Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-8071-1_P289-cd

Creating 3D Models of Bridges Using Different Data Sources and Machine Learning Methods

Kwasi Nyarko Poku-Agyemang [1,2*]
E-mail: kwasi.nyarko@inatech.de

Maximilian Kellner [1,2 *]
E-mail: maximilian.kellner@inatech.de

Annette Schmitt [1,2]
E-mail: annette.schmitt@ipm.fraunhofer.de

Alexander Reiterer [1,2]
E-mail: alexander.reiterer@ipm.fraunhofer.de
1 University of Freiburg, Department of Sustainable Systems Engineering INATECH, Germany
2 Fraunhofer Institute for Physical Measurement Techniques IPM, Germany
** contributed equally*

In today's world, aging and worn bridges pose an increasing risk to transportation infrastructure. In the worst case, old, poorly maintained bridges can collapse at any time. But complex and expensive maintenance work on the bridges causes traffic jams, which can lead to accidents or delivery problems. Therefore, bridges require intelligent and individual maintenance, which leads to a higher demand for documentation. One way to facilitate documentation is Building Information Modeling (BIM), which is based on a 3D model of the construction. For most of the German bridges no 3D data is available. So, it is necessary to create a 3D model as a base for the BIM by Scan-to-BIM processes. The 3D data for this process can come from a wide variety of sources like laser scanning, photogrammetry or analog 2D plans. A concept for automated 3D modelling with data from diverse sources and machine learning methods is presented. Point clouds of the bridges captured with cameras and/or laser scanners and 2D plans are used as data base for the 3D model, which is created by machine learning methods from the fused point clouds by calculating surfaces. The resulting model can be used for BIM and AR/VR applications.

Keywords: Machine Learning, Point Cloud, 3D Reconstruction, Deep Learning, bridge

1. Introduction

Nowadays traffic infrastructure is one of the most sensitive parts of our world. Without an intact traffic infrastructure, supply chains will collapse and our daily lives are thrown off course. Therefore, it is more important than ever to maintain our transport infrastructure. Within the traffic infrastructure, bridges are of special importance. A partially or fully blocked bridge causes long and preventable traffic jams. Therefore, it is necessary to recognize the necessity of maintenance work early and plan the works wisely. A helpful planning tool for maintenance processes is building information modeling (BIM). But the base of such a model is a digital geometrical 3D representation of the bridge. Especially older bridges with a high maintenance effort do not have such a model.

In this work, we present a workflow for a 3D reconstruction of bridges using different data sources and machine learning (ML) approaches. Two sensor systems and a 2D planebased method are introduced, as well as a 3D segmentation pipeline. The workflow is illustrated by an example. We choose a pedestrian bridge in Freiburg, Germany, across the river Dreisam to test the process. The so called Ottiliensteg, shown in Figure 1, is approximately 35 m long, 2 m wide and has a height of about 4 m.

Fig. 1.: Ottiliensteg in Freiburg, Germany

2. Related Work

The base for building information modeling is always a digital geometrical 3D model of the complex structure, which is mostly not available. First of all the structure has to be captured by suitable sensors or a digitization of plans is necessary.
There are different sensors for capturing complex 3D structure available on the market. Bornaz and Rinaudo (2004) use for example terrestrial lasers scanner (TLS). On the other hand, mobile laser scanners (MLS) are used by Kukko et al. (2012). Additionally, photogrammetry could be used to create a 3D point cloud, but this is not part of this work.
A different approach for creating the 3D model of the complex structure, is using 2D plans. Poku-Agyemang and Reiterer (2023) describe a semi-automatic method to generate a 3D point cloud from 2D design plans. In this case, the individual components of the bridge, like the superstructure, pillars or abutments, are reconstructed individually. At the end of the process, the point clouds of the different elements are fused to one complete 3D point cloud of the bridge.
The next step in the 3D model generation is 3D segmentation, which is tackled by different approaches today. A general overview is given in Guo et al. (2019). Mainly, these approaches can be divided into projection-based, discretization-based, point-based, and hybrid methods. In Milioto et al. (2019) the point cloud gets projected into a spherical image and processed by 2D convolutions. Sparse 3D tensors are introduced in Choy et al. (2019) to allow processing discretized large scenes with sparse 3D convolutions. KPConv by Thomas et al. (2019) defines kernel points which can be used to directly process point clouds. Another way to work directly on the point cloud has been introduced in Zhao et al. (2021). Here, attention mechanisms are used instead of convolutions. The spherical and bird's eye projections are fused by using a KPConv layer in Kellner et al. (2022).

3. Method

3.1. *3D data capturing*

As already mentioned, a workflow for creating 3D models of bridges is shown based on the example of Ottiliensteg. Three different kind of data sources are used to create point clouds of the Ottiliensteg bridge. First of all the terrestrial laser scanner Leica RTC360 by Leica Geosystem AG (2023b) has been used to capture the bridge from different position on and around the bridge. The scans from different position are registered using the Leica Cyclone post-processing software to create a complete reconstruction of the bridge.
Additionally, the handheld mobile laser scanner Leica BLK2GO by Leica Geosystem AG (2023a) has been used to capture the bridge by continuously walking on and around the bridge. Based on the so called GrandSLAM technique, the BLK2Go creates the 3D point cloud in realtime. The GrandSLAM fuses simultaneous location and mapping from visual and lidar with IMU data.
At last, 2D plans provided from Stadt Freiburg, are used to create a point cloud of the individual elements of the bridge. The components of the bridge on the 2D plan are the north and south abutment, railing and the concrete superstructure. The individual point clouds of the elements are created with the pipeline introduced by Poku-Agyemang and Reiterer (2023) and fused into a complete point cloud of the bridge.

3.2. *3D Deep Learning*

Because of possible differences between as built and as planned elements, we do not want to rely the segmentation solely on the point cloud generated from 2D plans. For this reason, we segment the captured point clouds from the 3D sensor by using KPConv. We choose $K = 15$ deformable kernel points with an input radius of

$r = 3$ m and $r = 5$ m leading to the neighborhood $N_x = \{x_i \in P \mid \|x_i - x\| \leq r\}$ with the center point x. To ensure that the number of points in the defined sphere N_x is not too large, the entire point cloud is reduced in advance by voxel downsampling using a voxel size of 0.03 m for the smaller radius and 0.1 m for the larger radius. Since we do not assume that color coding of points is given in all cases, we will train the Neural Network (NN) only on the 3D points themselves and not use colors as input features.

The classes to be identified are reduced to the general basic structures as well as to auxiliary classes. The fundamental basis is the dataset from Lu et al. (2019) containing 10 RC highway bridges around Cambridgeshire, United Kingdom. We adapted the data and enriched it by two more bridges recorded in Freiburg leading to the dataset defined in Table 1. It is worth noting that the ratio for the distribution of points in both the training and test data is very atypical. The reason for this is that the data we recorded was taken with a different sensor and has a higher density of points. Since these bridges are used for testing purposes, this distribution occurs.

Table 1.: Semantic segmentation class list with amount of points belonging to it. The data belonging to train are used to train the model, while the data belonging to test are used only to evaluate the performance of the trained model.

Class	ID	Train	Test
Unlabeled	0	1.463.051	153.963
Ground	1	21.673.976	35.406.474
High vegetation	2	12.193.852	32.584.318
Abutment	3	21.091.544	7.133.976
Superstructure	4	42.809.109	17.156.002
Road surface	5	23.300.776	10.381.405
Railing	6	7.261.507	2.932.185
Traffic signs	7	13.451	3.741
Pillar	8	6.893.551	7.498.267

3.3. *Fusion*

The complexity of a bridge structure makes it very difficult to fully record it with just one data capturing technique. The various 3D data acquisition techniques have their advantages and limitations in the reconstruction of the bridges. For example, the TLS- and MLS-based methods could capture just the visible parts of the bridges. The base or big parts of the abutment are not visible, therefore the point clouds created from 2D plans are the perfect addition. The 3D point clouds of all three data sets are fused by registering them into the same coordinate system. The fused point cloud provides a fully reconstructed point of the bridge including the visible and invisible- underground parts of the bridge.

4. Experiments

4.1. *3D data capturing*

To reconstruct the 3D model from the laser scanner measurements, the 10 TLS scans on and around Ottiliensteg footbridge are processed in Leica Cyclone software. Approximately 185 million registered points create a dense 3D model of the bridge and its environs as shown in Figure 2a. The 3D point cloud from the MLS was obtained directly from the Leica BLK2go with approximately 4 minutes of measurement on and around the bridge. The point cloud reconstructed by the MLS consists of approximately 42 million points as shown in Figure 2b. Both point clouds was obtained by the TLS and MLS consist of the geometric coordinates and RGB-data from the internal cameras from the sensor systems.

The 2D plans from the Stadt Freiburg were used to reconstruct individual components of the bridge as designed. The reconstruction was done at a resolution of 1 cm interval and a unique RGB information was provided for each bridge component. The reconstruction provides both visible and invisible parts of the bridge as shown in Figure 2c. Although the RGB data provide additional information about the bridge, they have no semantic meaning. For this purpose, the points must be assigned to the individual components of a bridge.

(a) TLS: point cloud

(b) MLS: point cloud

(c) Bridge Plan point cloud

Fig. 2.: 3D reconstruction of Ottiliensteg Bridge

4.2. *3D Deep Learning*

For evaluation we use the mean intersection-over-union (mIoU). As the classes within the dataset are imbalanced (see Table 1) we will use a weight for each class $w_c = \frac{1}{\sqrt{f_c}}$ leading to the weighted cross-entropy loss $\mathcal{L}_{\text{wce}}$:

$$\mathcal{L}_{\text{wce}} = -\frac{1}{\sum_{c \in C} w_c} \sum_{c \in C} w_c y_c \log \hat{y}_c \quad (1)$$

The Lovász-Softmax loss $\mathcal{L}_{\text{ls}}$ Berman et al. (2018) allows for optimizing the IoU metric. It is defined as:

$$\mathcal{L}_{\text{ls}} = -\frac{1}{|C|} \sum_{c \in C} \Delta_{J_c}(\boldsymbol{e}(c)) \quad (2)$$

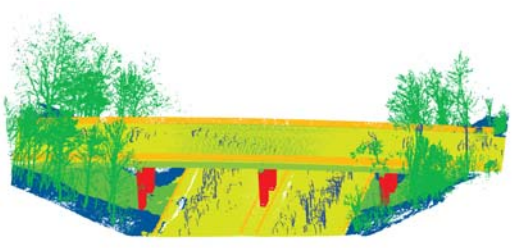
Where $\boldsymbol{e}(c)$ is the vector of pixel errors for class c and Δ_{J_c} is the Lovász extension of the IoU.
To optimize for the pixel wise accuracy and the IoU we use a liner combination of both losses $\mathcal{L} = \mathcal{L}_{\text{wce}} + \mathcal{L}_{\text{ls}}$. To avoid overfitting the data gets augmented. First, we drop a random amount of $\mathcal{U}(0, 0.3)$ points. Afterward, the x, y, and z position of each point gets shifted by the value of $\mathcal{U}(-2, 2)$, the point cloud gets rotated around the z-axis by an angle between $30°$ to $330°$, each point cloud gets scaled by a value of $\mathcal{U}(0.8, 1.2)$ and we add noise with $\mathcal{N}(\mu = 0, \sigma^2 = 0.01)$. Each augmentation but the first gets applied independently with a probability of 0.5.
The test results are shown in Table 2. Due to the increase of the receptive field, the metric has slightly decreased. However, this may be due to the lower resolution of the points. Because of the high memory requirements, it is not practical to increase the radius while keeping the voxel size constant. The metric can be further optimized by finding the right balance between receptive field, voxel size, and memory requirements.

An example bridge and its associated prediction is illustrated in Figure 3. It is important to mention that we did not use the color as input feature for training the NN.

Table 2.: IoU for each class on the test set with different input radius r. The overall mIoU decreases slightly with an increasing radius, but that may also be due to the lower resolution.

Class	$r = 3$ m	$r = 5$ m
Unlabeled	0.48	0.51
Underground	0.77	0.68
High vegetation	0.93	0.93
Abutment	0.66	0.5
Superstructure	0.91	0.92
Top surface	0.55	0.6
Railing	0.88	0.86
Traffic signs	0.02	0.01
Pillar	0.98	0.95
mIoU	0.69	0.66

Input bridge



Prediction of our trained NN

Fig. 3.: Example test bridge. The trained NN is able to make a good prediction for unseen bridges. No color was used for the input of the NN, it is only for illustration for the reader.

4.3. *Fusion*

The segmented semantic point cloud from the 3D reconstruction from the sensor system and the 3D reconstruction of the 2D plans are fused in the same coordinate system. The coordinate system of the TLS is used as the reference system for the registration process. Visible keypoints for example the ends of the hand railing on the sides of the bridge are used to compute the translation and rotation parameter. The RMS errors for the registration is about $0.03\ m$ and $0.11\ m$ for the registration between the MLS and the 2D plan respectively. The fused semantic point cloud consists of both the visible and invisible parts of the bridge above and beneath the earth's surface. The final fused point cloud and the bridge separated from the environment using the semantic information can be seen in Figure 4.

Full fused bridge

Bridge separated from environment

Fig. 4.: Fused semantic point cloud of Ottiliensteg bridge. Supplementing the data from the scanners with the data from the 2D plans is valuable because it gives the final model information that was not originally visible. It can be seen most clearly at the abutment.

5. Conclusion

We have shown how a homogeneous point cloud can be generated from different data and sensor sources. Using the semantic information from the 2D plans and the NN, the point cloud can be reduced to the object itself and individual relevant components can be separated.

The next step in this reconstruction method has to be the implementation of a suitable geometric modeling process. In further research, the NN can be further optimized, for example, by using an adaptive receptive field to link global and local context. A greater variation of different bridges within the data can also improve generalization. Furthermore, the fusion can be optimized and it can be investigated in which time point the fusion is most suitable to achieve the best result.

Acknowledgement

The work has been funded by the BMBF as part of the mfund-project “Partially automated creation of object-based inventory models using multi-data fusion of multimodal data streams and existing inventory data - mdfBIM+“ (FKZ: 19FS2021B) and by the German Research Foundation (DFG) as part of the project “Modeling of civil engineering structures with particular attention to incomplete and uncertain measurement data by using explainable machine learning - MoCES” (Project no. 501457924), which is a subproject of ”SPP 2388: Hundred plus - Extending the Lifetime of Complex Engineering Structures through Intelligent Digitalization” (Project no. 461030501). The authors thank Ruodan Lu and Ioannis Brilakis for sharing the data.

References

Berman, M., A. Rannen Triki, and M. B. Blaschko (2018). The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4413–4421.

Bornaz, L. and F. Rinaudo (2004). Terrestrial laser scanner data processing. In *XXth ISPRS Congress Istanbul*.

Choy, C., J. Gwak, and S. Savarese (2019). 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3075–3084.

Guo, Y., H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun (2019). Deep learning for 3d point clouds: A survey. *CoRR abs/1912.12033*.

Kellner, M., B. Stahl, and A. Reiterer (2022). Fused projection-based point cloud segmentation. *Sensors 22*(3).

Kukko, A., H. Kaartinen, J. Hyyppä, and Y. Chen (2012). Multiplatform mobile laser scanning: Usability and performance. *Sensors 12*(9), 11712–11733.

Leica Geosystem AG (2023a). Leica blk laser scanner. https://shop.leica-geosystems.com/de/leica-blk/blk2go.

Leica Geosystem AG (2023b). Leica rtc360 3d laser scanner. https://leica-geosystems.com/products/laser-scanners/scanners/leica-rtc360.

Lu, R., I. Brilakis, and C. R. Middleton (2019). Detection of structural components in point clouds of existing rc bridges. *Computer-Aided Civil and Infrastructure Engineering 34*(3), 191–212.

Milioto, A., I. Vizzo, J. Behley, and C. Stachniss (2019). RangeNet++: Fast and Accurate LiDAR Semantic Segmentation. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*.

Poku-Agyemang, K. N. and A. Reiterer (2023). 3d reconstruction from 2d plans exemplified by bridge structures. *Remote Sensing 15*(3).

Thomas, H., C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas (2019). Kpconv: Flexible and deformable convolution for point clouds. *Proceedings of the IEEE International Conference on Computer Vision*.

Zhao, H., L. Jiang, J. Jia, P. H. Torr, and V. Koltun (2021). Point transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16259–16268.