3 Beyond OSM – Alternative Data Sources and Approaches Enhancing Generation of Road Networks for Traffic and Driving Simulations

Andreas Richter, Hartmut Friedl, Michael Scholz; German Aerospace Center
{Andreas.Richter, Hartmut.Friedl, Michael.Scholz}@DLR.de

Abstract

Often OpenStreetMap data is used in the creation of the anderlying road networks for traffic simulations; sometimes it can be enriched with navigation data. Due to the fact that traffic simulation plays a growing role for the development of driver assistance and automation systems more detailed road networks are necessary because traffic simulations are coupled to driving simulators. This detailed information on lane-level and road infrastructure is not available in OSM or navigation data or modelled in an inappropriate way.
This paper will discuss alternative data sources like cadastral or mobile mapping data, their advantages and disadvantages compared to OSM and navigation data and will give an idea on how inappropriate data could be improved. We will show approaches of how to fuse this data to a road network using the OpenDRIVE format resulting in a suitable output for the use in traffic and driving simulations.

Keywords: traffic and driving simulations, road network generation, data sources

3.1 Introduction

The development of driver assistance and automation systems is still the next big thing in the automotive domain. It is important to differ from other car manufactures but furthermore to shape the future of mobility (1). These systems are currently getting more and more complex due to their growing functionality and connection between other vehicles and to the outer environment of the vehicle. To test all this complexity the driving simulation is an essential tool to reduce time and cost effort. Therefore a driving simulator needs a detailed and highly precise road network often accompanied by an appropriate environment model.
Connecting driver assistance and automation systems to each other and interchanging data with other participants raises the need for testing of applications and systems in traffic simulations with larger road networks and more traffic participants, too. Examples are multimedia systems with global and micro navigation functionality or the upcoming vehicle-2-X communication. Therefore the coupling of driving and traffic simulations is a good idea to combine both testing purposes. Due to the different requirements on the anderlying road network the coupling of driving and traffic simulation shows some issues. For example, positioning of the traffic participants differs due to different ways of calculating the movement. Also different ways of linking the road networks can cause a

varying behaviour in route finding. Challenges in the coupling of a driving simulation with a traffic simulation are described in detail using the example of connecting SUMO with VirtualTestDrive (2).

3.2 State of the Art

3.2.1 Traffic Simulation

OpenStreetMap (OSM) is one of the favourite data sources for road network generation for the traffic simulation SUMO because of its extensive world-wide coverage; it is more or less continuously updated and provided for free. Besides OSM also road networks from other traffic simulators like VISUM, Vissim and MATSim are used (3), which often base on OSM data, too.

OpenStreetMap

All these networks have in common that they describe a logically linked road network with some attribute data such as speed limits or other traffic rules but also integrates interesting information such as the number of lanes and additional road furniture. To understand OSM it's good to know, that geometry and all descriptions are stored in an xml formatted text file. Each feature is described with tags. A tag consist of a key and a value, e.g. highway=residential. In most cases tagging follows guidelines to describe defined features, but in general tagging is free and therefore good and easy for crowd sourcing. To process OSM for other application data is often translated in tables good for reading in relational databases such as PostgreSQL. This is easy for single features but can get complicated for larger data extractions. The following figure 3.1 shows a JOSM screenshot, displaying a complex intersection in Braunschweig, Germany. It shows road edges, lanes including traffic rules, traffic signs, additional road furniture and the description of the way highlighted in red in the included table.

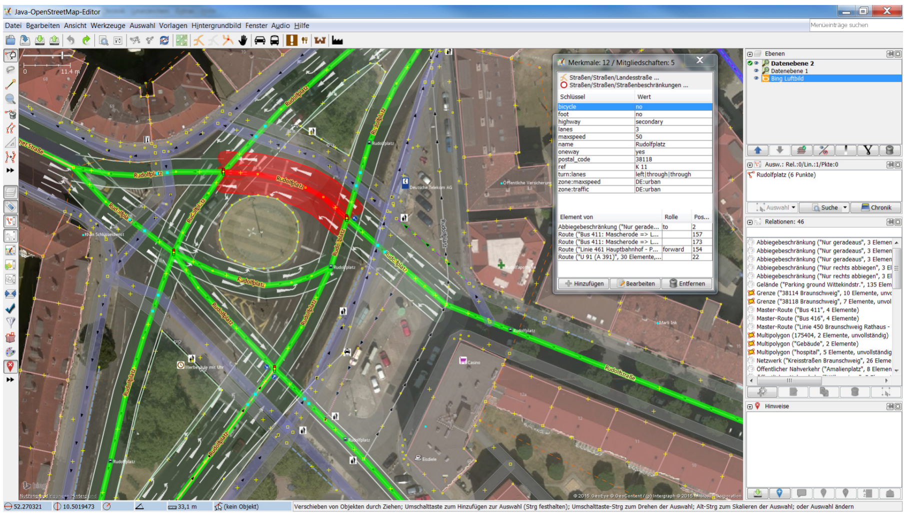

Figure 3.1: complex intersection with information about road edges, lanes including traffic rules, traffic signs and additional road furniture

The amount of additional data about road infrastructure and lanes grew in the last years. Data was added especially for intersections. Nevertheless OSM data has some major disadvantages for driving simulators. The following sections are going more into the technical details of the data. All following statistics were collected with http://taginfo.openstreetmap.org on 24th of February 2016. Crossroads are often simplified for map representation. The above pictured intersection is a good exception due to the street island. Less sophisticated intersections are not modelled in such a precise way as a second intersection in Braunschweig shows in figure 3.2.

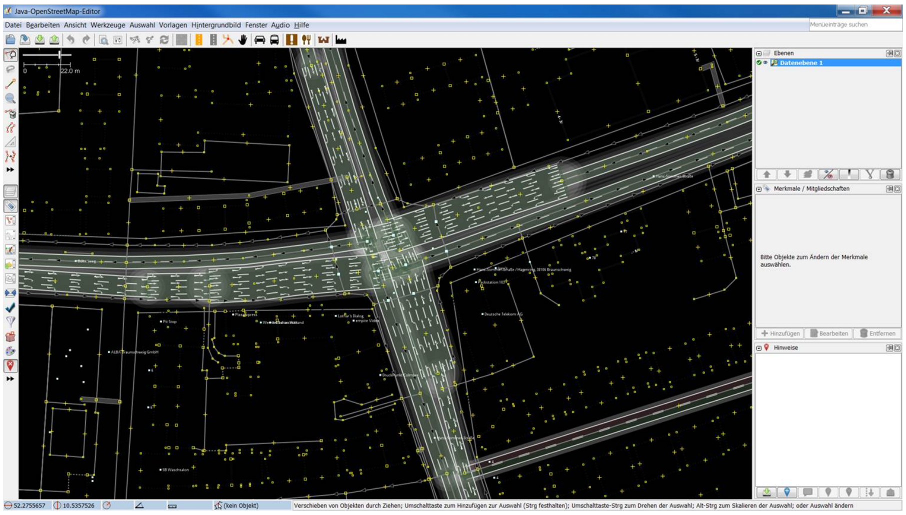

Figure 3.2: larger intersection with simplified layout

In general there is mostly a defined way of tagging and adding additional data to the OSM database but practically there are still different ways of tagging that information. Especially for new kinds of data a common way has to be established first, e.g. proposed for railway infrastructure (4). Concerning the tags for [lanes] or [turns] mostly more than 99% of the tags are correct, but tags like [width] must be cleaned because of information on measurement units being often included. Additionally, it can be confusing how to derive the information from different tags. Tags such as [turn], [turn:lanes], [turn:lanes:forward], [turn:lanes:backward], [turn:lanes:both_ways], [turn:forward], [turn:backward] or, in typically rare but existing exceptions [lanes:turnleft], come up with the layout information of an intersection. The [turn:lanes:forward] and [turn:lanes:backward] tags have the disadvantage that they can only describe the linkage at the beginning and at the end of a way and all intersections in between are not considered. That is a drawback especially for long roads (5).
Moreover the latter tag [lanes:turnleft] follows a different logic, describing the amount of lanes turning left, while the previously mentioned tags are describing the lane turn types like [left|through|through|right][1].
Additionally, typos and careless editing lead to a network that is not usable out of the box, in particular for the road network (5; 6). It has to be cleaned and corrected. The following figure 3.3 shows a complex intersection in Stuttgart, Germany, where some lanes in the north-west were tagged with a future validity date, which has already been reached at the time of browsing. Furthermore, these lanes are connected to nowhere.
A driving simulator depends on the topographical representation of the roads. But just 1% of all highways provide a [width] tag, or 4% of [highways] with a [lane] tag also define a [width] tag. Only 0.03% of all [lanes] provide a [width:lane] tag. Information on connections in-between lanes does not exist but can be estimated/constructed in most cases. There is also no information on detailed road markings (transverse markings, no passing zones) but may be constructed from the [change:lanes] tag. There is not always a need for this tag, so it is hard to say how complete this information is. At least more than 20% of [turn:lanes] are tagged.
There are lots of approaches to detect, enhance or correct locations of road networks like OSM (7). For driving or traffic simulation there is little gain in correcting locations, even though locations in OSM offer sufficient quality. For simulation approaches enhancing topographic data, such as the

[1] http://taginfo.openstreetmap.org/tags/turn%3Alanes%3Aforward=left%7Cthrough%7Cthrough%3Bright

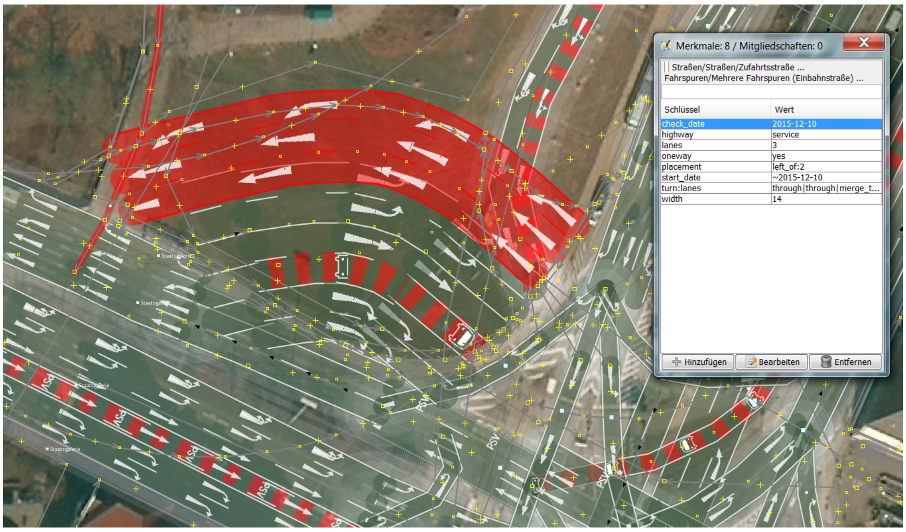

Figure 3.3: intersection with incorrect lane information

width of roads, is more prospering. Road width could be calculated from very detailed town plans including curbstones (8) or from ortho images (9).

Navigation Databases and Cadastral Data

The pre-processing can be minimised by using official data provided by map makers such as HERE or TomTom and already prepared road networks from other traffic simulation tools. The standard navigation data lacks information about the amount of lanes and lane width, too. The following figure 3.4 shows a HERE road network with underlying road topography loaded as Shapefile. It clearly shows the mismatch of lane widenings, which leads to a wrong representation of the capacity of turning lanes.

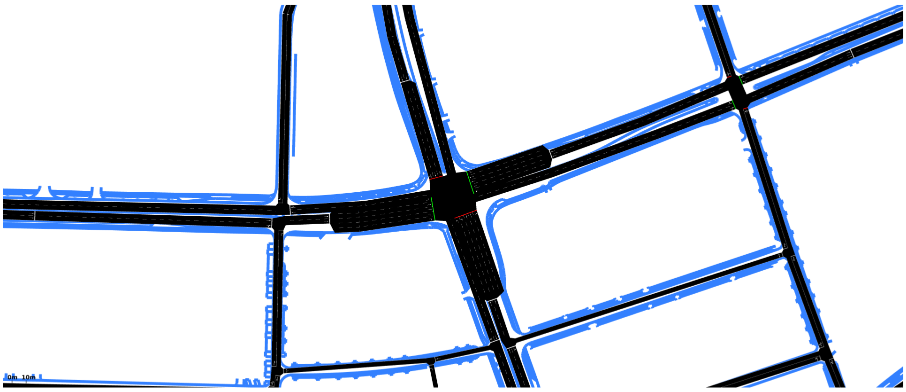

Figure 3.4: road network based on navigation data compared with topography data

But there exists so called High Definition or ADAS data, which is enriched with elevation profiles and road curvatures as additional information for assistance and automation systems. Actually they contain recently tagged information about lane widening and linkage. But again, most of these data lack topographical information about width of the lanes (5). For example, in the case of a lane widening for a left-turning lane, it is of importance at which point this new lane offers enough space for vehicles not to block the neighbouring lane any more.
An additional drawback is that these navigation data sets have to be purchased and sometimes are not available nationwide.

Official German cadastral data DLM5 does contain information about the lane width for some road categories. The following figure 3.5 shows the same intersection as above, highlighting the roads with lane width information. Only major roads are represented in detail, but there is also no information about lane widening available and intersections are generalized.

Figure 3.5: highlighted roads with information about lane width in cadastral data

3.2.2 Driving Simulation

Data bases for driving simulations are getting bigger and bigger due to their usage in model- or system-in-the-loop tests. Additionally, more and more real world data is needed in the driving simulator to close the gap between simulated and real world test (10).
A road network for a driving simulator does not only contain the logical linkage of the roads and their lanes. The major part is its topographical description of the geometrical road appearance. A third component is the description of road infrastructure such as road signs, traffic lights and important road objects. Therefore specified formats are used, for instance OpenDRIVE (11), which evolved as a de-facto standard in the driving simulator domain.
Currently, suitable data is poorly available. Most of such OpenDRIVE data bases are manually built based on areal images as guidance or even totally artificially built. Both approaches are time-consuming processes. A mobile mapping road survey is able to deliver exactly the needed data in a very accurate way. The following figure 3.6 depicts the same intersection as in the images above, modelled with the help of mobile mapping data.

However, if the area of interest is growing and contains many complex situations, the surveying of roads can get very expensive and time consuming as well. The use of cadastral data can fill the gap because it can deliver a good coverage, e.g. city-wide, but sometimes does also lack interesting data for lane-level description or information about the road infrastructure (12).

3.3 Road Network Generation

The project "Virtual World" builds upon the experience of the "SimWorld$^{\text{URBAN}}$" project, which fused survey data with cadastral and land use data as well as areal images to generate a detailed road network (13). To process the cadastral data a computer graphics approach combined with geographic information systems functionality was used to generate the basic information, such as reference lines and lane borders, and to add road infrastructure.
The following paragraph gives a brief overview about how a road network can be generated. A more

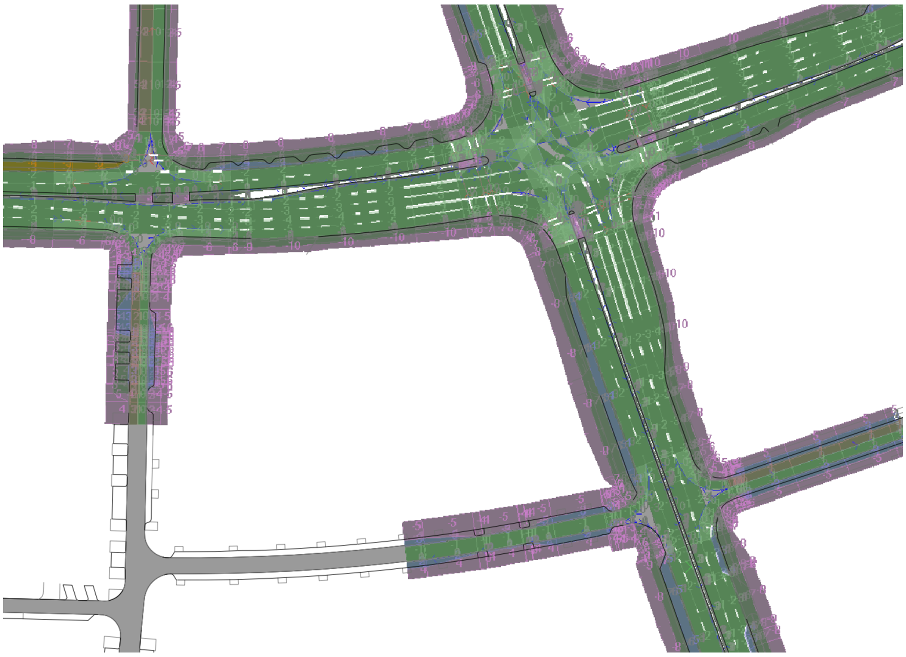

Figure 3.6: intersection based on mobile mapping data

detailed description can be found in (14).
As starting point a geo-dataset of curb stones from cadastral street topography is used. The data comes with geometrical inconsistencies and has to be cleaned and integrated in the first step. At this point there is also no information about driving lane properties present as this dataset consists of plain linear geometries without any semantics of the road surface area other than its boundary. The whole, cleaned line network is polygonised to determine potential asphalt zones for motorised traffic. The appropriate area can be selected with a combination of cadastral road centrelines and centrelines from OSM/HERE. Reference lines serve as anchoring of all further street information such as lanes, signals, and elevation. A reference line is ideally located in the centre of a road. It describes the road's course in detail as a mathematical function. From the polygonised asphalt zones such reference lines are derived as skeletons of the polygons. The outcome is a linked line network which describes the course topographically correct. A potential road junction exists where the skeletonised reference lines join in one or multiple neighbouring points. The linkage in such junction areas is still geometrically inappropriate and does not describe the real bend of curves. Therefore line segments within these junction areas are removed and exchanged by more realistic arc-, spline-, and spiral-based connection paths. For broad streets with centre islands and complex lane layout the reference lines are translated onto the centre street islands for correct lane reference. Information about these centre islands is also gripped from the polygon dataset. For single-lane roads an appropriate translation of the reference line is performed as well. One-way roads can be considered also if appropriate meta-information is extracted from OSM or HERE, for example. Each reference line is associated a polynomial elevation model derived from digital raster-based elevation data. Such digital terrain models are available from photogrammetric or radar-based remote sensing or can be generated from laser scanning. Based on the corrected reference lines the road lanes are generated. A rule-set interpolates the count and width of lanes incorporating distances between reference lines and asphalt borders and width thresholds. At junctions logical relationships between incident lanes are added generically and can be corrected with metadata from OSM or HERE again. Those rudimentarily generated lanes already serve well for basic representation of the street network and can automatically be furnished with markings. Also grass borders, bike lanes and pedestrian walks could be estimated from cadastral data to enrich the road network.
If there is more information available in the input data, e.g. lane markings, the approach works better

and generates a more realistic road network (8; 15). A similar approach can be used to generate the environment, e.g. the city model (14; 16) or road infrastructure. The following figure 3.7 shows the same intersection as before in a fully equipped driving simulator environment.

Figure 3.7: 3D driving simulator environment

As descripted the road network generation can be improved by introducing OSM or HERE data to support the generation of road axes and to enrich these roads with road furniture. Road infrastructure, such as signs and traffic lights, is added with the help of extracted data from appropriate cadastral sources. At the end the road network is generated in the OpenDRIVE format.
Even though a lot of information, such a city models, road topography, road infrastructure, is already available in a digital way, in most cases this data is not machine-readable and thus cannot be processed in a proper way. Additionally, these data can also lack accuracy due to careless editing and different ways of gathering of the raw data, such as digitalisation of analogue maps or processing of aerial surveys.
Nevertheless the DLR is working together with major German car manufactures on guidelines to survey road networks in a simplified but enhanced way to meet cadastral and simulation requirements. The guidelines will also cover the gathering of attribute data, such as the relation and the type of lines and points, in order to support a later transformation of these data into specific driving simulator or navigation formats like Navigation Data Standard (NDS). In the future cadastral data should be streamlined for additional usage and not just for digital visualisation. Therefore a growing database should be available to generate detailed road networks out of cadastral data.

3.4 Dual Use

SUMO's netconvert application is able to read OpenDRIVE and transform it into SUMO's internal road format (3). It takes over all necessary information like lane linkages and infrastructure, such as traffic lights (2). Therefore the OpenDRIVE network is fully functional and can be used by SUMO and the SUMO suite tools. The following figure 3.8 shows the OpenDRIVE road network from the two previous pictures loaded in SUMO with underlying road topography loaded as Shapefile.
The representation is more or less equal and differs only because of the less detailed internal road description format of SUMO. Most important is that all identifiers and connections stay unchanged and make a coupling between traffic and driving simulation possible. The driving simulator and the traffic simulator can work on the same road network without the need to translate road network elements. Still the positioning of vehicles in the driving simulator has to be interpolated with the

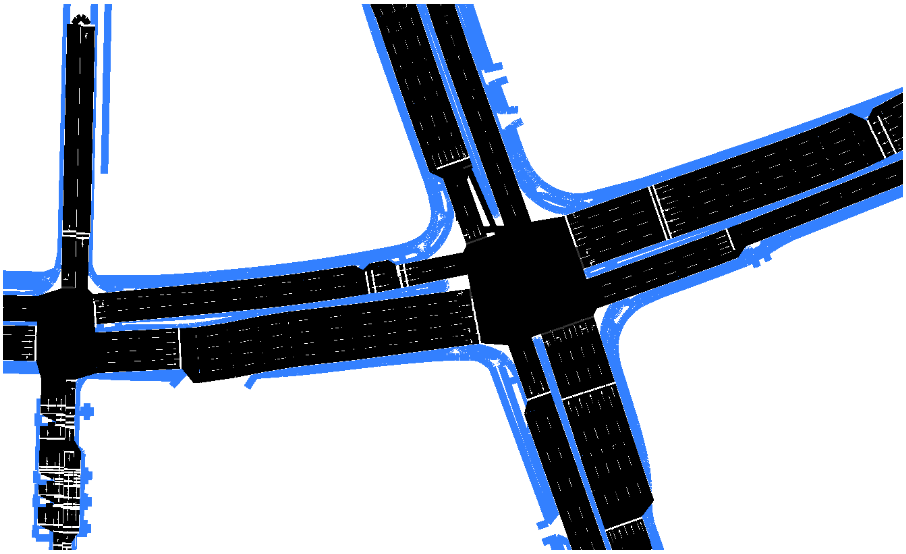

Figure 3.8: OpenDRIVE road network compared with topography data

help of the road identifiers.
As side benefit SUMO's netconvert can also export OpenDRIVE networks. Therefore it is possible to create a huge road network database for driving simulators. Nevertheless, it will show all the drawbacks of SUMO's internal road network format, for instance, the mathematical road curvature being discretised as line strings.

3.5 Summary

We showed that it is possible to generate road networks based on different kinds of geodata. In general the fusion of heterogeneous sources is necessary to reach the goal of a precise road model. OpenStreetMap can be such a data source but does not have to remain the only one. However, OSM currently cannot fulfil all requirements of upcoming use cases. For example, for the test of emergency vehicle rescue lanes or continuous lane change, it is necessary to model the width change of lanes in a proper way. For the coupling of driving simulation and traffic simulators more detailed information will be required. OSM cannot provide this information, but it can be derived from cadastral data. To model this information in a suitable manner DLR will announce guidelines for a generally accepted way of representing all required information.

3.6 References

[1] Maurer, Markus and Gerdes, J. Christian and Lenz, Barbara and Winner, Hermann (Hrsg.), (2015) *Autonomes Fahren. Technische, rechtliche und gesellschaftliche Aspekte.*, DOI 10.1007/978-3-662-45854-9

[2] Krajzewicz, Daniel and Richter, Andreas and Behrisch, Michael and Erdmann, Jakob (2013) *Abbildung des Umgebungsverkehrs in einem Fahrsimulator.* VDI Verlag. 4. Berliner Fachtagung Fahrermodellierung, 13.-14. June 2013, Berlin, Germany. ISBN 978-3-18-303522-9. ISSN 1439-958X

[3] Krajzewicz, Daniel and Erdmann, Jakob and Behrisch, Michael and Bieker, Laura (2012) *Recent Development and Applications of SUMO - Simulation of Urban MObility.* International Journal On Advances in Systems and Measurements, 5 (3&4), (pp. 128-138). ISSN 1942-261x.

[4] Rahmig, Christian and Simon, Augusto (2014) *Extracting Topology and Geometry Information from OpenStreetMap Data for Digital Maps for Railway Applications.* 10th ITS European Congress 2014, 16.-19. June 2014, Helsinki, Finland.

[5] Sämann, Robert (2014) *Bestimmung einer Bewertungsmetrik zum Vergleich digitaler Straßennetze für Verkehrsflusssimulation und Routing von Einsatzkräften.* master thesis, Leibniz University Hanover.

[6] Krumnow, Mario (2014) *Modeling Mobility with Open Data* http://www.sumo.dlr.de /2014/Presentation_Keynote_Open%20data%202014_MarioKrumnow.pdf SUMO - Simulation of Urban Mobility 2014, 15.-16. May 2014, Berlin, Germany.

[7] Butenuth, Matthias (2010) *Geometric refinement of road networks using network snakes and SAR images.* In Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International (pp. 449-452). IEEE.

[8] Richter, Andreas (2015) *Geodaten: Vom Kataster in die Fahrsimulation.* 1. Fachtagung Fahrsimulatoren, 07. July 2015, Stuttgart.

[9] Mattyus, Gellert and Wang, Shenlong and Fidler, Sanja and Urtasun, Raquel (2015) *Enhancing Road Maps by Parsing Aerial Images Aroand the World.* Computer Vision (ICCV), International Conference (pp. 1689-1697). IEEE. International Conference on Computer Vision, 13-16 Dez. 2015, Santiago de Chile, Chile.

[10] Richter, Andreas and Fischer, Martin and Frankiewicz, Tobias and Schnieder, Lars and Köster, Frank (2015) *Reducing the gap between simulated and real life environments by introducing high-precision data.* Driving Simulator Conference 2015 Europe, 16.-18. Sep. 2015, Tübingen, Germany. ISBN 978-3-9813099-3-5.

[11] *OpenDRIVE® Format Specification, Rev. 1.4 Issue H* (2015) http://www.opendrive.org/docs/OpenDRIVEFormatSpecRev1.4H.pdf

[12] Friedl, Hartmut and Richter, Andreas (2012) *Fusion heterogener Geodaten zur Erstellung realer 3D-Welten am Beispiel einer Fahrsimulation.* Shaker. GEOINFORMATIK 2012, 28.-30. March 2012, Braunschweig. ISBN 978-3-8440-0888-3. ISSN 1618-1034

[13] Richter, Andreas and Friedl, Hartmut and Guraj, Vitalij and Ruppert, Thomas and Köster, Frank (2011) *Developing a toolchain for providing automatically highly accurate 3D database.* Verlag der Bauhaus-Universität Weimar. ConVR2011, 03.-04. Nov. 2011, Weimar, Germany. ISBN 9783860684580.

[14] Richter, Andreas and Scholz, Michael and Friedl, Hartmut and Ruppert, Thomas and Köster, Frank (2016) *Challenges and Experiences in Using Heterogeneous, Geo-Referenced Data for Automatic Creation of Driving Simulator Environments.* Transactions of The Society for Modeling and Simulation International, Special Issue Driving Simulation 2016. Prepublished 08. April 2016, DOI: 10.1177/0123456789123456

[15] Richter, Andreas (2015) *Master the Challenge of Generating OpenDRIVE Road Networks based on Real World Data.* 8. OpenDRIVE User Meeting, 15. Oct. 2015, Stuttgart, Germany.

[16] Richter, Andreas and Friedl, Hartmut (2015) *Parametrisierte Stadtmodelle für Fahrsimulatoren.* GIS Talk 2015, 19.-21. May 2015, Unterschleißheim, Germany.