. The homography mapping is used to warp the raw image of a scene rectangle of size 1.5m $\times$ 1.5m on the ground plane in front of a vehicle into a rectified image, such that the corners of the scene rectangle and the corners of the rectified image correspond. The rectified image is then used as an input to a convolutional neural network.

To label the image captured at time t_0, the corresponding surface roughness $\rho(t_0)$ is found. We define a fixed measurement distance d, set as $d = 0.75$m in all our experiments. Then time t_0 when the vehicle travels on the surface over the measurement distance d, is calculated using the vehicle velocity as $t_0 = d/v(t)$.

The size of the input image, which is used by the CNN to predict the surface roughness, was found experimentally. Too small region around the measurement distance would not provide enough context, while too large region would capture an irrelevant background.

In principle, image rectification is unnecessary. Nevertheless, this is an easy way to normalize images and possibly combine different cameras with different viewpoints without retraining. In all our experiments, the homography is estimated offline. Any deviations from the estimated mapping, that may occur due to camera tilting because of mounting on the vehicle body, are neglected. The resulting inaccuracy is filtered in Eq. (1). Online estimation of the camera tilt (from, e.g., IMU or stereo) would probably increase the spatial sensitivity of the method.

C. Convolutional neural network

Using the process above, we collected a labeled dataset $\{(I_1, \rho_1), \ldots, (I_n, \rho_n)\}$. RESNET-50 [41] was used as the backbone, with the input of the rectified RGB image I resized to 224×224 px receptive field. The network has a single scalar output after ReLU, the estimated surface roughness $\text{CNN}(I; \Theta) = \hat{\rho}$. Vector Θ concatenates all network weights.

L_2-regression loss $L(\Theta) = \sum(\rho_i - \hat{\rho}_i)^2$ was used as the loss function penalizing differences between the ground-truth and predicted labels of the surface roughness. Batch-normalization layers [42] were inserted and network weights Θ were found by ADAM optimizer [43]. As data augmentation, we used color jitter (contrast, brightness, hue), and random horizontal flipping. Only small $\pm4^\circ$ image rotations were made, since direction of the image texture may be important for the perceived surface roughness. The training converged after about 50 epochs.

Fig. 3. Sub-scale vehicle platform used in our experiments.

Fig. 4. Accelerometer mounting location.

IV. Experiments

A. Experimental sub-scale vehicle platform

The data collection and real-world experiments were performed using a sub-scale experimental platform based on a commercial RC car, the Losi®1:5 DBLX-E. The platform is showcased in Fig. 3.

Necessary mechanical modifications were done to prepare the platform for intended experiments. The center differential was blocked, and the front drive train was dismounted. Thus, the front wheels are not driven and could be used for vehicle speed measurement.

The platform is equipped with multiple computational units that collectively control the vehicle, read and record all sensor data, and ensure the safe operation of the vehicle. The high level ECU is Intel *NUC7i7BNK* mini PC, allowing rapid prototyping and high level data collection using *Matlab & Simulink* and Python environment. The mini PC solution enables direct visual data collection from StereoLabs ZED stereo camera. Note that in the experiments, we use a single camera of the rig.

The low-level computational units are Raspberry Pi 4, Arduino Nano, and STM32 Nucleo microcontroller (STM32L432). The Raspberry Pi controller augmented with Navio2 Autopilot HAT unit, is the vehicle main hardware-level controller. The measurement of vehicle motion is based on a dual 9-degree-of-freedom inertial measurement unit (namely, MPU9250 and LSM9DS1) combined with U-blox M8N Global Navigation Satellite System (GNSS) module. The Navio2 unit has a nine channel-independent PWM generator used to command the steering servos and the e-motor control unit. The Adruino and STM32 microcontrollers were added to collect wheel-level data. The road roughness is measured using Analog Devices ADXL326 3-axis accelerometer mounted at the front left wheel axle. The location of the sensor is presented in Fig. 4. Each wheel is equipped with an RPM sensor. The incremental magnetic sensor consists of static ALLEGRO A1325 low noise, linear Hall effect sensor with analog output, and 3D printed discs with uniformly distributed ten permanent magnets attached to the wheel shaft. Finally, the safety of operation is guaranteed by build-in emergency stop functionality and system redundancy.

B. Dataset

The dataset was collected by driving the vehicle on various surfaces. We acquired about 5 hours of synchronous raw recordings of all vehicle sensors and control signals, including the camera and accelerometer data. Test rides were made in about 5 different locations, each containing a collection of different surfaces. Recordings were repeated at different days and day times to capture the effect of illumination, so we have images for sunny/cloudy days at noon or at dusk. A simple automatic cruise control was set for test rides to maintain a more or less constant speed of 2.5 m/s. The labeling process described in Sec. III assumes the vehicle goes approximately straight, so the data of significant deviations from the straight direction were discarded in the postprocessing.

We took images every 0.5s to keep the dataset of a reasonable size with enough diversity. The resulting dataset, automatically labeled with the accelerometer-based surface roughness, includes about 10k samples. The dataset was split into disjoint subsets: the training subset (70%), the validation subset (20%) to select the best training epoch, and the test subset (10%) used for evaluation. We made sure the test set does not overlap with the training data by using location and temporal metadata.

A sample of our dataset is shown in Fig. 5. Fig. 6 presents a distribution of the ground-truth labels of surface roughness.

	MAE	RMSE	Corr	P_{95}	$e_{0.1}$
baseline	0.1415	0.1920	0.3302	0.4282	49.2%
CNN	0.0522	0.0720	0.9013	0.1494	86.9%

TABLE I
ERROR STATISTICS

C. Evaluation and results

Before we evaluate the proposed method, we introduce a simple baseline to compare.

Baseline method. The baseline is based on the idea that the roughness of the surface is related to its texture. A very smooth surface does not usually have a strong texture. And vice versa, rough surfaces have to manifest by some texture, due to normal direction changes. The texture strength was measured by using image gradients. In particular, the input image (the same as for the CNN method) was converted to a gray scale, and the average magnitude of pixel-wise image gradients was calculated. The range is finally adjusted by multiplication with a constant, such that the standard deviations of the ground-truth surface roughness and the average image gradient magnitudes calculated over the training set were equal. Besides the range fit, there is no learning involved in the baseline method.

All evaluation statistics, computed on the independent test split of our dataset, are summarized in Tab. I. The statistics are Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Correlation coefficient (Corr)[1], Percentile-95 (P_{95}) meaning the prediction error lower or equal for 95% of test samples, and Error-0.1 ($e_{0.1}$) showing the percentage of samples having prediction error lower than or equal to 0.1. It is clearly seen that the proposed CNN method outperforms the baseline by a large margin in all evaluation statistics.

Besides the statistics, we show the scatter plots in Fig. 7 visualizing the correlation between the ground-truth and the predicted surface roughness labels. In Fig. 8, we show the cumulative histograms of absolute errors to provide an insight on the error distributions. Statistics P_{95} and $e_{0.1}$ are easily seen in the plots.

To provide further intuition, in Fig. 9, we present several images with a large difference between the two methods, the CNN and the baseline. It is seen that the proposed CNN is not disturbed by shadows or weak illumination. On the other hand, the naive baseline unsurprisingly overestimates the grass and the textured but smooth tarmac, and at the same time, underestimates the cobblestone paving.

D. Roughness maps – qualitative results

The following experiment demonstrates that the trained model generalizes to a slightly different problem, predicting a coarse map of the local surface roughness.

Our CNN-model was trained to predict the local surface roughness in the center of the given image, i.e., a single

[1]Note that the range fitting in the baseline method has no effect on the correlation coefficient.

Fig. 5. Samples from our dataset sorted from the lowest to highest surface roughness. The labels are depicted.

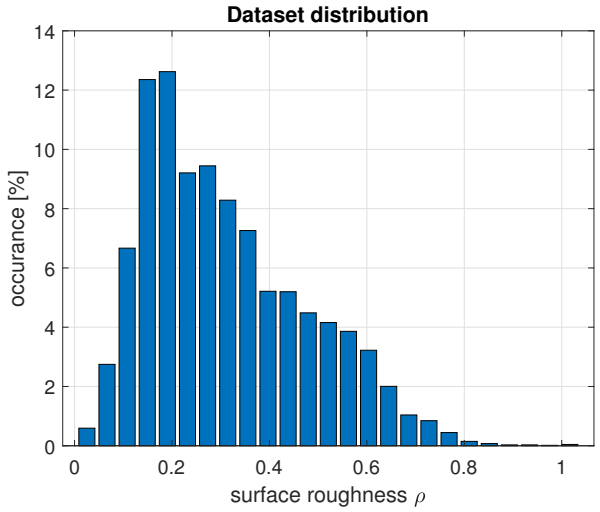

Fig. 6. Surface roughness distribution in our dataset

scalar for an image. The input to the CNN represents a square of $1.5\text{m}\times1.5\text{m}$ in the scene. Therefore, we scan a larger area of the surface seen by the camera with the CNN evaluated at many locations. In particular, the scanning is done in the rectified bird's-eye view, and partially overlapping images of $1.5\text{m} \times 1.5\text{m}$ windows are one by one fed to the CNN. The outputs are stored, the center window locations are colorized based on the surface roughness, and finally back-projected to the raw camera image.

Results are shown in Fig. 1 and Fig. 10 for several raw images of our test set. We can see a smoother interlocking pavement in colder colors of lower roughness, while old cobblestone paving of higher roughness appears in warmer colors. Similarly, a bumpy road with pebbles correctly appears rougher than a smooth grass.

This experiment is presented as a qualitative result, since we do not have the ground-truth data for other locations than the measurement point (0.75m in front of the vehicle for an image). The CNN was not trained for the other

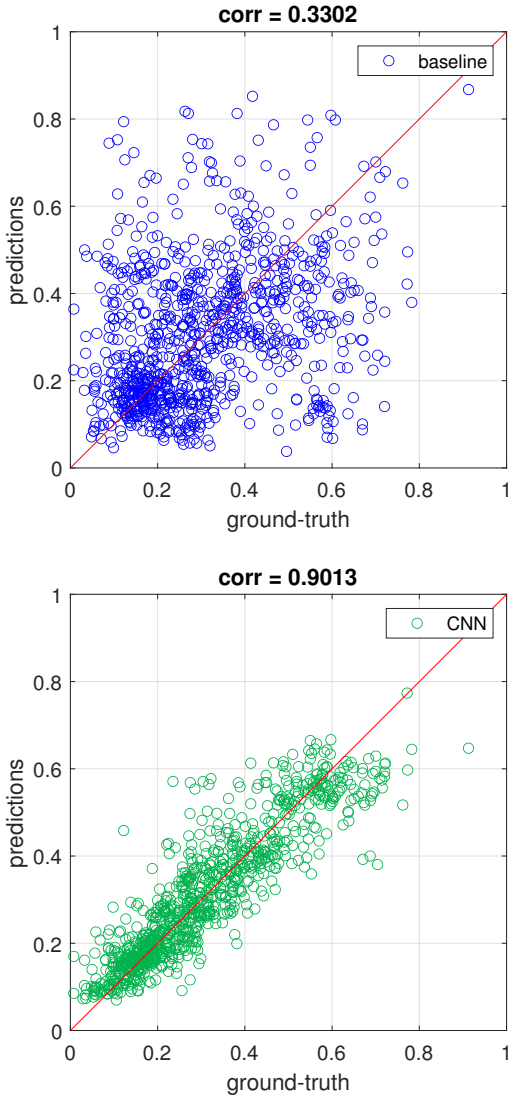

Fig. 7. Scatter plots for the baseline and the proposed CNN-based method. Ideal predictions would lie on the red diagonal line. Correlation coefficients are shown in titles of the plots.

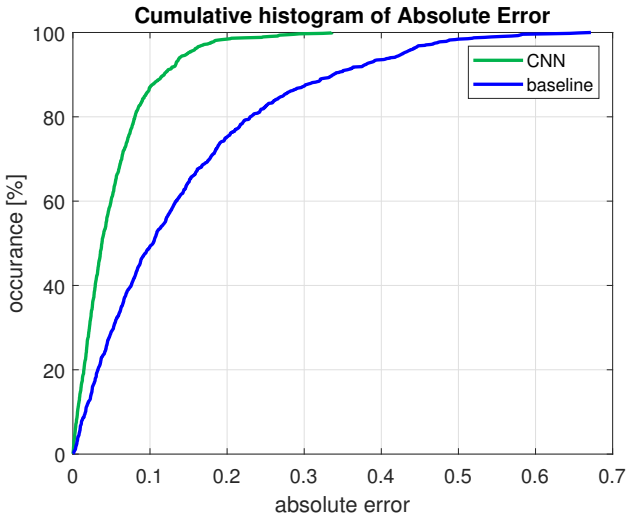

Fig. 8. Cumulative histograms of absolute error.

locations, and it is rather surprising that it generalizes well for far distant locations (up to 5 meters), where a loss of resolution degrades the image quality. From this perspective, particularly challenging is the last image in Fig. 10, where the rougher surface is in the distance.

The performed experiment demonstrates the accuracy and spatial consistency of the predictor. Computational aspects were not considered here. Of course, different architectures are used for semantic map prediction problems, e.g., [44], [45]. To train the network to predict the local surface roughness map, we could simply use the labels generated as in this experiment or treat the problem as a semi-supervised learning, where only certain regions have the labels assigned.

V. Conclusions

We presented a method, where a visual predictor of surface roughness was trained by cross-modal supervision without any manually annotated data. The surface roughness measured by a wheel axle mounted accelerometer was used to label images captured by the camera. An automatically labeled dataset of about 10k images was collected, and a convolutional neural network was trained. The method was evaluated on the independent test split of the dataset, achieving accurate predictions and clearly outperforming the baseline method. As a qualitative result, we showed coarse roughness maps by scanning the input image with the trained CNN predictor.

A limitation of the presented method is that all the training is done offline. It means a large dataset is collected first, and the predictor is trained and then kept fixed. In the future, we will investigate methods that would allow online learning, i.e., the predictor would learn novel surfaces immediately when they are first encountered. Detection of the surfaces to be learned should be straightforward. The surfaces are those where the visual prediction of roughness significantly differs from the roughness measured by the accelerometer.

Acknowledgement

The research was supported by Toyota Motor Europe, and by CTU student grant under project SGS20/171/OHK3/3T/13.

References

[1] A. Yunusov, S. Eshkabilov, D. Riskaliev, and N. Abdukarimov, "Estimation and evaluation of road roughness via different tools and methods," in *Proc. Transport Problems*, 2019.

[2] H. B. Salau, A. J. Onumanyi, A. M. Aibinu, E. Onwuka, J. Dukiya, and H. Ohize, "A survey of accelerometer-based techniques for road anomalies detection and characterization," *International Journal of Engineering Science and Application*, vol. 3, no. 1, 2019.

[3] A. Gonzalez, E. J. O'brien, Y.-Y. Li, and K. Cashell, "The use of vehicle acceleration measurements to estimate road roughness," *Vehicle System Dynamics*, vol. 46, no. 6, 2008.

[4] Romdas System, "Laser profilometer," 2021, https://romdas.com/romdas-laser-profiler.html.

[5] G. Bitelli, A. Simone, F. Girardi, and C. Lantieri, "Laser scanning on road pavements: A new approach for characterizing surface texture," *Sensors*, vol. 12, no. 7, pp. 9110–9128, 2012.

[6] M. Yadav, B. Lohani, and A. K. Singh, "Road surface detection from mobile lidar data," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 4, no. 5, pp. 95–101, 2018.

[7] A. Mahmoudzadeh, A. Golroo, M. R. Jahanshahi, and S. F. Yeganeh, "Estimating pavement roughness by fusing color and depth data obtained from an inexpensive rgb-d sensor," *Sensors*, vol. 19, no. 7, 2019.

[8] F. Marinello, A. R. Proto, G. Zimbalatti, A. Pezzuolo, R. Cavalli, and S. Grigolato, "Determination of forest road surface roughness by kinect depth imaging," *Annals of Forest Research*, vol. 60, no. 2, 2017.

[9] M. Sarker, S. Hadigheh, and D. Dias-da-Costa, "Stereoscopic modelling and monitoring of roughness in concrete pavements," in *Proc. ACMSM25*, 2020.

[10] S. Nakashima, S. Aramaki, Y. Kitazono, S. Mu, K. Tanaka, and S. Serikawa, "Application of ultrasonic sensors in road surface condition distinction methods," *Sensors*, vol. 16, no. 10, 2016.

[11] P. P. Smith and K. Zografos, "Sonar for recognizing the texture of pathways," *Robotics and Autonomous Systems*, vol. 51, no. 1, 2005.

[12] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, K. Lau, C. Oakley, M. Palatucci, V. Pratt, P. Stang, S. Strohband, C. Dupont, L.-E. Jendrossek, C. Koelen, C. Markey, C. Rummel, J. van Niekerk, E. Jensen, P. Alessandrini, G. Bradski, B. Davies, S. Ettinger, A. Kaehler, A. Nefian, and P. Mahoney, "Stanley: The robot that won the DARPA grand challenge," *Journal of Field Robotics*, vol. 23, no. 9, pp. 661–692, 2006.

[13] C. Brooks and K. Iagnemma, "Vibration-based terrain classification for planetary exploration rovers." *IEEE Transactions on Robotic*, vol. 21, no. 6, pp. 185–191, 2005.

[14] S. Sattar, S. Li, and M. Chapman, "Road surface monitoring using smartphone sensors: A review," *Sensors*, vol. 18, no. 11, 2018.

[15] G. Singh, D. Bansal, S. Sofat, and N. Aggarwal, "Smart patrolling: An efficient road surface monitoring using smartphone sensors and crowdsourcing," *Pervasive and Mobile Computing*, vol. 40, pp. 71–88, 2017.

[16] Y.-A. Daraghmi, T.-H. Wu, and T.-U. Ik, "Crowdsourcing-based road surface evaluation and indexing," *IEEE Trans. on Intelligent Transportation Systems*, 2021, in Press.

[17] S. Wu and A. Hadachi, "Road surface recognition based on deepsense neural network using accelerometer data," in *Proc. IEEE Intelligent Vehicles Symposium*, 2020.

[18] M.-T. Cao, Q.-V. Tran, N.-M. Nguyen, and K.-T. Chang, "Survey on performance of deep learning models for detecting road damages using multiple dashcam image resources," *Advanced Engineering Informatics*, vol. 46, 2020.

[19] D. Lydon, S. E. Taylor, M. Lydon, and J. Early, "A review of vision based methods for pothole detection and road profile analysis," in *Proc. Civili Engineering Research in Ireland (CERI)*, 2020.

[20] V. Slavkovikj, S. Verstockt, W. De Neve, S. Van Hoecke, and R. Van de Walle, "Image-based road type classification," in *Proc. ICPR*, 2014.

Fig. 9. Example of images having a large difference in predictions between the CNN and the baseline methods. The captions include (from left to right): the ground-truth surface roughness (in brackets), CNN-predicitions, and baseline-predictions.

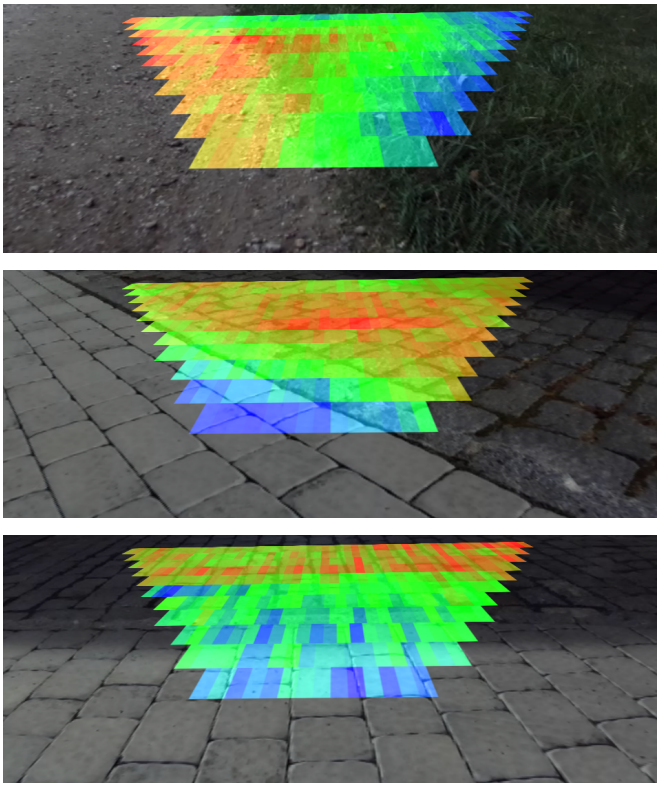

Fig. 10. Color-coded surface roughness calculated by the trained CNN executed in a scanning window over the input image rectified to bird's-eye view, colorized by the estimated roughness, and finally back-projected to the raw camera image.

[21] J.-M. Dai, T.-A. J. Liu, and H.-Y. Lin, "Road surface detection and recognition for route recommendation," in *Proc. IEEE Intelligent Vehicles Symposium (IV)*, 2017.

[22] A. Riid, D. L. Manna, and S. Astapov, "Image-based pavement type classification with convolutional neural networks," in *Proc. 24th International Conference on Intelligent Engineering System*, 2020.

[23] D. Lee, S. Kim, H. Lee, C. C. Chung, and W.-Y. Kim, "Paved and unpaved road segmentation using deep neural network," in *Proc. ACPR*, 2019.

[24] V. Tumen, O. Yildirim, and B. Ergen, "Recognition of road type and quality for advanced driver assistance systems with deep learning," *Elektronika ir Elektrotechnika*, vol. 24, no. 6, 2018.

[25] E. Sabanovic, V. Zuraulis, O. Prentkovskis, and V. Skrickij, "Identification of road-surface type using deep neural networks for friction coefficient estimation," *Sensors*, vol. 20, no. 3, 2020.

[26] M. Bahnik, D. Filyo, D. Pekarek, M. Vlasimsky, J. Cech, T. Hanis, and M. Hromcik, "Visually assisted anti-lock braking system," in *Proc. IEEE Intelligent Vehicles Symposium (IV)*, 2020.

[27] K. Li, L. Yu, S. Wang, and P.-A. Heng, "Towards cross-modality medical image segmentation with online mutual knowledge distillation," in *Proc. AAAI*, 2020.

[28] H. Alwassel, D. Mahajan, B. Korbar, L. Torresani, B. Ghanem, and D. Tran, "Self-supervised learning by cross-modal audio-video clustering," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[29] T. Afouras, A. Owens, J. S. Chung, and A. Zisserman, "Self-supervised learning of audio-visual objects from video," in *Proc. ECCV*, 2020.

[30] D. Eigen and R. Fergus., "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in *Proc. ICCV*, 2015.

[31] Y. Wang, W.-L. Chao, D. Garg, B. Hariharan, M. Campbell, and K. Q. Weinberger, "Pseudo-LiDAR from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving," in *Proc. CVPR*, 2019.

[32] C. Chun and S.-K. Ryu, "Road surface damage detection using fully convolutional neural networks and semi-supervised learning," *Sensors*, vol. 19, no. 24, 2019.

[33] T. Lee, C. Chun, and S.-K. Ryu, "Detection of road-surface anomalies using a smartphone camera and accelerometer," *Sensors*, vol. 21, no. 2, 2021.

[34] M. O. Shneier, T. Chang, T. Hong, W. P. Shackleford, R. V. Bostelman, and J. S. Albus, "Learning traversability models for autonomous mobile vehicles," *Autonomous Robots*, vol. 24, pp. 66–86, 2008.

[35] M. Bajracharya, A. Howard, L. H. Matthies, B. Tang, and M. Turmon, "Autonomous off-road navigation with end-to-end learning for the lagr program," *Journal of Field Robotics*, vol. 26, no. 1, pp. 3–25, 2009.

[36] K. van Hecke, G. de Croon, L. van der Maaten, D. Hennes, and D. Izzo, "Persistent self-supervised learning: From stereo to monocular vision for obstacle avoidance," *International Journal of Micro Air Vehicles.*, vol. 10, no. 2, pp. 186–206, 2018.

[37] Loeber Motors blog, "What is Mercedes-Benz Magic Body Control?" 2015, https://www.loebermotors.com/blog/mercedes-benz-magic-body-control/.

[38] J Groves. Car Magazine, "Audi predictive active suspension: does it work?" 2020, https://www.carmagazine.co.uk/car-news/tech/audi-predictive-active-suspension-how-does-it-work/.

[39] Jaguar Land Rover, "Pothole detection technology research announced by jaguar land rover," 2020, https://www.landrover.com/experiences/news/pothole-detection.html.

[40] The Car Connection, "New tech could alert drivers to potholes," 2019, https://www.thecarconnection.com/news/1123080_new-tech-could-alert-drivers-to-potholes.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, 2016.

[42] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. International Conference on Machine Learning*, 2015.

[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization." in *Proc. ICLR*, 2015.

[44] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[45] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. ECCV*, 2018.