Hauke Brunken* and Clemens Gühmann

Pavement distress detection by stereo vision

Straßenzustandserkennung durch stereoskopische Bildverarbeitung

DOI 10.1515/teme-2019-0046

Abstract: For road maintenance up-to-date information about road conditions is important. Such information is currently expensive to obtain. Specially equipped measuring vehicles have to perform surface scans of the road, and it is unclear how to automatically find faulty sections in these scans.
This research solves the problem by stereo vision with cameras mounted behind the windshield of a moving vehicle so that the system can easily be integrated into a large number of vehicles. The stereo images are processed into a depth map of the road surface. In a second step, color images from the cameras are combined with the depth map and are classified by a convolutional neural network. It is shown that the developed system is able to find defects that require knowledge about surface deformations. These defects could not have been found on monocular images. The images are taken at usual driving speed.

Keywords: Pavement distress detection, stereo vision, convolutional neural network, fault recognition.

Zusammenfassung: Für die Straßenerhaltung sind aktuelle Informationen über den Straßenzustand wichtig. Zurzeit ist die Beschaffung dieser Informationen jedoch teuer und zeitaufwendig. Speziell ausgestattete Messfahrzeuge müssen Oberflächenscans des Straßennetzes erstellen und es ist unklar, wie fehlerhafte Abschnitte in diesen Daten gefunden werden können.
Diese Arbeit löst die Aufgabe durch stereoskopische Bildverarbeitung von Kameras, die hinter der Windschutzscheibe eines Fahrzeugs montiert werden. Dadurch kann das System leicht in eine große Anzahl von Fahrzeugen integriert werden. Die stereoskopischen Bilder werden zu Tiefenbildern der Straßenoberfläche verarbeitet. In einem zweiten Schritt werden Farbbilder mit den Tiefenbildern kombiniert und durch ein faltendes neuronales Netz klassifiziert.
Das entwickelte System ist in der Lage Fehlstellen zu finden, für die Wissen über die Oberflächenverformungen nötig ist. Auf monokularen Bildern hätten sie nicht gefunden werden können. Die Bilder werden bei normalen Fahrgeschwindigkeiten aufgezeichnet.

Schlüsselwörter: Straßenzustandserkennung, stereoskopische Bildverarbeitung, faltendes neuronales Netz, Fehlererkennung.

***Corresponding author: Hauke Brunken,** Fachgebiet elektronische Mess- und Diagnosetechnik, Technische Universität Berlin, Einsteinufer 17, 10587 Berlin, E-Mail: hauke.brunken@tu-berlin.de
Clemens Gühmann, Fachgebiet elektronische Mess- und Diagnosetechnik, Technische Universität Berlin, Einsteinufer 17, 10587 Berlin, E-Mail: clemens.guehmann@tu-berlin.de

1 Introduction

In order to maintain road networks, it is important to know which sections need attention the most. If developing defects are found and fixed on time, severe damages might even be prevented. Defects can be divided into two categories: Those whose detection requires knowledge about the shape of the surface and those whose detection does not. Examples for the first category are surface depressions and rutting. Cracks and potholes fall into the second category, as cracks are not necessarily accompanied by surface deformations and potholes can be found without such information from the shape alone.

In [2] an overview about pavement distress detection methods is given. Surface defects for whose detection knowledge about surface deformations are not important can be detected by analyzing camera images. In [4] a dataset for the purpose of training deep learning algorithms is given. By utilizing a deep neural network, good classification results for the defects cracks, potholes, patches and open joints are shown. The images are taken from cameras that point perpendicularly to the road surface. That eliminates perspective distortion, but requires external installations to the vehicle.

A common approach to measuring surface deformations of roads is to use measurement vehicles equipped with LIDAR and laser triangulation devices [2, 4, 5]. Although they provide detailed depth maps, the disadvantages are the high cost and the need for specially equipped vehicles [7]. The question of how to automatically detect faulty sections in these depth maps has received little attention so far.

This paper focuses on the detection of defects that require knowledge of surface deformation. Instead of relying on complex measuring vehicles, two cameras are used that can be mounted behind the windscreen of any vehicle. Structured light can be used to extract 3D information from cameras, as is shown in [9] for the inspection of free-form metal surfaces. However, this would require an external light source. By using the cameras as a stereo pair, detailed depth maps of the road in front of the vehicle can be extracted. The problem of automatically finding faulty sections is addressed by a deep convolutional neural network.

2 Previous work

The review papers [2] and [7] show that a lot of work has been published on the detection of surface cracks and potholes by analyzing camera images. For surface deformations, LIDAR and laser triangulation devices are widely used to provide depth maps or two-dimensional surface profiles. The automatic detection of faulty sections, however, has received little attention.

In [11] and [12] a depth map is fed into a convolutional neural network for the purpose of crack detection. In [3] sequential two-dimensional road profiles are converted to piecewise standard deviations of height measurements and are then concatenated to a two-dimensional array. This array is then fed into a convolutional neural network in order to predict the level of road degradation. A classification into types of defects is not carried out. In [10] the watershed algorithm, which virtually fills depth maps with water, is used to find potholes in 3D surface scans. The classification is carried out by analyzing the depth and the size of the covered area. One can assume that the distinction between potholes and other surface defects, like depressions, poses a problem. It is also difficult to use a watershed algorithm if the border of a surface defect is not clearly visible. In [13] different types of surface defects are found in surface depth maps by thresholding against a reconstructed, error-free surface. Connecting regions are found and depending on the size, the bounding box aspect ratio and if the part is below or above the mean road surface, the defective region is classified.

The remainder of this paper is structured as follows: In Section 3 the setup of the measurement system is described. The depth map extraction is covered in Section 4 and in Section 5 the recognition of faulty section in these maps is shown. Results are presented in Section 6.

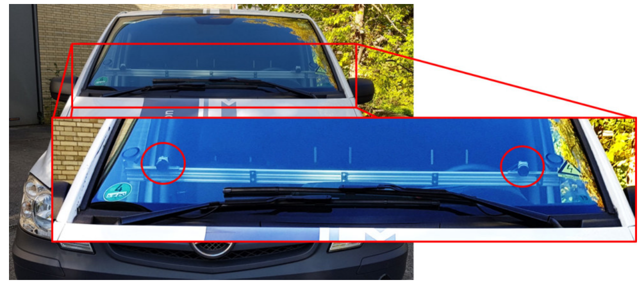


Fig. 1: Stereo cameras are installed behind the windshield of the test vehicle.

3 Measuring system

The measuring system consists of two Basler acA1920-150uc global shutter color cameras. For easy integrability they are mounted behind the windshield of the test vehicle (Fig. 1). For the detection of surface defects in the millimeter range, a measurement with high depth resolution is necessary. Since in stereo imaging the depth resolution increases with a wider baseline (i. e. the distance between camera centers), the baseline should be chosen to be as wide as possible, which is 1.08 m in this case. To cover a lane width of 2 m at a distance between 4 m and 11 m in a single stereo image pair, 25 mm lenses are employed. In order to increase the intersection of camera views, they are not aligned in parallel, but inclined towards each other by 6° each and tilted to the ground by 13°. The cameras are triggered externally in order to record the pictures as synchronously as possible.

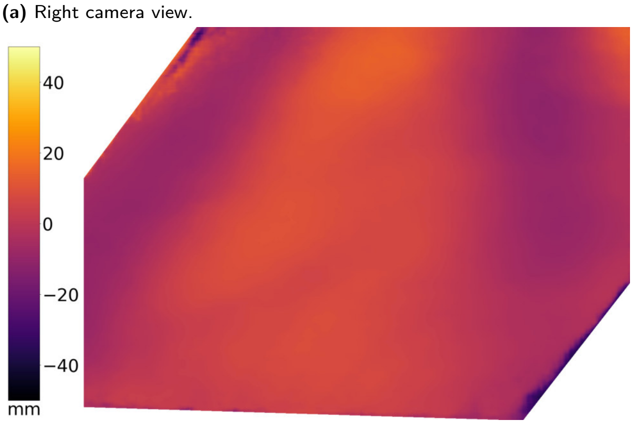
The image sensors have an optical size of 2/3 inches and a resolution of 1920 pixel × 1200 pixel. This results in a depth resolution between approximately $1\frac{\text{mm}}{\text{pixel}}$ at a distance of 4 m and $2.5\frac{\text{mm}}{\text{pixel}}$ at a distance of 11 m above ground from the cameras.

4 Depth extraction

The task of extracting depth from stereo images can be summarized to matching pixels between a left and a right camera image. The distance between the position of matching pixels together with the geometric setup corresponds to the position of the corresponding object point in 3D space. In this work, the algorithm that we published in [1] is utilized for this purpose.

It consists of a neural network that converts stereoscopic images of an approximately flat surface into a depth map, as can be seen in Fig. 2. Due to the perspective distortion of the camera images and thus of the depth maps,

(a) Right camera view.



(b) Corresponding depth map to the view from (a). The depth is measured in relation to the mean road surface.

Fig. 2: Asphalt road with rutting.

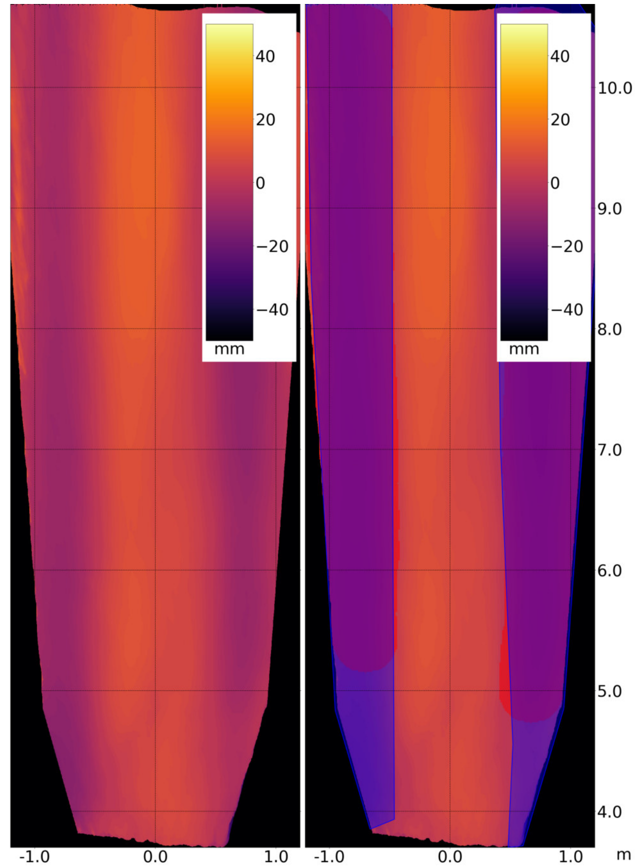


Fig. 3: Depth map from Fig. 2b, converted to a bird's-eye view, with and without labels. The manual annotation is shown in blue, the prediction in purple.

the same surface defect looks different depending on the position in the image. For this reason, the depth map and the color image are converted into a bird's-eye view. For each pixel of the depth map the coordinates are found in 3D space. The x-y plane is arranged in such a way that it corresponds to the mean road surface. The x and y coordinates of points are rounded to integer values. They are used as indices for an array into which the z and RGB values are copied. The resulting array can be handled as a regular image.

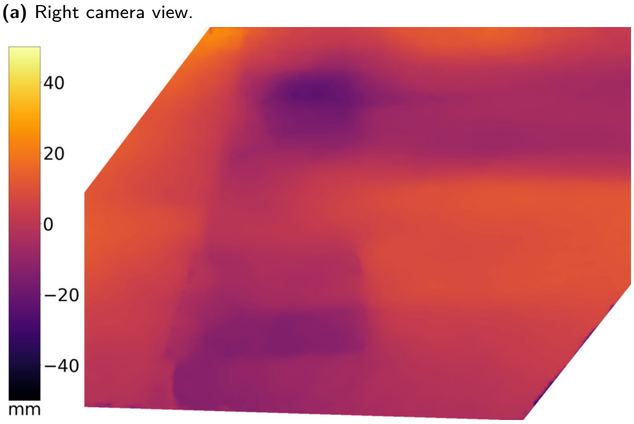
5 Fault recognition

Although in [13] good results with a handcrafted classification algorithm are shown, such algorithms tend to fail in situations that have not been considered during the design. It is believed that a convolutional neural network is more suitable for this purpose. The drawback of neural networks and machine learning algorithms in general is the need for labelled training data, which is not easy to acquire. Labelling the data is a tedious and time-consuming task. For that reason, the objective in the design of the neural network has to be a well utilization of the training data.

The other aspect in creating the dataset is the labelling of surface defects itself. Especially for surface depressions it is difficult to define objective criteria that decide if a deformation should be classified as such. The problem can be seen in Fig. 4. A threshold depth value is unsuitable for marking a surface depression because it is unclear in relation to what height the depth should be measured. Usually, the depression does not have a clear boundary and if it is decided that the deformation is a depression, the question is where it starts and where it ends. This consideration also led to the idea of utilizing a neural network instead of a handcrafted classification algorithm.

In [6] a neural network for semantic segmentation is presented. It is trained on a dataset for urban scene understanding, consisting of frames of a video sequence. Only 367 frames were used for training, 101 for validation

(a) Right camera view.



(b) Corresponding depth map to the view from (a). The depth is measured in relation to the mean road surface.

Fig. 4: Asphalt road with surface depression.

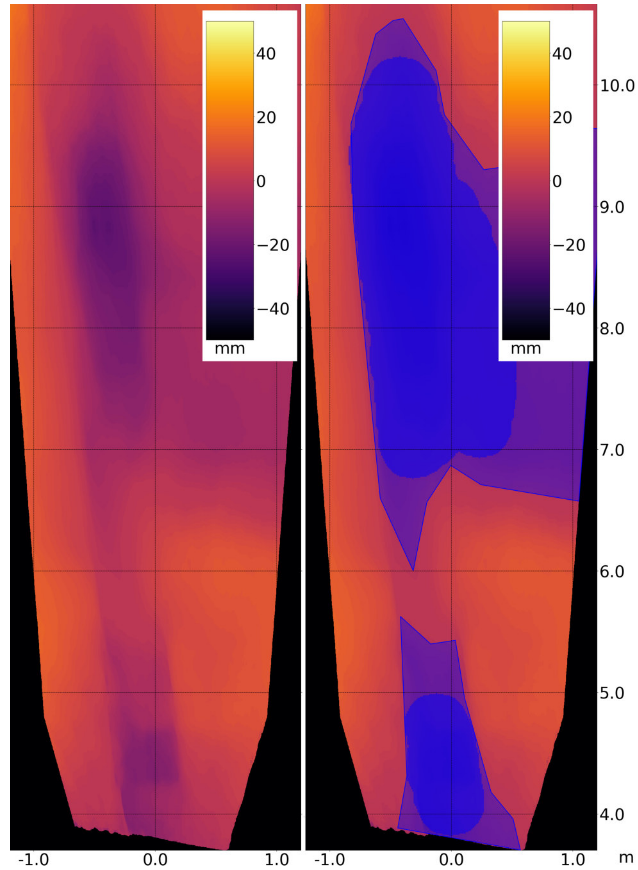


Fig. 5: Depth map from Fig. 4b, converted to a bird's-eye view, with and without labels. The manual annotation is shown in blue, the prediction in purple.

and 233 for testing. The dataset is not only very small, sequential images are also very similar, which reduces the amount of information. Nevertheless, it performs very well. Therefore, the neural network from [6] is adapted to the purpose of defect detection.

As input for the neural network, the RGB values are converted to greyscale and then concatenated with the green and the depth channels. The network from [6] is adapted in the following way:

– Defects manifest themselves differently in the color image and in the depth map. Therefore, the output of the first convolutional layer does not mix the input channels.
– Different surface defects can occur together. E. g. a long rut can contain a local deeper depression. To train a neural network through backpropagation and gradient decent, it is important that the training data is not contradictory. In order to avoid contradictions in the labelled data, every label is classified against the background. Thus, the network has twice as many outputs as labels, such that each label has its own background. The SoftMax function is applied pairwise on the outputs.
– The sum of the pairwise binary cross entropy is used as a loss function for training.

A morphological opening is applied as a post-processing step to remove small patches, which are considered noise.

6 Results

Fig. 3 shows the result for a road surface with rutting. The predicted label is similar to the manual annotation. Only at the bottom of the picture the prediction differs from the annotation. Fig. 5 shows the result of a road with surface depressions. The shapes of the prediction and annotation differ, but the locations are correct. The reason is probably an inconsistent labelling of the data. If one looks at the depth map in Fig. 4b, it is difficult to decide what the correct size of the depressions should be.

7 Conclusion and outlook

The developed system covers the entire measurement chain from recording images, through processing stereo images into depth maps to segmenting and classifying the data. The stereo camera setup produces surface measurements of roads that are comparable to laser scans for the purpose of pavement distress detection. Tests have been performed at vehicle speeds of up to $80\,\mathrm{km\,h^{-1}}$ and higher speeds should be possible. Thus far, 193 images are used for training, 18 for testing and only two kinds of defects are considered. Nevertheless, the results look promising. The segmentation and classification results show a high consistency with the annotated data. The system is completely automatic, which makes it possible to map surface defects of large road networks.

In order to quantify the results, more images need to be labelled and more types of defects should be considered. Since it is often unclear how to label the data, the annotations can be regarded as uncertain. This can be taken into account during training with cost functions for uncertain training data, such as shown in [8]. Experiments have to be carried out to test its effectiveness on the existing data set.

Acknowledgment: We would like to thank Daimler AG for providing the test vehicle (Vito E-Cell).

References

[1] H. Brunken and C. Gühmann. Incorporating Plane-Sweep in Convolutional Neural Network Stereo Imaging for Road Surface Reconstruction. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP*, pages 784–791. SciTePress, 2019. ISBN 978-989-758-354-4. 10.5220/0007352107840791.

[2] T. B. Coenen and A. Golroo. A review on automated pavement distress detection methods. *Cogent Engineering*, 4(1), Sept. 2017. ISSN 2331-1916. 10.1080/23311916.2017.1374822. URL https://www.cogentoa.com/article/10.1080/23311916.2017.1374822.

[3] M. Eguchi, A. Kawamura, K. Tomiyama, and O. Shinichiro. A Simplified Method of Detecting Spot Surface Defects by using Quasi-3d Data from a Conventional Road Profiler. *Transportation Research Record: Journal of the Transportation Research Board*, page 036119811983456, June 2019. ISSN 0361-1981, 2169-4052. 10.1177/0361198119834567. URL http://journals.sagepub.com/doi/10.1177/0361198119834567.

[4] M. Eisenbach, R. Stricker, D. Seichter, K. Amende, K. Debes, M. Sesselmann, D. Ebersbach, U. Stoeckert, and H.-M. Gross. How to get pavement distress detection ready for deep learning? A systematic approach. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 2039–2047, Anchorage, AK, USA, May 2017. IEEE. ISBN 978-1-5090-6182-2. 10.1109/IJCNN.2017.7966101. URL http://ieeexplore.ieee.org/document/7966101/.

[5] German Federal Road Research Institute. Measuring vehicles and systems, 2018. URL https://www.bast.de/durabast/DE/durabast/Messfahrzeuge/fahrzeuge_node.html.

[6] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio. The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1175–1183, July 2017. 10.1109/CVPRW.2017.156.

[7] A. Ragnoli, M. De Blasiis, and A. Di Benedetto. Pavement Distress Detection Methods: A Review. *Infrastructures*, 3(4):58, Dec. 2018. ISSN 2412-3811. 10.3390/infrastructures3040058. URL http://www.mdpi.com/2412-3811/3/4/58.

[8] S. Reed, H. Lee, D. Anguelov, C. Szegedy, D. Erhan, and A. Rabinovich. Training Deep Neural Networks on Noisy Labels with Bootstrapping. *arXiv:1412.6596 [cs]*, Dec. 2014. URL http://arxiv.org/abs/1412.6596. arXiv: 1412.6596.

[9] M. Strohmeier, M. Schröder, and C. Faber. Optical inline inspection detecting 3d defects on complex free-form surfaces in harsh production environments. *tm-Technisches Messen*, 86(6):335–344, 2019.

[10] Y.-C. J. Tsai and A. Chatterjee. Pothole Detection and Classification Using 3d Technology and Watershed Method. *Journal of Computing in Civil Engineering*, 32(2):04017078, Mar. 2018. ISSN 0887-3801, 1943-5487. 10.1061/(ASCE)CP.1943-5487.0000726. URL http://ascelibrary.org/doi/10.1061/%28ASCE%29CP.1943-5487.0000726.

[11] A. Zhang, K. C. P. Wang, B. Li, E. Yang, X. Dai, Y. Peng, Y. Fei, Y. Liu, J. Q. Li, and C. Chen. Automated Pixel-Level Pavement Crack Detection on 3d Asphalt Surfaces Using a Deep-Learning Network: Pixel-level pavement crack detection on 3d asphalt surfaces. *Computer-Aided Civil and Infrastructure Engineering*, 32(10):805–819, Oct. 2017. ISSN 10939687. 10.1111/mice.12297. URL http://doi.wiley.com/10.1111/mice.12297.

[12] A. Zhang, K. C. P. Wang, Y. Fei, Y. Liu, S. Tao, C. Chen, J. Q. Li, and B. Li. Deep Learning–Based Fully Automated Pavement Crack Detection on 3d Asphalt Surfaces with an Improved CrackNet. *Journal of Computing in Civil Engineering*, 32(5):04018041, Sept. 2018. ISSN 0887-3801, 1943-5487. 10.1061/(ASCE)CP.1943-5487.0000775. URL http://ascelibrary.org/doi/10.1061/%28ASCE%29CP.1943-5487.0000775.

[13] D. Zhang, Q. Zou, H. Lin, X. Xu, L. He, R. Gui, and Q. Li. Automatic pavement defect detection using 3d laser profiling technology. *Automation in Construction*, 96:350–365, Dec. 2018. ISSN 09265805. 10.1016/j.autcon.2018.09.019. URL https://linkinghub.elsevier.com/retrieve/pii/S0926580518304813.