

# IMAGE-BASED REALITY-CAPTURING AND 3D MODELLING FOR THE CREATION OF VR CYCLING SIMULATIONS

W. Wahbeh, M. Ammann, S. Nebiker, M. van Eggermond, A. Erath

School of Architecture, Civil Engineering and Geomatics  
FHNW University of Applied Sciences and Arts Northwestern Switzerland, MuttENZ, Switzerland -  
(wissam.wahbeh, manuela.ammann, stephan.nebiker, michael.vaneggermond, alexander.erath)@fhnw.ch

## Commission IV, WG 9

**KEY WORDS:** Mobile Mapping, Biking Simulation, Urban Mobility, Virtual Reality, 3D Modelling, Urban Scene Reconstruction.

### ABSTRACT:

With this paper, we present a novel approach for efficiently creating reality-based, high-fidelity urban 3D models for interactive VR cycling simulations. The foundation of these 3D models is accurately georeferenced street-level imagery, which can be captured using vehicle-based or portable mapping platforms. Depending on the desired type of urban model, the street-level imagery is either used for semi-automatically texturing an existing city model or for automatically creating textured 3D meshes from multi-view reconstructions using commercial off-the-shelf software. The resulting textured urban 3D model is then integrated with a real-time traffic simulation solution to create a VR framework based on the Unity game engine. Subsequently, the resulting urban scenes and different planning scenarios can be explored on a physical cycling simulator using a VR helmet or viewed as a 360-degree or conventional video. In addition, the VR environment can be used for augmented reality applications, e.g., mobile augmented reality maps. We apply this framework to a case study in the city of Berne to illustrate design variants of new cycling infrastructure at a major traffic junction to collect feedback from practitioners about the potential for practical applications in planning processes.

## 1. INTRODUCTION

Creating reality-based virtual environments of complex urban spaces is a challenging and complex task. This is particularly the case if Virtual Reality (VR) scenes – as in the case of our cycling simulation scenario – are firstly meant to communicate future urban planning scenarios to the local population and politicians and are secondly intended to aid the political decision-making process. In these cases, VR scenes support the credibility of the depicted planning scenarios. A special requirement of VR environments for cyclists and pedestrians is the necessity of an eye-level perspective and for a high-definition representation of the environment of the walking or cycling paths. The spatial immersion allows for cognitive and emotional awareness of scale, depth and movement. Additionally, it is possible to account for the passing of time while being spatially immersed. It is possible to create dynamic environments by either moving objects in the environment, or by moving the individual. This distinguishes these VR environments from the prevailing bird's-eye-view perspective of current state-of-the-art mesh-based 3D urban models.

The main contribution of this paper is an end-to-end solution for creating and exploiting professional reality-based VR environments of complex streetscapes in combination with a realistic multimodal traffic simulation. We outline a novel image-based 3D reality capturing process a) by using street-level imagery from in-house developed backpack- and vehicle-based multi-camera solutions and b) by providing 3D imagery with georeferencing accuracies at the dm level or better. We further document an image-based 3D reconstruction of large and complex urban scenes using commercial off-the-shelf multi-view stereo software and ground-based street-level imagery – as opposed to the prevailing reconstructions from airborne imagery. And we subsequently propose a workflow for integrating georeferenced street-level imagery, derived mesh-based 3D

models, existing geospatial 2D and 3D data – such as a city model – with traffic assets into an interactive 3D scene suitable for real-time VR applications. Finally, we demonstrate the exploitation of the VR model by using a physical VR cycling simulator and by generating 360-degree videos to highlight the advantages and differences of various street design scenarios.



**Figure 1.** Biking Simulator: left: Bike with motion sensors and VR glasses (Photo source: Tobias Wootton); right: the simulator in action and the view during the test for the presented use case.

This interdisciplinary research brought together experts from photogrammetry / computer vision, 3D modelling / architecture and traffic simulation with the goal to advance the process of creating and exploiting high-fidelity reality-based 3D environments in VR and applying these environments in transport and urban planning scenarios.

## 2. RELATED WORKS

### 2.1 Image-based Reality Capturing and Reconstruction

Over the last decade there has been an abundance of research on urban reconstruction using data from different sensors and platforms. Musialski et al. (2013) provide a comprehensive overview of typical reconstruction approaches, input data and challenges. They point out that airborne data is more suitable for coarse building models and ground-based data is more useful for individual buildings and facade details. In their recent review, Wang et al. (2018) provide an overview of research in urban reconstruction from LiDAR point clouds. An inherent problem of LiDAR-based reconstruction – in contrast to image-based approaches – is the spatial and temporal coherence of model geometry and texture (Nebiker et al., 2015) since they are derived from different and spatially disparate sensors.

Our research on large-scale reconstruction of urban scenes from street-level imagery benefits from a long series of related work. Some of the computational foundations for reconstructing large-scale scenes were laid in earlier works on dense 3D reconstruction from aerial and close-range imagery (Lafarge et al., 2013; Rothermel et al., 2012; Vu et al., 2012). Further research focused on urban reconstructions from oblique imagery (Cavegn et al., 2014; Haala et al., 2015) and from UAV imagery (Strech et al., 2015), which brought along the support for large depth ranges and for fisheye sensor models. Many of these developments have since become part of commercial multi-view matching and 3D reconstruction software such as nFrame SURE, Bentley ContextCapture, or Pix4D products. A recent project closely related to our work is the Stuttgart City Walk by Schmohl et al. (2020). They also aim to create a virtual city model for interactive exploration in a VR framework. However, they use airborne imagery for their scene reconstruction process, which puts some limitations on the resolution and level of detail of the resulting street space model.

In our ambition of obtaining highly detailed and accurate reconstructions of urban environments for cycling simulations, we rely on high-quality street-level imagery. The earlier developments in image-based mobile mapping and the potential of accurately georeferenced (3D) imagery are discussed in (Nebiker et al., 2015). Among the latest developments in image-based portable mapping are the works by Blaser et al. (2018), who introduce the high-performance mobile mapping backpack BIMAGE, originally designed for indoor use. To obtain georeferencing accuracies at the sub-decimetre accuracy level in challenging environments, new image-based georeferencing methods have recently been presented (Cavegn, 2020). Blaser et al. (2020) demonstrated that cm-level accuracies can be achieved using a combination of high-end image-based portable mapping systems with state-of-the-art georeferencing techniques – even in challenging urban and forested areas.

### 2.2 3D Modelling

VR applications in the construction field, heritage and city planning are varied. Some of them use models which capture the reality (Bekele et al., 2018), and some of them are based on models for new projects which visualise imagined spaces and structures (Wen and Gheisari, 2020). There is a huge difference

between the work pipelines for those use cases to obtain the VR model. In this study, we combine the two concepts. We have also previously investigated some aspects of direct urban modelling from 360-degree panoramic imagery in which “reality capturing” and “3D modelling” are conceptually combined (Calvano and Wahbeh, 2014a). Moreover, we investigated how project data and images of the city can be combined to visualise unrealised projects in the urban context through 360-degree renderings (Calvano and Wahbeh, 2014b).

### 2.3 Integration of traffic simulation in Virtual Reality

Traffic simulations set out to model detailed vehicle and pedestrian movement to evaluate the performance of transport facilities, varying from highways, intersections to airports and train stations. Simulations model the position of each vehicle or pedestrian every tenth of second and use well-established car-following and pedestrian interaction models. Practitioners worldwide apply microsimulations with software from different manufacturers. Examples include PTV Vissim, Aimsun and the open-source software SUMO (Lopez et al., 2018). While initial traffic microsimulations included simple two-dimensional visualizations for the validation and communication of simulation results, state-of-art microsimulations allow for 3D visualization of traffic flows. However, these visualizations do not result in photorealistic virtual environments for virtual reality purposes, nor were they intended to. To this end, efforts have been made to integrate traffic microsimulation with game engines to create realistic virtual environments.

Nazemi et al. (2021) used trajectories resulting from a traffic microsimulation in virtual reality-based surveys investigating alternative street design options to accommodate cycling. One shortcoming of this study is that there is no interaction between the player (avatar) and the traffic microsimulation.

Kaths et al. (2019) seek to overcome this shortcoming by allowing for the interaction of the first person with simulated vehicles using SUMO. Recent versions of PTV Vissim do allow for the communication between the game engine Unity and the simulation.

Given the availability of traffic microsimulations within authorities and planning agencies, as well as the familiarity with traffic microsimulation, using traffic microsimulation models and their well-documented car-following models is preferred by practitioners. Nevertheless, it is possible to directly model pedestrians behaviour in game engines (e.g. Unity, 2021) as well as model cars. One example can be found in the software VIRE (Farooq et al., 2018) that modelled the interaction between pedestrians and vehicles.

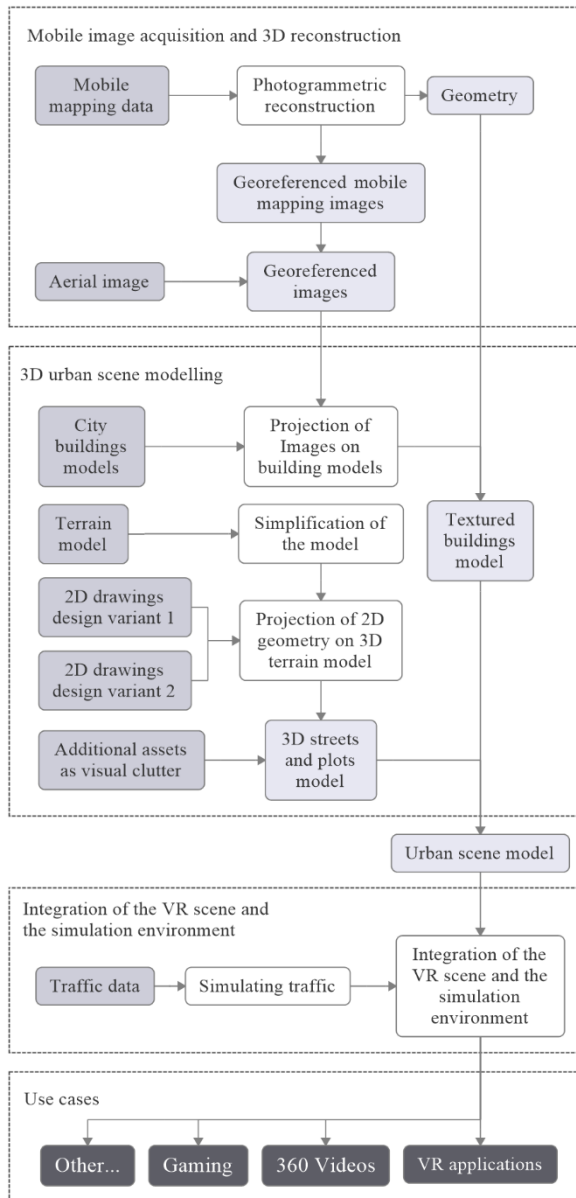
## 3. METHODOLOGY

In our research, we create and investigate two different types of urban models: a hybrid model combining the geometry and semantics of multiple geodata sources, such as existing city models, terrain models and 2D GIS data, and a largely dense mesh-based 3D model reconstructed from street-level imagery and enriched with additional geodata.

We combine image-based automated urban reconstruction from street-level imagery and interactive modelling techniques using various geodata sources with multimodal traffic data and assets to create reality-based VR environments for cycling simulation. Georeferenced street-level imagery from portable and vehicle-based systems plays a vital role in our process and is used for multiple purposes: firstly, for the semi-automated texturing of the hybrid urban model; secondly, for the automated 3D reconstruction of the mesh-based urban model; and thirdly, for the interactive capturing of geospatial objects such as traffic

infrastructure assets (e.g., streetlights, traffic signs), and street furniture, etc., as additional elements for the VR scene. Our workflow for creating VR cycling simulation environments is illustrated in Figure 2. It encompasses the following main steps:

- 1) Mobile image acquisition and 3D reconstruction
- 2) 3D Urban scene modelling
- 3) Creation of traffic model and integration of the VR scene



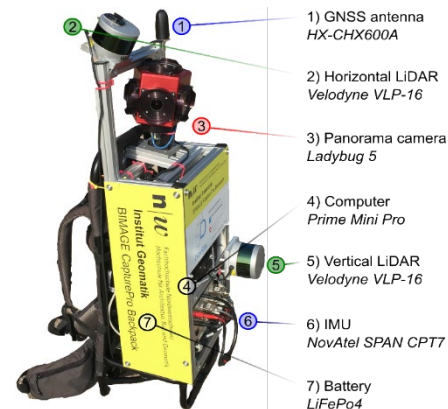
**Figure 2.** Representation of the main activities and work packages as workflow to obtain the VR simulation environment

### 3.1 Mobile Image Acquisition and 3D Reconstruction

#### 3.1.1 Acquisition of street-level imagery:

The environment for a bicycle simulator is best reconstructed or modelled with images acquired from a similar position to that of the bicyclist. Thus, street-level imagery – captured either using a vehicle-based or a portable mobile mapping system (MMS) such as our BIMAGE Backpack (Blaser et al., 2020, 2018) shown in figure 3 – is ideally suited. The BIMAGE Backpack includes state-of-the-art and high-end sensors, such as the GNSS- and IMU-based navigation unit, NovAtel SPAN CPT7, with tactical grade performance, two multi-beam LiDAR scanners Velodyne VLP-16, as well as the multi-head panoramic camera FLIR

Ladybug 5. Furthermore, the system supports precise sensor synchronization and is accurately calibrated using state-of-the-art calibration techniques.



**Figure 3.** BIMAGE Backpack (Blaser et al., 2020).

In the case of complex urban scenarios, a combination of vehicle-based and portable MMS is advisable to provide complete coverage and reconstruction. Ideally, the street-level imagery would be complemented by imagery acquired with a UAV, thus capturing the scene from every angle. The additional capturing of treetops, roofs and balconies from a bird's-eye-view perspective contributes to a complete 3D reconstruction.

#### 3.1.2 Georeferencing

Most MMS, including the BIMAGE Backpack, support direct georeferencing based on GNSS and IMU measurements. In our case, we apply advanced georeferencing either based on LiDAR SLAM or image-based techniques. Blaser et al. (2020) evaluated the georeferencing performance of the BIMAGE Backpack in a city centre and demonstrated that georeferencing accuracies of a few cm can be achieved using state-of-the-art image-based georeferencing. This is important for our case, where the imagery is to be used for automatic texturing of 3D objects, on the one hand, and where the reconstructed scene is to be combined with other geodata, on the other hand.

In our case, we either use the commercial Structure from Motion software Metashape from Agisoft or COLMAP (Schonberger and Frahm, 2016) for image-based georeferencing. One of the reasons is their support for camera rig constraints in the bundle adjustment which ensures more accurate and robust results with multi-head panoramic cameras. If the images have partial view obstructions, e.g., caused by the frame of the backpack, creating a mask before the images are processed is helpful for further work like aligning or reconstructing. The masks are created per camera and are saved in the transparency channel. Once all images have been successfully georeferenced, they can be used for the subsequent 3D reconstruction or in the texturing process described in Section 3.2.3.

#### 3.1.3 3D Reconstruction

The reconstruction of a 3D city model includes all oriented images to cover the entire scene. In our research, the large-scale 3D reconstruction process is performed and evaluated using two commercial off-the-shelf software packages: ContextCapture by Bentley and SURE by nFrames. They both use the image orientations from the image-based georeferencing process in Metashape, and then they compute the tie points. The masks are used in this processing as well to eliminate interfering objects. With software settings for geometric accuracy of one pixel for the reconstruction, the geometric accuracy is between 0.07 and 5.5 cm per pixel.

The reconstructed 3D urban model can subsequently be used to represent the ‘as-is’ urban scene. If the reconstructed model is to be combined with a planned bicycle route and street environment, only the buildings are used.

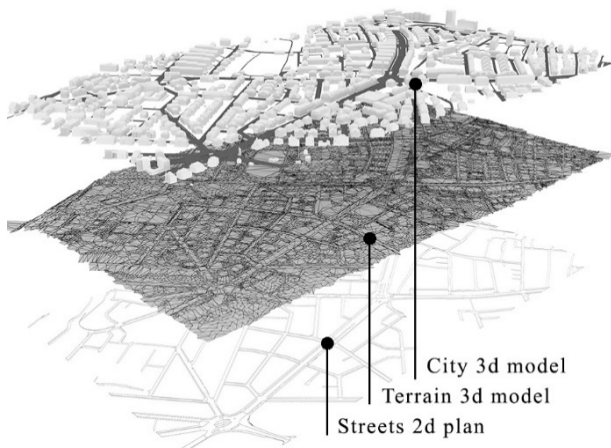
### 3.2 3D Urban Scene Modelling

#### 3.2.1 Collection and Geometric Transformation of Existing 3D and 2D Geodata

Normally, this includes several types: 3D dense meshes such as terrain models (DTM), simplified 3D meshes of structures from photogrammetric reconstructions, and 2D data representing design and project details of buildings, road networks and infrastructures. Using different geodata requires the correct georeferencing. Moreover, considering that 3D modelling software are sensitive in accuracy to the size of the coordinates used, the use of small coordinates is preferable. So, the translation of models into local coordinate systems is essential. Accordingly, the definition of the local origin and the definition of the translation parameters is one of the first project standards to be defined.

#### 3.2.2 Semi-Automated Urban 3D Modelling

In general, two main types of models exist for urban areas. On the one hand, unstructured 3D models with explicit and non-continuous object descriptions. Typically, they are relatively complex 3D meshes (Blaha et al., 2016). Such polygonal reconstructions are normally noisy and, in some cases, memory intensive due to their high number of sub objects (faces or vertices). On the other hand, project drawings which are structured with implicit descriptions. They are typically 2D plans represented by polylines and splines (Figure 4).



**Figure 4.** to obtain the 3D model we combined mainly three data collections: city 3D model (only buildings), terrain model, and streets 2d plans of two variants (existing and proposed solutions).

Digital terrain models (DTM) are complex models and consist of a highly dense mesh. This results in difficulty in further modelling and texturing works. A simplification of the terrain model is important for two reasons: first, to obtain a low-poly model; and second, to convert it to a geometrically and semantically structured model as implicit representation of the streets, pavements and building plots with simple surfaces ready to be textured and to receive the projection of the 2D drawings. The transformation of 2D project drawings into a 3D environment is possible through this spatial projection. In computer graphics, we can differentiate two main representation technologies of surfaces and solids applied to construction models, namely Non-Uniform Rational B-Splines

(NURBS) and polygonal representation (Meshes) (Wang et al., 2018). NURBS is a precise mathematical representation of surfaces and curves which, unlike mesh, offers a continuous description of geometry and many other advantages (Piegl, 1991). This makes the NURBS representation an ideal solution, which provides reliable results for mathematical operations such as Boolean operations and projections. To project the curves and surfaces in order to obtain an automatic transformation from 2D to 3D objects two bases must be prepared:

- The simplified streets surface as mathematical representation (NURBS).
- 2D closed boundaries of streets surfaces, pavements, bike paths, building plots and all the details of street signs on the pavement were created as closed polylines.

Projections could be executed automatically by algorithms. This automation is helpful to save modelling time during the project. The algorithm executes the projection by extruding the boundaries along z axis creating volumes; then intersecting the volumes with the simplified street surface generating new 3D surfaces. Some newly generated surfaces could be extruded afterwards to rise volumes from the street-level. For instance, the sidewalks and beets.

Finally, individual projected layers, defined by object categories and scenarios, are exported in an exchange model format such as FBX or OBJ to be imported into Unity.

#### 3.2.3 Texturing of the 3D city model

Our strategy consists of combining existing city models and georeferenced photos as textures to optimise the model. i.e., to achieve a low-complexity model with a great visual impact due to photorealistic textures. For this purpose, street-level imagery with high georeferencing accuracy is needed.

After processing the images to get the photogrammetric reconstruction which produce high poly meshes and oriented images. The high poly meshes of the buildings in the photogrammetric project are to be replaced by the simplified existing city models. Then the georeferenced images are to be projected on them to create the texture. In our experiments with fully automatic texturing, the very high density of overlapping images from the MMS backpack proved to be very challenging. It turned out to be favourable to extract the textures from only few images with optimal object coverage and where possible a perpendicular orientation to the building facade so that balconies, or other protruding elements of the building are visible from the front.

### 3.3 Creation of the Traffic Model and Integration of the VR Scene and the Simulation Environment

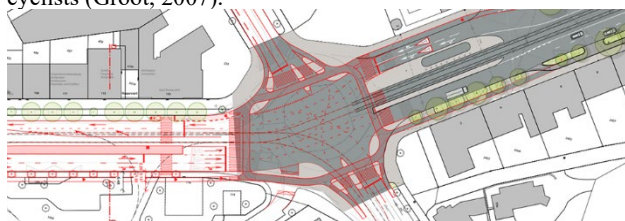
For this research, use was made of the software PTV Vissim and the game engine Unity. Vehicles in PTV Vissim traverse links; pedestrians walk along pedestrian areas. Interaction between vehicles and pedestrians is possible at crossings. PTV Vissim offers the possibility to export vehicle and pedestrian trajectories for a pre-defined step and pre-defined timeframe. Additionally, it is possible to export traffic signal timings. While it is possible to account for height differences in traffic microsimulations to model possible differences in acceleration / deceleration, our traffic simulation model did not account for height differences. Instead, trajectories of vehicles and pedestrians are projected on to a terrain model that excludes objects as trees, traffic lights and lamp posts. It would be possible to perform this projection in real-time, this pre-processing step reserves computational resources for visualization. These scripts have been published and are available open-source software. The avatar of the player (i.e. the individual immersed in VR) is projected onto the terrain model in real-time. In the traffic simulation, each vehicle belongs to a

vehicle class (e.g. car, bus, articulated bus, tram, lorry) or pedestrian class (male, female). Vehicle assets and pedestrian assets are assigned by the game engine on the vehicle resp. pedestrian class defined in the simulation. To provide a realistic environment and full immersion, vehicles are equipped with sound, dependent on the position of the player and each vehicle, rotating wheels and braking lights. Movement through the virtual environment is provided with a cycling simulator. The first iteration of this cycling simulator is documented by Schramka et al. (2017) and has been subsequently used for travel behavior studies by Nazemi et al. (2021). This simulator consists of an instrumented bicycle, equipped with steering, braking, wheel and pedaling bluefruit sensors, combined with immersive virtual reality. Virtual Reality is provided by an HTC Valve Index which is rendered on a high-end desktop computer. The computer is equipped with an Intel Core i9 with 32 GB of RAM and has a GeForce RTX 2080 TI graphics card. Approximately 100 frames per second are rendered; a requirement to avoid simulator sickness in VR.

#### 4. CASE STUDY

##### 4.1 Test Location

For this study, an intersection in the city of Berne, Switzerland, was selected. The intersection is known as the Burgernziel junction and is located at the intersection of the Thunstrasse, Ostring and Muristrasse. The intersection is characterized by high traffic volumes and multiples bus and tram lines traversing the intersection. Intersections are of special interest, as most collisions with cyclists occur at junctions (e.g. Dill, 2009) and proper design of junctions contributes to lower crash risk for cyclists (Groot, 2007).



**Figure 5.** Burgernziel junction drawing of Project variant (Source: Metron AG and Stadt Bern)

We chose the Burgernziel junction as test location for a variety of reasons. First, the government of the City of Berne already developed two design alternatives in collaboration with traffic planners to improve the safety and comfort of cyclist using this junction. This allowed us to focus on the development of the VR environment. Second, tree lined roads surrounding the junction provide additional challenges for texturing the facades due to the occlusion caused by the trees. Third, we had access to two different 3D city models, vehicle-based MMS data and aerial imagery which provided an ideal environment to test the utility of those various data sources to generate a VR environment.

##### 4.2 Existing Data Sets

For our study case we had two 3D city models, the cadastral surveying map, an aerial image, the elevation model and MMS data from the Ostring area.

One city model is a CityGML LOD2.5 model with roof structure and the other one is a LOD2 model, where the facades are more accurate. The aerial orthoimage of the area has a resolution of 10cm/pixel. The MMS data consists of georeferenced street level 3D imagery (Nebiker et al., 2015), which was captured in the years 2018 and 2020 using a multi-stereo mobile mapping vehicle by iNovitas AG.

##### 4.3 Data Acquisition

Since the involved streets are lined by dense rows of trees covering large parts of the building facades, an exclusively vehicle-based mobile mapping campaign would not yield satisfactory 3d reconstructions of the urban scene. Thus, it was decided to use the BIMAGE Backpack shown in Figure 3 for capturing 360° imagery along the sidewalks. This should ensure a detailed and complete reconstruction of the building facades and the vegetation as well.



**Figure 6.** Trajectories (shown in blue) of the backpack image acquisition along the sidewalks on both sides of the streets.

The measuring campaign took place on the 23.07.2020. Data of the 6 cameras was captured at 3'174 positions along the test area, resulting in a total of 19'044 images for modelling. Control and check points consisted of visible points from the cadastral survey, of points measured in the field using GNSS and of points measured in the cloud-based infrastructure service Infra3D by iNovitas.

# Exposures	# Images	# Control Points	# Check Points
3174	19'044	18	6

**Table 1.** Summary of the mapping campaign with the BIMAGE Backpack

Under ideal circumstances, UAV imagery would have been captured and incorporated, too. However, since the test site is in the vicinity of Berne airport, UAV flights are prohibited in the area. All in all, there are 21'118 images from the Backpack and the vehicle MMS in this use case. For image orientation we used the software Metashape from Agisoft.

##### 4.4 Georeferencing of the Street-Level Imagery

The georeferencing was split in two parts at the roundabout in the centre of the area to ensure faster processing. The part Thunstrasse covers the part westward of the roundabout and the part Ostring the eastward part. The images were oriented in Metashape. Table 2 lists the reprojection error and root mean square error (RMSE) with backpack images only, and with combined images from backpack and vehicle.

	Reprojection error [pix]	Control Points RMSE [cm]	Check Points RMSE [cm]
<u>Georeferencing of backpack images</u>			
Thunstrasse	0.34	1.47	8.63
Ostring	0.30	0.72	2.74
<u>Georeferencing of combined images (backpack and vehicle)</u>			
Thunstrasse	0.34	0.69	6.60
Ostring	0.40	0.73	2.02

**Table 2.** Georeferencing: Reprojection error and RMSE in X,Y,Z for control and check points

#### 4.5 3D Reconstruction

In this study, different reconstructions were investigated: a) two different software solutions, namely Bentley Context Capture and nFrames SURE and b) reconstructions using backpack imagery only versus a combination of backpack- and vehicle-based imagery. With ContextCapture the geometry could successfully be reconstructed (Figure 7). Once with the backpack data only and once with backpack and vehicle data.



**Figure 7.** Reconstructed 3D scene using backpack and vehicle data in ContextCapture

Processing time for the Ostring part with just the backpack images was 52 hours: 1 hour for the alignment and 51 hours for the 3d reconstruction. For the combination of backpack and vehicle data, overall processing time was 98 hours: 2 hours for the alignment and 96 hours for the 3D reconstruction.

	#images	Aligning[h]	Reconstruction[h]
Backpack	6240	1	51
Backpack&vehicle	7208	2	96

**Table 3.** Summary of processing time from the Ostring part

To process the data, we used a workstation with two 20 Cores CPUs, 512 GB RAM and two Nvidia GeForce 2080 Super. 3D reconstructions of the backpack imagery with SURE were more challenging. This is because SURE does not support fisheye camera models. Therefore, the backpack images first had to be converted to the perspective camera model and cropped for their central part. This eliminates some of the image overlaps. A second issue is that SURE is currently optimized for aerial and close-range image acquisition patterns with sensor motion mostly perpendicular to the viewing direction.

#### 4.6 3D Modelling and Texturing

For this case study we used different modelling software to complete the process. for 3D editing and modelling 3dsMax software was used as polygonal modeller, as a NURBS modeller we used Rhinoceros and its extension Grasshopper as algorithm editor to create the projection algorithm. Also, for this study case, a simplification of the complex terrain and streets model into few different NURBS surfaces has been made.

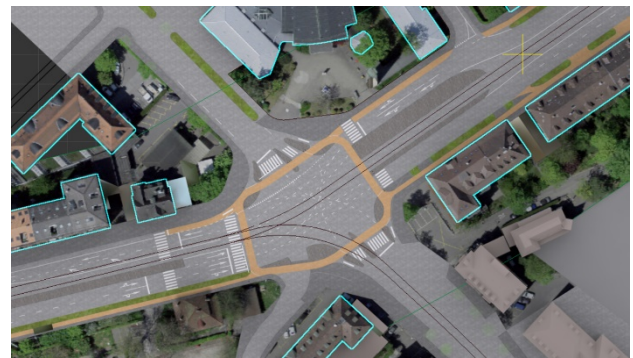


**Figure 8.** City model facades textured by MMS images

The automation of the projection was important in this project because details related to the different scenarios or changes are created in the 2D environment and then displayed and examined

in the 3D environment. This automation allowed to almost eliminate the time to introduce the third dimension to the 2D drawings. To texture the city model, the georeferenced images could be directly projected onto the 3D model with Metashape. Best texturing images were manually chosen. Depending on the size of the facade we used different quantities of images. The images in front of the building and between two buildings are manually chosen, so that the images cover every part of the building and have no interfering objects. Throughout the facades of the model are textured from 180 images from the backpack data.

Typically, a flat façade in the city model is composed of a few mesh faces. The projection of highly inclined images produces a deformation in the texture due to the low mesh detail. A subdivision increasing the number of vertices in the facade solves this problem. For texturing the roofs, the aerial orthoimage is used (Figure 9).



**Figure 9.** The roofs are textured by aerial images

#### 4.7 Extraction of Traffic-Related object Assets from Cloud-based 3D Image Service

The traffic assets (signposts or another street inventory) were digitized and modelled based on mobile mapping data hosted in the cloud-based infrastructure service Infra3D by iNovitas. Infra3D provides georeferenced 3D images of the mapped road environments, allowing accurate 3D measurements through a web client. This way, dimension and position of every visible traffic asset (e.g. a road sign) can be extracted, including its texture. The assets were subsequently created in the modelling software 3Ds Max directly at the right position and with the correct orientation and subsequently textured with the extracted images. Thus, photorealistic traffic assets can be created.

#### 4.8 Traffic simulation

Traffic microsimulations for the intersection were provided by a leading engineering consultancy and created with the software PTV Vissim. In addition to the intersection, the model extends up to 3 kilometres from the intersection. Two scenarios were modelled: the current scenario as well as a newly designed junction. We choose to export the vehicle and pedestrian positions for every fourth of a second for a period of 15 minutes. Vehicles were modelled by means of a car-following model. Pedestrians were modelled with the social force model. Traffic light timings were taken from the actual signal program, but no detectors were included that could account for variations in traffic volumes or could provide public transport priority. The initial traffic simulation considered peak hour vehicular and pedestrian flows. These flows were reduced by 50% to ensure that traffic would not be a distracting factor. On other, pedestrian and bicycle flows were increased slightly, and pedestrian areas were modelled along the sidewalk.

## 5. RESULTS & DISCUSSION

The resulting 3D reconstructed scenes show a good quality with detailed reconstructions of facades, sidewalks, vegetation and street furniture. Most objects are well defined and easily recognizable. However, before the reconstructed 3D scene can be used in the simulation environment, a manual cleaning step is still required.



**Figure 10.** The resulting Model ready for VR application.

The scene reconstruction investigations comparing backpack-only versus combined imagery from backpack- and vehicle-based imagery revealed the major challenges of image-based reconstructions using imagery of different epochs. In our case, the backpack imagery was acquired in 2020 and the vehicle-based imagery in 2018 and 2020. While the reconstruction from a single epoch worked well, combining the two data sets proved to be only partially successful. To have a fully reconstructed model, when the vehicle data are used for the street part, then the backpack images are too far away.



**Figure 11.** VR views of protected bicycle lane in different positions of different scenario models: Above: existing situation; Below: new design of the intersection.

## 6. CONCLUSION & OUTLOOK

We appreciate that the feasibility of using VR environments in the planning process is directly dependent on the ability to generate them in a cost-effective manner. Here, the combination of an automated 3D reconstruction based on data collected with an MMS backpack and with 3D models of the road infrastructure directly derived from BIM models provides strong potential to substantially simplify the generation of the VR environment. We are currently investigating the benefits of using different types of product models (Fig. 14) and the effectiveness and limitations in technical terms and in relation to user experience.

Regarding applications in research, we consider both videos and the VR bicycle simulator as promising research tools, but for different types of applications. Videos and pictures can be very useful to showcase different design options in web-based surveys and visual choice experiments. The problem of participants becoming motion sick when driving around corners severely restricts the feasibility of VR bicycle simulator applications to evaluate behaviour in junctions. However, we still see potential for research applications in the field of traffic safety and human behaviour such as the measurement of reaction times.



**Figure 12.** VR view using the two different building models produced. above: the automatic reconstruction with MMS data (dense mesh), below: city models (low poly) textured with georeferenced MMS photos.

Planners who tried the simulator agreed that videos generated based on such VR environments are probably the most relevant for practical applications. It was highlighted that the storyline of such videos must be carefully developed to convey the message.

## 7. ACKNOWLEDGEMENTS

We would like to thank Michael Joos and Muhammad Salihin Bin Zaol-kefli for their support with developing the Unity environment, Oliver Hasler for the creation of 3D assets, the City of Berne for the provided data and cooperation, Metron AG for providing the 2D plans, Rudolf Keller AG for providing the traffic microsimulation and iNovitas AG for the acquisition of the vehicle-based street-level imagery and for providing access to their infra3D cloud platform.

## 8. REFERENCES

- Bekele, M.K., Pierdicca, R., Frontoni, E., Malinverni, E.S., Gain, J., 2018. A Survey of Augmented, Virtual, and Mixed Reality for Cultural Heritage. *J. Comput. Cult. Herit.* 11, 1–36. <https://doi.org/10.1145/3145534>
- Blaha, M., Vogel, Ch., Richard, A., Wegner, J.D., Pock, Th., Schindler, K., 2016. Large-Scale Semantic 3D Reconstruction: An Adaptive Multi-resolution Model for Multi-class Volumetric Labeling, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Las Vegas, NV, USA, pp. 3176–3184. <https://doi.org/10.1109/CVPR.2016.346>
- Blaser, S., Cavegn, S., Nebiker, S., 2018. Development of a Portable High Performance Mobile Mapping System Using the Robot Operation System. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences IV-1*, 13–20. <https://doi.org/10.5194/isprs-annals-IV-1-13-2018>

- Blaser, S., Meyer, J., Nebiker, S., Fricker, L., Weber, D., 2020. Centimetre-Accuracy in Forests and Urban Canyons – Combining A High-Performance Image-Based Mobile Mapping Backpack with New Georeferencing Methods. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* V-1–2020, 333–341. <https://doi.org/10.5194/isprs-annals-V-1-2020-333-2020>
- Calvano, M., Wahbeh, W., 2014a. Disegnare la Memoria. L'immagine della città attraverso la rappresentazione integrata-Drawing the Memory. The image of the city through the integrated representation. *DISEGNARECON* 7. <https://doi.org/10.6092/issn.1828-5961/4102>
- Calvano, M., Wahbeh, W., 2014b. Drawing the Imagined City. Latina as a urban representation workshop. *Disegnare Idee Immagini-Ideas Images* 25, 80–90.
- Cavegn, S., 2020. Integrated georeferencing for precise depth map generation exploiting multi-camera image sequences from mobile mapping. Universität Stuttgart.
- Cavegn, S., Haala, N., Nebiker, S., Rothermel, M., Tutzauer, P., 2014. Benchmarking High Density Image Matching for Oblique Airborne Imagery. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-3, 45–52. <https://doi.org/10.5194/isprsarchives-XL-3-45-2014>
- Dill, J., 2009. Bicycling for Transportation and Health: The Role of Infrastructure. *Journal of Public Health Policy* 30, S95–S110.
- Farooq, B., Cherchi, E., Sobhani, A., 2018. Virtual Immersive Reality for Stated Preference Travel Behavior Experiments: A Case Study of Autonomous Vehicles on Urban Roads. *Transportation Research Record* 2672, 35–45. <https://doi.org/10.1177/0361198118776810>
- Groot, R., 2007. *Design manual for bicycle traffic*. CROW
- Haala, N., Rothermel, M., Cavegn, S., 2015. Extracting 3D urban models from oblique aerial images, in: 2015 Joint Urban Remote Sensing Event (JURSE). pp. 1–4. <https://doi.org/10.1109/JURSE.2015.7120479>
- Kaths, H., Keler, A., Kath, J., Busch, F., 2019. Analyzing the behavior of bicyclists using a bicycle simulator with a coupled SUMO and DYNA4 simulated environment. Presented at the *SUMO User Conference* 2019, pp. 199–191. <https://doi.org/10.29007/dcmp>
- Lafarge, F., Keriven, R., Bredif, M., Vu, H.-H., 2013. A Hybrid Multiview Stereo Algorithm for Modeling Urban Scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 5–17. <https://doi.org/10.1109/TPAMI.2012.84>
- Lopez, P.A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., Wießner, E., 2018. Microscopic traffic simulation using SUMO, in: The 21st IEEE International Conference on Intelligent Transportation Systems (ITSC). IEEE.
- Musialski, P., Wonka, P., Aliaga, D.G., Wimmer, M., Van Gool, L., Purgathofer, W., 2013. A survey of urban reconstruction. *Computer Graphics Forum* 32, 146–177. <https://doi.org/10.1111/cgf.12077>
- Nazemi, M., van Eggermond, M.A.B., Erath, A., Schaffner, D., Joos, M., Axhausen, K.W., 2021. Studying bicyclists' perceived level of safety using a bicycle simulator combined with immersive virtual reality. *Accident Analysis & Prevention* 151, 105943. <https://doi.org/10.1016/j.aap.2020.105943>
- Nebiker, S., Cavegn, S., Loesch, B., 2015. Cloud-Based Geospatial 3D Image Spaces—A Powerful Urban Model for the Smart City. *ISPRS International Journal of Geo-Information* 4, 2267–2291. <https://doi.org/10.3390/ijgi4042267>
- Piegl, L., 1991. On NURBS: A Survey. *IEEE Computer Graphics and Applications* 11, 55–71. <https://doi.org/10.1109/38.67702>
- Rothermel, M., Wenzel, K., Fritsch, D., Haala, N., 2012. SURE: Photogrammetric Surface Reconstruction from Imagery, in: *Proceedings LCD Workshop*. Berlin, p. 9.
- Schmohl, S., Tutzauer, P., Haala, N., 2020. Stuttgart City Walk: A Case Study on Visualizing Textured DSM Meshes for the General Public Using Virtual Reality. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 88, 147–154. <https://doi.org/10.1007/s41064-020-00106-z>
- Schonberger, J.L., Frahm, J.-M., 2016. Structure-from-motion revisited. Presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4104–4113.
- Schramka, F., Arisona, S., Joos, M., Erath, A., 2017. Development of Virtual Reality Cycling Simulator, in: 3rd International Conference on Virtual Reality. Presented at the 3rd International Conference on Virtual Reality, Hong Kong.
- Strecha, C., Zoller, R., Rutishauser, S., Brox, B., Schneider-Zapp, K., Chovancova, V., Krull, M., Glassey, L., 2015. Quality Assessment of 3D Reconstruction Using Fisheye And Perspective Sensors. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W4, 215–222. <https://doi.org/10.5194/isprsannals-II-3-W4-215-2015>
- Unity, 2021. Navigation and Pathfinding: Unity User Manual <https://docs.unity3d.com/Manual/Navigation.html> (21.4.2021)
- Vu, H.-H., Labatut, P., Pons, J.-P., Keriven, R., 2012. High Accuracy and Visibility-Consistent Dense Multiview Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 889–901. <https://doi.org/10.1109/TPAMI.2011.172>
- Wang, R., Peethambaran, J., Chen, D., 2018. LiDAR Point Clouds to 3-D Urban Models: A Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11, 606–627. <https://doi.org/10.1109/JSTARS.2017.2781132>
- Wen, J., Gheisari, M., 2020. Using virtual reality to facilitate communication in the AEC domain: a systematic review. *Constr. Innov.* 20, 509–542. <https://doi.org/10.1108/CI-11-2019-0122>