# Structure from motion - Quantitative Evaluation of structure from motion software for the 3D-reconstruction of traffic accidents

**Tim KAISER, Christian CLEMEN and Robert KADEN, Germany**

**Key words**: Structure from Motion, traffic accident, 3D scene reconstruction

## SUMMARY

In case of severe traffic accidents, German Police is required to provide a scaled sketch of the accident scene. Therefore, police authorities need tools helping them to measure the affected scene in a short period of time to reduce the traffic obstructions while maintaining the demanded statutory accuracy. With the progress in the fields of computer vision and digital photogrammetry new and simple low-cost methods can be established for accurate data acquisition.

In an ongoing research project at the Dresden University of Applied Sciences Structure from Motion techniques are used to generate a 3D-reoconstruction of the accident location as a first step. Afterwards the information contained in the gained dense point cloud has to be extracted and converted into a 2-dimensional vector-based map or 3-dimensional model that can be used in court. This requires that distinctive objects for the mapping of the traffic scene can be located with a high accuracy and on the spatially correct location.

This paper examines a series of Structure from Motion software packages regarding their applicability for traffic accident reconstruction. Special focus was put on the investigation of the geometric accuracy of the registered camera stations since these results are crucial for the following step of generating a dense point cloud. Besides of that other criteria such as hardware requirements or processing time were considered. In the end, a decision for the further used Structure from Motion tool is presented.

# Structure from motion - Quantitative Evaluation of structure from motion software for the 3D-reconstruction of traffic accidents

**Tim KAISER, Christian CLEMEN and Robert KADEN, Germany**

## 1. INTRODUCTION

In case of severe traffic accidents, German police authorities are required to provide a sketch of the incident location that has to contain the most important features of the scene such as the final location of cars, traffic signs, skid marks or road markings. In the state of Saxony this sketch is prepared in the scale of 1:200 so that it can be used in court (Sächsisches Staatsministerium des Innern 2016). This means that accurate reconstruction of the scene using appropriate measurement methods is essential. Clearly, the field of geodesy and surveying is offering various techniques to fulfill this task. Nevertheless, there is always the area of conflict between the reachable accuracy and the ease of use of the respected method since assigned police officers are not intensively trained experts in surveying.

In addition, the exact and accurate surveying of the location is in great conflict with the demand of clearing the affected roads as fast as possible. Closing roads for multiple hours for mapping the scene has negative impact on the remaining traffic. There is an estimation that by saving one minute for the time of road clearance affected motorists will also save four to five minutes on their primary way (Cooner, Balke 2000).

In a prior research project, the Dresden University of Applied Sciences in cooperation with the company SOEASY has developed a specific CAD application for police authorities in Saxony. The software is specifically tailored to the requirements in the respective field of use. The software features a direct connection to a total station with the aim of trying to make the whole mapping workflow as easy as possible. However, police officers still have to be trained extensively for operating the devices. Additionally it would be very cost-intensive to supply each police unit with their own total station and the necessary training for the officers.

The recent and rapid progress in the domain of photogrammetric Computer Vision and especially in the field of Structure from Motion (SfM) influenced the idea to make use of such measuring techniques for the 3D-reconstruction of traffic accidents. Therefore, the overall aim of the current ongoing research project is to implement a SfM-workflow generating a 3D point cloud that in the end can be used to generate an accurate sketch of the accident scene. Since a complete new implementation of a robust SfM-workflow is very complex and time-consuming, we decided to search for an open source pipeline. The overall aim of this paper was to agree on one of the investigated SfM tools for our further research based on this evaluation. Some of the characteristics / questions that were specifically evaluated are among other things:

– Accuracy of the registered camera views

- Hardware requirements and ease of use
- To which extent affect the implemented feature detection algorithms the results of the SfM-workflow?

## 2. RELATED WORK

### 2.1 3D Reconstruction and Mapping of Traffic Accidents

The application of “traditional” geodetic and photogrammetric measurement methods for the reconstruction and mapping of traffic accidents is not something new. Police units around the world make use of various techniques for capturing the incident scene. This includes very basic measuring methods such as surveying wheels or tape measures (Kaden 2005). However, using these devices is time consuming and the accomplished results are negatively affected by error sources like curbs or little hills. In addition, these analogous methods are very error-prone. During the stressful situation of accident reconstruction when officers are under time pressure transposing of digits or reading errors are likely to happen. Therefore, a complete digital data processing workflow can ease the accident reconstruction

That is why many police authorities make use of total stations for the capturing of accident scenes (Agent et al. 2005; Stáňa et al. 2017). This can speed up the reconstruction process especially if the accident scene is widespread. The Saxon police also noted the benefits of a total station so that the development of the customized CAD software ForensicSurvey started.

In recent years, Terrestrial Laser Scanners (TLS) gained more and more attraction for applications in the forensic sector. Buck et al. (2013) for example showed the benefits of TLS data for the reconstruction of the circumstances of fatal casualties. Whereas Pagounis et al. (2006) directly investigated the use of laser scanners for the application of accident reconstruction. They successfully showed that the captured 3D data can be turned into a 2D map respectively sketch of the accident scene. However, TLS produces a huge amount of data that needs to be filtered. This requires a notable amount of time for training the officers. Besides of that the respective gear is expensive so that it might not be feasible to equip every police unit.

Besides of using total stations or Terrestrial Laser Scanning a couple of research projects already investigated the capabilities of photogrammetric methods for the reconstruction of traffic accident scenes. For example Fraser et al. (2008) and Du et al. (2009) demonstrated that photogrammetry can deliver accurate reconstruction results using consumer grade cameras and a self-calibrating process. However, a big downside of the proposed methods are that specific coded targets and/or special scale bars have to be placed at the accident location. The need to prepare the scene properly before the image acquisition is one of the major drawbacks of the photogrammetric measurement technique and significantly increases the burden for a successful application within police departments. Police officers for example have to make sure that they always capture the coded target markers with enough detail so that they can be identified in multiple images. In addition, the present light conditions at the scene have to be considered in a way that the correct matching of the markers becomes possible. Since accident locations are outdoors this can be a challenging task. Furthermore, a notable amount of time is needed for the installation of the markers that in turn extends the duration of road closures and obstructions.

One advantage that both TLS and photogrammetric methods have in common is the fact that by producing a complete 3D reconstruction of the scene it becomes possible to map important objects that might have been missed during the initial data acquisition. After the accident location is once cleared, a later mapping becomes impossible when using a total station or a surveyor’s wheel for the capturing process. The location can be revisited, at least virtually, though when TLS or photogrammetry techniques were applied instead.

These findings show that photogrammetric accident reconstruction can deliver good results. However, it is very important to keep the reconstruction process for the police officers as simple as possible in order to raise the acceptance and to foster the utilization for the particular measuring method. That is why our reconstruction idea makes use of the Structure from Motion approach. As can be seen in the next chapters, Structure from Motion needs neither specialized coded targets nor specialized measurement cameras making the raw data acquisition a lot easier. The most important aspect to get good results is to respect some basic principles regarding the image capturing process and the resulting image network geometry. The teaching of police members in these principles is a lot easier compared to the required training for a total station.

### 2.2 Structure from Motion

Despite of differing in some details throughout the specific pipeline, all investigated software packages perform similar steps for the photo-based 3D reconstruction. The first step always comprises of extracting homologous points (also called keypoints) from the input images. The most common approach for this task is the so-called SIFT-Operator (Scale Invariant Feature Transform) presented by Lowe (2004). All in our evaluation investigated software systems make use of SIFT-Features or slightly variants thereof, except regard3D that uses so-called AKAZE-Features (Alcantarilla et al. 2013). Since the SIFT algorithm is patented (using it for research purposes is free though) AKAZE-Features might be a good alternative for the keypoint extraction process as there are no legal restrictions.

After matching the extracted key points, they are used to determine the cameras orientation and position. Robust parameter estimation is required to filter wrongly matched outliers. In most SfM tools this is done with the help of the RANSAC algorithm (Random Sample Consensus, Fischler, Bolles (1981)) or one of its variations. After determining the camera intrinsics and extrinsics as well as the 3D location of the extracted feature points, a bundle block adjustment is performed to refine the initial values. This step concludes the actual SfM workflow.

In most applications the resulting camera parameter and especially the sparse point cloud contains to less information in terms of measured 3D points. Therefore, dense matching or multi-view stereo algorithms (MVS) are applied to generate a much denser point cloud with much more information. In a first step, the algorithms use the calculated camera parameters to undistort the images. Afterwards the dense point cloud can be generated using the distortion free images.

## 3. EVALUATION OF SOFTWARE PACKAGES

### 3.1 Selected Software Packages

In recent years, numerous open-source software tools and software packages dealing with structure from motion have been developed and released. However, these systems vary in respect to their targeted use case and usability. The first published SfM tools originated from the computer vision discipline. They put their main emphasis on robust reconstruction even when using an inhomogeneous set of input images and not on generating a very accurate 3D object reconstruction as known from photogrammetric tools.

Since it is not possible to evaluate all available SfM-packages at once, we have chosen four systems for further investigation. The four packages are:

- VisualSFM (Wu 2011)
- Colmap (Schonberger, Frahm 2016)
- openMVG (Moulon et al. 2017)
- regard3D (Regard3D 2018)

Furthermore, the tests were carried out with the commercial Software Agisoft Photoscan. Due to its easy handling and usability, Agisoft Photoscan is a very widely used SfM tool in the field of geosciences and can be regarded as a reference tool. For the better comprehension of the results, it is important to note that all tests were performed using the default parameters of the respective software tool. By adjusting some of the changeable parameters, it might be possible to get other results.

The resulting camera positions and point clouds are typically only referenced in a local arbitrary coordinate system. Additionally they are not up to scale. In order to correctly georeference the results, Ground Control Points (GCP) are needed. The various SfM software tools differ in respect to the processing step when the GCP based transformation is performed. Some tools allow to integrate the GCP information into the bundle adjustment step as further constraints. Whereas with other systems the transformation is performed by searching the corresponding point locations in the dense point cloud. This way the parameters for a seven-parameter Helmert transformation can be estimated. In general, the approach to supply the GCP data during bundle adjustment is the preferred way as the step of selecting the corresponding points in the dense point cloud can be a challenging task and has to be regarded as an additional error source.

#### 3.1.1 Implemented camera models

In computer vision and photogrammetry, various options for the modelling of a camera exist. Depending on the chosen camera model a different number of distortion parameters are considered. Using simpler camera models can have negative effect on the obtained results during the MVS stage. Eltner, Schneider (2015) for example demonstrated that SfM pipelines that implement only one radial distortion parameter (such as VisualSFM and Colmap) can produce a dense point cloud with great deviations from the real world object.

To overcome this effect, they undistorted the initial images with their own camera parameters obtained from a separate camera calibration using a specialized test procedure. After running the pipeline with the separately undistorted images better results could be obtained. However, this calibration procedure would be an additional step for the police officers to deal with. Since deeper photogrammetric knowledge is required for camera calibration this can negatively influence the willingness and acceptance to use SfM as the basic reconstruction tool. To sum it up, it can be said that the implemented camera model influences the result of the dense point cloud as undistorted input images are needed for the dense matching. Therefore, a camera model that considers additional distortion parameters should be preferred over a rather simplistic camera model.

#### 3.1.2 Basic characteristics of the software systems

This section shall summarize some basic specifications for each software system (Table 1). The used default settings are underlined. Concerning the implemented camera models, parameters abbreviated with k are radial distortion coefficients and the ones with p are tangential distortion coefficients.

*Table 1: Characteristics of the evaluated software tools*

| | **VisualSFM** | **openMVG** | **Colmap** | **regard3D** | **Agisoft Photoscan** |
|---|---|---|---|---|---|
| **Dense Matching** | PMVS/CMVS | openMVS | integrated (CUDA required) | openMVS | integrated |
| **License** | Freeware (not for commercial purposes) | Mozilla Public License 2 | GNU General Public License v3 | MIT License | Proprietary |
| **Camera Model parameters (in addition to fx, fy, cx, cy)** | k1<br>(one set of parameters for each image) | various models;<br><br>k1,k2,k3 | various models;<br><br>k1 | various models;<br><br>k1 | k1, k2, k3, p1, p2<br>one model for all images |
| **Interface (GUI or CLI)** | GUI + CLI | CLI | GUI+CLI | GUI | GUI+CLI |

| | | | | | |
|---|---|---|---|---|---|
| **Feature Detector** | SIFT | SIFT + AKAZE | SIFT | AKAZE | Proprietary |
| **Geo-referencing** | With external tool after dense matching | During bundle adjustment | With external tool after dense matching | With external tool after dense matching | During bundle adjustment |
| **Hardware specialties** | Video card required | - | CUDA card required | - | - |

### 3.2 Sample Image Data

In order to test the different software tools a set of images was acquired. The set covers an intersection in a densely build urban area (see Fig. 1). Since real photos of an accident scene were not available, we used this setting to simulate a real accident location as good as possible. Totally 61 photos were taken using a Canon EOS 5D Mark III with a sensor size of 5670 x 3840 pixel. The photos were captured circular around the intersection in order to achieve a high overlapping between the images. During the image acquisition step, no additional GCPs were measured.

*Fig. 1: Overview of the location on the intersecton of Hochulstraße and Reichenbachstraße, Dresden. (Map data ©OpenStreetMap contributors, CC-BY-SA)*

### 3.3 Overall Results

All evaluated software tools reconstructed the full set of 61 images. For the reconstruction process a computer with an Intel Core i7@2,8 GHz CPU, 16 GB RAM and a CUDA enabled video card was used. The following table (Table 2) gives an overview of the basic results and processing times:

*Table 2: Processing results*

| | VisualSFM | openMVG | Colmap | regard3D | Agisoft Photoscan |
|---|---|---|---|---|---|
| **Registered Images** | 61 | 61 | 61 | 61 | 61 |
| **Processing Time [min.]** | 6,5 | 25 | 5 | 148 | 5,5 |
| **Nr. Points in Sparse Cloud** | 4456 | 32778 | 5879 | 27484 | 22471 |

The results show that openMVG and regard3D needed more time for the reconstruction than the other tools, since they do not use the video card for processing. Nevertheless, the processing time is passable in respect to the amount of images. They also produce a higher amount of points in the sparse point cloud.

### 3.4 Investigation of Geometric Accuracy of Registered Views

The main interest of the evaluation was put on the investigation of the geometric accuracy of the registered camera stations. Since the accuracy of the estimated camera stations and the camera parameters are crucial for the further processing step of calculating the dense point cloud, the eventually chosen software tool has to provide good and accurate results.

A common procedure to compare the results of a complete SfM pipeline (that means the generated dense point cloud) is to use terrestrial laser scanning (TLS) data as ground truth. In a first step Ground Control Points (GCP) are used to transform the generated point clouds in a common coordinate system. In case of the absence of GCPs the Iterative Closest Point (ICP) algorithm can be used to register the point clouds. After having completed this step, it becomes possible to determine the deviation between the reference and the SfM clouds.

However, our intention was to investigate the results of the actual SfM process. That means that a different approach was necessary. All investigated SfM tools reconstruct the 3D scene in their own local coordinate system. Therefore, the first step is to transform the calculated camera coordinates into one common system. We applied a principal axis transformation to all SfM results in order to have a common reference system as the base for a comparison. By only applying a principal axis transformation on the point clouds, they still have not the same scale in every reconstruction. To overcome this limitation we have selected two out of the 61 cameras and set an approximate distance between these two views. This way it becomes possible to scale all resulting camera views and point clouds the same way.

As a result, all calculated camera views are defined in the same coordinate system. In the next step, deviations in the camera center position are investigated. As reference for the five different bundle adjustment results, the median (out of five) camera center position for all 61 images was calculated. By calculating the median as reference observation, outliers and gross errors are not taken into account as strong as by using the average as the reference value. The median is then used to get the distance error for every image and for every software tool (Fig. 2). The results of VisualSFM show a significant higher error compared to the other systems. Colmap's deviations are also quite high. On the other side, openMVG and regard3D deliver acceptable results. Evidently, Agisoft Photoscan delivers the best results. The main reason for investigating regard3D was that the tool by default uses AKAZE-features. The statistical examination shows that the results are comparable to the results of openMVG.

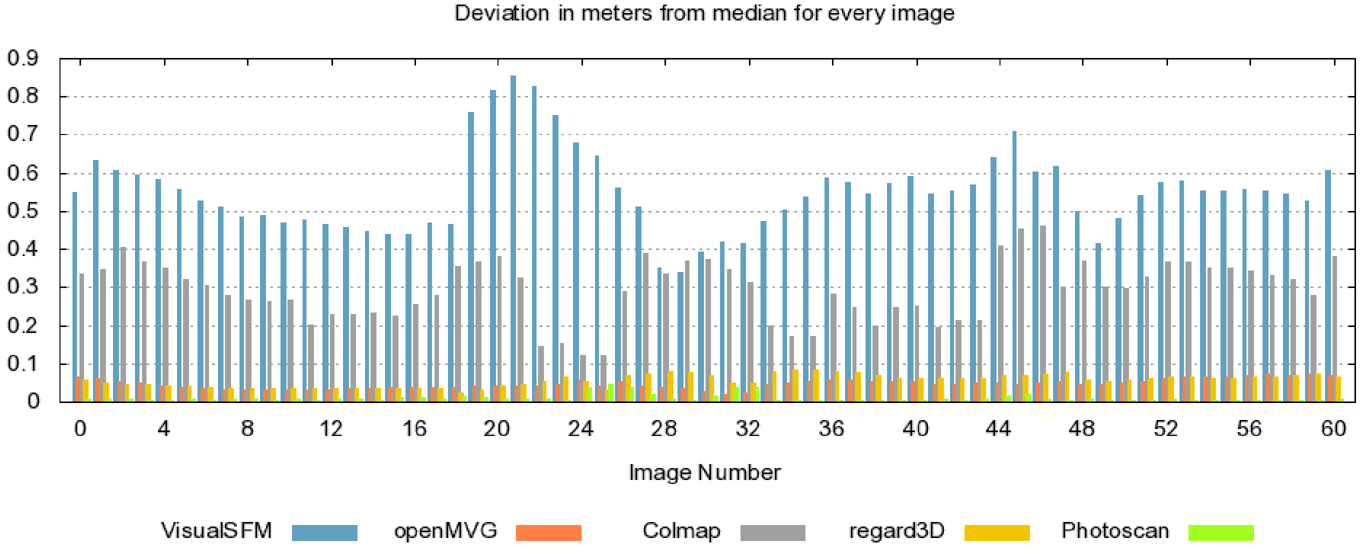


*Fig. 2: Deviaton in meters from median for every image*

## 4. DISCUSSION AND CONCLUSION

The results of section 3.4 show that especially VisualSFM but also Colmap have strong deviations from the median for the calculated camera centers. This might be attributed to their applied feature extraction process. If images with a higher resolution of 3200 pixels are used, both tools resize the images accordingly for faster processing. As a negative effect of the downsizing, less keypoints are computed leading to the presented results.

One problem all software tools have in common is the detection of keypoints on road areas. Asphaltic roads usually look very homogeneous and uniform without a lot of texture. This makes it hard for feature detection algorithms to find keypoints. Fig. 3 shows one of the sample images with the detected keypoints. The red dots indicate the keypoints found by Colmap, the yellow dots are the keypoints calculated by openMVG and the blue dots show the calculated AKAZE-features. It clearly shows that very few keypoints are detected on asphaltic road segments, while the most keypoints are found on objects of the built environment. The figure also demonstrates that there is no big difference between the two feature detection algorithms SIFT and AKAZE. Both methods cannot find many points on the street due to the lack of textured areas.

*Fig. 3: Feature point detection on road. Feature points in red are detected by Colmap in yellow by openMVG and in blue by regard3D*

Concerning the registration process of two images, the best results are achieved when correctly matched feature points are evenly distributed over the whole image. Within the application of traffic accident mapping and hence the need to work with road image data this distribution might not always be guaranteed. When thinking about a reconstruction on a country road where objects from the built environment are missing the task of accurate camera reconstruction can become difficult. Consequently, it has to be investigated in future work, if the basic feature detection and matching algorithms of the used tools can be modified accordingly in order to increase the amount of detected keypoints on country road scenes.

The intention of the conducted investigation was to select an appropriate SfM tool for the accident scene reconstruction. After finishing the study, we agreed on using openMVG for our further research project. An important reason was that openMVG performed quite good regarding the geometric accuracy of the registered views. In addition, it implements more sophisticated camera models compared to for example VisualSFM. openMVG also can be considered as a complete SfM software library additionally providing useful features that go beyond the basic SfM pipeline. Furthermore, the openMVG community features an active development and improvement of the library. Another benefit of openMVG is that it does not require a special hardware / video card as all calculation are executed on the CPU. In fact, this makes the reconstruction process last a bit longer compared to VisualSFM or Colmap but as police departments usually are limited concerning their hardware infrastructure the waiver of GPU-computing might be also an advantage. Since openMVG can be operated completely from its command line interface, it would become possible to implement a customized GUI for the police around the openMVG executables. This way it would be ensured to keep the SfM processing for police authorities as easy as possible.

## ACKNOWLEDGMENTS

The European Social Fund (ESF) and the Free State of Saxony (funding no. 100316427) financially support this research project.

## REFERENCES

Agent, J.R.; Barrett, M.L.; Agent, K.R. (2005): Evaluation of Methods to Limit the Time Taken to Investigate Crash Sites. Kentucky Transportation Center.

Alcantarilla, P. F.; Nuevo, J.; Bartoli, A. (2013): Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. In : British Machine Vision Conf. (BMVC).

Buck, U.; Naether, S.; Räss, B.; Jackowski, C.; Thail, M. J. (2013): Accident or homicide - Virtual crime scene reconstruction using 3D methods. In *Forensic Science Internationl* (225), pp. 75–84.

Cooner, S. A.; Balke, K. N. (2000): Use of Photogrammetry for Investigation of Traffic Incident Scenes. Texas Department of Transportation.

Du, X.; Jin, X.; Zhang, X.; Shen, J.; Hou, X. (2009): Geometry features measurement of traffic accident for reconstruction based on close-range photogrammetry. In *Advances in Engineering Software* 40 (7), pp. 497–505.

Eltner, A.; Schneider, D. (2015): Analysis of Different Methods for 3D Reconstruction of Natural Surfaces from Parallel-Axes UAV Images. In *Photogram Rec* 30 (151), pp. 279–299.

Fischler, M. A.; Bolles, R. C. (1981): Random sample consensus. A paradigm for model fitting with applications to image analysis and automated cartography. In *Commun. ACM* 24 (6), pp. 381–395.

Fraser, C.; Hanley, H.; Cronk, S. (2008): Close-range photogrammetry for accident reconstruction. In *Proceedings of XXI ISPRS congress commission V, WG V*, pp. 125–128.

Kaden, R. (2005): Entwicklung eines digitalen Aufnahmeverfahrens zur räumlichen Erfassung von Verkehrsunfallstellen. Diplomarbeit. Hochschule für Technik und Wirtschaft, Dresden.

Moulon, P.; Monasse, P.; Perrot, R.; Marlet, R. (2017): OpenMVG: Open Multiple View Geometry. In Bertrand Kerautret, Miguel Colom, Pascal Monasse (Eds.): Reproducible Research in Pattern Recognition. 1st ed.: Springer International Publishing.

Pagounis, V.; Tsakiri, M.; Palaskas, S.; Biza, B.; Zaloumi, E. (2006): 3D Laser Scanning for Road Safety and Accident Reconstruction. In : Proceedings of the XXIIIth international FIG congress, pp. 8–13.

Regard3D (2018). Available online at www.regard3d.org.

Sächsisches Staatsministerium des Innern (2016): Verwaltungsvorschrift Verkehrsunfall. VwV Verkehrsunfall, revised 10/7/2016.

Schonberger, J. L.; Frahm, J.-M. (2016): Structure-from-Motion Revisited. In : Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016., pp. 4104–4113.

Stáňa, I.; Tokař, S.; Bucsuházy, K.; Bilík, M. (2017): Comparison of Utilization of Conventional and Advanced Methods for Traffic Accidents Scene Documentation in the Czech Republic. In *Procedia Engineering* 187, pp. 471–476.

Wu, C. (2011): VisualSFM: A Visual Structure from Motion System. Available online at http://ccwu.me/vsfm/.


## BIBLIOGRAPHICAL NOTES

Tim Kaiser, born 1992. Graduated in 2017 as a M.Eng. in Geomatics/Management from the HTW Dresden, University of Applied Sciences. Since then he is a PhD Student at the Faculty of Geoinformation.

Christian Clemen, born 1976. Graduated in 2004 as a Dipl.-Ing. in Surveying from the Technical University of Berlin. From 2004 to 2010 he was an Assistant at the Department of Geodesy and Geoinformation, Technical University of Berlin. He then worked as a GNSS software engineer for Alberding GmbH, Wildau. Since 2013 he has been a full professor at the HTW Dresden, University of Applied Sciences. Within FIG he is national delegate for commission 10 and chairing the JWG BIM.

Robert Kaden, born 1979. Graduated in 2009 as a M.Sc. in Geodesy and Geoinformation Science from the Technical University of Berlin. From 2009 to 2012 he was an Reseach Assistant at the Department of Geodesy and Geoinformation Science, Technical University of Berlin and from 2013 to 2015 at the Chair of Geoinformatics, Munich University of

Technology. He then worked as a office manager for the Surveying Company Kaden in Dresden. Since 2016 he is Postdoc at the HTW Dresden, University of Applied Sciences.

## CONTACT

Tim Kaiser
PhD student, Faculty of Geoinformation
HTW Dresden, University of Applied Sciences
Friedrich-List-Platz 1
D-01069 Dresden, Germany
Tel: +49 351 462-3164
Fax: +49 351 462-2191
E-Mail: tim.kaiser@htw-dresden.de
Web site: http://www.htw-dresden.de/geo

Prof. Dr.-Ing. Christian Clemen
Chair of CAD and Virtual Reality, Faculty of Geoinformation
HTW Dresden, University of Applied Sciences
Friedrich-List-Platz 1
D-01069 Dresden, Germany
Tel: +49 351 462-3199
Fax: +49 351 462-2191
E-Mail: christian.clemen@htw-dresden.de
Web site: http://www.htw-dresden.de/geo

Dr.-Ing. Robert Kaden
Postdoc Geodesy and BIM, Faculty of Geoinformation
HTW Dresden, University of Applied Sciences
Friedrich-List-Platz 1
D-01069 Dresden, Germany
Tel: +49 351 462-2050
Fax: +49 351 462-2191
E-Mail: robert.kaden@htw-dresden.de
Web site: http://www.htw-dresden.de/geo