ISPRS Int. J. Geo-Inf. **2015**, *4*, 2267-2291; doi:10.3390/ijgi4042267

OPEN ACCESS

ISPRS International Journal of

Geo-Information

ISSN 2220-9964

www.mdpi.com/journal/ijgi/

Article

Cloud-Based Geospatial 3D Image Spaces—A Powerful Urban Model for the Smart City

Stephan Nebiker*, Stefan Cavegn and Benjamin Loesch

Institute of Geomatics Engineering, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Gründenstrasse 40, 4132 Muttenz, Switzerland; E-Mails: stefan.cavegn@fhnw.ch (S.C.); benjamin.loesch@fhnw.ch (B.L.)

* Author to whom correspondence should be addressed; E-Mail: stephan.nebiker@fhnw.ch; Tel.: +41-61-467-4336; Fax: +41-61-467-4460.

Academic Editors: Jochen Schiewe and Wolfgang Kainz

Received: 28 July 2015 / Accepted: 14 October 2015 / Published: 26 October 2015

Abstract: In this paper, we introduce the concept and an implementation of *geospatial 3D image spaces* as new type of *native urban models*. 3D image spaces are based on collections of georeferenced RGB-D imagery. This imagery is typically acquired using multi-view stereo mobile mapping systems capturing dense sequences of street level imagery. Ideally, image depth information is derived using dense image matching. This delivers a very dense depth representation and ensures the spatial and temporal coherence of radiometric and depth data. This results in a high-definition WYSIWYG ("what you see is what you get") urban model, which is intuitive to interpret and easy to interact with, and which provides powerful augmentation and 3D measuring capabilities. Furthermore, we present a scalable cloud-based framework for generating 3D image spaces of entire cities or states and a client architecture for their web-based exploitation. The model and the framework strongly support the smart city notion of efficiently connecting the urban environment and its processes with experts and citizens alike. In the paper we particularly investigate quality aspects of the urban model, namely the obtainable georeferencing accuracy and the quality of the depth map extraction. We show that our image-based georeferencing approach is capable of improving the original direct georeferencing accuracy by an order of magnitude and that the presented new multi-image matching approach is capable of providing high accuracies along with a significantly improved completeness of the depth maps.

Keywords: smart city; urban modeling; mobile mapping; stereovision; image matching; georeferencing; cloud computing; 3D monoplotting; augmentation

1. Introduction

The original concept of "smart city" was first postulated in the late 1980s and was focused on the role of information and communication technologies (ICT) with regard to modern urban infrastructures. Since then it has evolved into a (multi-dimensional) more general concept relying on the use of ICT to enhance quality and performance of urban services, to reduce costs and resource consumption, and to engage more effectively and actively with its citizens. Albino *et al.* [1] provide a good overview of the evolution, dimensions, and definitions of different variants of a smart city. Most definitions contain elements that are closely related to geospatial concepts, in general, and to the new urban modeling approach introduced in this paper. Hall *et al.* [2], for example, emphasize the monitoring of critical infrastructures, such as roads, bridges, tunnels, rails, subways, major buildings, *etc.*, in order to optimize resources, plan preventive maintenance activities, and monitor security aspects with the intent to maximize services to its citizens. Harrison *et al.* [3] emphasize the connectivity of the physical infrastructure, the IT infrastructure, the social infrastructure, and the business infrastructure to leverage the collective intelligence of the city. Cretu [4], last but not least, identifies governance and economy as important drivers for smart cities and highlights the need for new thinking paradigms. A common denominator of all smart city definitions is the employment of ICT concepts and infrastructures allowing people to smartly interact with real-world objects and processes. Such ICT solutions, again, require models of the real world in our case urban models, in order to represent, interact with, analyze or simulate the urban environment and processes. Since a large part of urban infrastructure and activity is closely linked to road corridors, streetside urban models are of particular importance in a smart city context. Today, citywide streetside environments can be routinely captured by vehicle-based mobile mapping systems [5–8]. Early research and visionary experiments, such as the Aspen Movie Maps project [9], date back to the late 1970s and were entirely image-based. In this visionary project, the image-based virtual urban streetside environment was used for interactively navigating through and interacting with the real world. It, thus, demonstrated many of the features, which more than 25 years later became part of popular street-level mapping services, in particular Google Street View [8]. Research in mobile mapping systems and sensors was originally focused on direct georeferencing of frame imaging sensors [10]. With the emergence of mobile LiDAR sensors the focus was almost completely shifted to mobile laser scanning (MLS) [11] which is currently dominating the market for engineering applications. However, due to tremendous progress in sensor technologies, photogrammetric, and computer vision algorithms, and due to new applications, namely indoor mobile mapping, image-based approaches have again become a strong research focus in different communities [7,12,13].

This paper introduces a new type of *native urban models* based on collections of georeferenced 3D imagery from multiview-stereo together with a fully functional implementation. The new model provides a high-fidelity representation of the streetside environment, accurate and robust 3D measurement capabilities, powerful options for capturing and augmenting urban infrastructure elements,

and is extremely easy-to-use. It, thus, combines the high accuracies of point cloud based urban models from MLS with the intuitive navigation and interpretation found in popular image-based streetside web services. This makes the 3D image-based urban model suitable for a wide range of smart city applications for professionals and citizens alike.

2. Related Work

Urban models can be considered as suitable digital representations of urban environments for capturing, managing, analyzing and visualizing specific urban processes. As pointed out in the introductory discussion on the notion of the "smart city", the spectrum of such processes and of their needs is extremely broad. Thus, there exists no "suits-it-all" urban model supporting all conceivable smart city contexts and applications. While there is no widely accepted taxonomy of (3D) urban models, there are a number of helpful classifications. Meilland *et al.* [14], for example, distinguish between 3D parametric models and image-based key-frame models, but they do not cover point cloud-based models, which play an important role in urban modeling. In an earlier overview of urban models [15], the authors distinguish between geometric 3D models, image-based models, and a rich point cloud model. They also provide a comparison between these three model types based on the following criteria: modeling concept, representation modeling, modeling strategy, prevailing acquisition strategy and coverage, georeferencing accuracy requirements, a typical modeling scope ranging from micro to macro scale, suitable visualization scenarios, as well as navigability.

While new geospatial sensors for acquiring urban models are emerging and rapidly evolving, there seems to be a convergence into two main types of urban models. We refer to them as *urban reconstructions* or derived 3D urban models on the one hand and *native urban models* on the other.

For a comprehensive overview of research activities in *urban reconstruction* in the fields of computer graphics, computer vision, photogrammetry, and remote sensing we refer the reader to the survey of Musialski *et al.* [16]. Despite the efforts and major progress in the automatic generation of accurate parametric 3D models [12,13,17] from airborne or ground-based imagery and point clouds, 3D reconstruction remains a complex and ill-posed tasks. In order to control complexity and to address inherent problems, such as occlusions and gaps in the data, state-of-the art urban reconstruction techniques employ priors or grammars for specific types of urban structures to be modeled. These approaches often lead to visually-appealing, photorealistic or abstracted, possibly even semantic, urban models [12,18,19], but they are not well-suited for unstructured environments or for urban models with high to very high accuracy requirements, e.g. for high-quality 3D measurements [7,20–22].

Native urban models, the second type of urban models, primarily consist of basic geospatial data types. These include: monoscopic, stereoscopic [7], panoramic [23] or RGB-D imagery [7,8,14], 3D point clouds [15] or combinations thereof [24]. Native urban models do not aim at complete 3D reconstructions of entire objects or urban scenes. As a consequence, native urban models are less complex, do not require a high level of scene understanding, and can be generated with a much higher level of automation and robustness than actual urban reconstructions. However, large-scale high-definition native urban models come at the cost of extremely large data volumes. Therefore, apart from major progress in kinematic positioning and mobile mapping sensor technologies, the dramatic increase in network bandwidths, as well as in cloud storage and cloud computing capacities have been important

enablers and drivers for research in native urban models [25]. Additionally, numerous research activities were spurred by successful commercial examples of native urban models, most of all Google Street View [8].

The main contributions of our paper are as follows. We first introduce the concept of 3D geospatial image spaces (Section 3). We then present a framework implementing all components for capturing, processing, and exploiting the new type of native urban model (Section 4). We subsequently address three important research questions related to 3D image spaces and discuss the respective results based on real-world experiments:

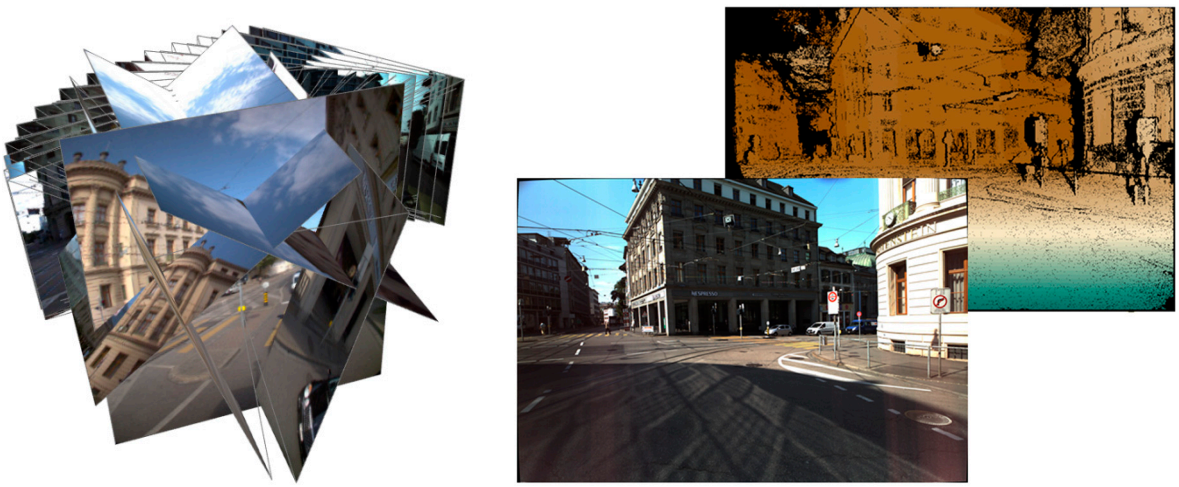
- Georeferencing strategies for the RGB-D imagery and the obtainable absolute measuring accuracies (Section 5);
- Depth map extraction strategies and the obtainable relative measuring accuracies (Section 6);
- The smart exploitation of the new urban model with respect to functionality and ease-of-use (Section 7).

In the final section, we provide conclusions on strengths and limitations of 3D image spaces and give an outlook on future developments and perspectives.

3. Geospatial 3D Image Spaces

3.1. Concept

We propose a simple but powerful new native urban model type, which we refer to as *Geospatial 3D Image Spaces*. Such 3D image spaces (Figure 1a) consist of collections of georeferenced RGB-D imagery, combining radiometric (RGB), and depth information (D) (Figure 1b).



(a) **(b)**

Figure 1. (**a**) Conceptual illustration of 3D image spaces consisting of collections of georeferenced multi-view RGB-D imagery; and (**b**) georeferenced RGB image with its associated depth map (D) containing a depth value for each pixel of the RGB image.

The native urban model of 3D image spaces shall fulfill the following requirements:

- Provide a high-fidelity metric photographic representation of the urban environment, which is easy to interpret and which can be augmented with existing or projected GIS data
- The RGB and the depth information shall be spatially and temporally coherent, *i.e.* the radiometric and the depth observation should ideally take place at exactly the same instance
- The depth information shall be dense, ideally providing a depth value for each pixel of the corresponding RGB image
- Image collections are usually ordered, e.g., in the form of images sequences, for simple navigation and shall efficiently be accessed via spatial data structures
- The model shall support metric imagery with different geometries, e.g., with perspective, panoramic or fish eye projections
- The model shall be easy-to-use and shall at least support simple, robust and accurate image-based 3D measurements using enhanced 3d monoplotting
- The model shall provide measures to protect privacy

The urban model of 3D image spaces can further be characterized by its scope, by typical imaging sensors, acquisition strategies, and by its specific exploitation features and techniques.

Scope—The scope of our proposed urban model is very broad. While the current focus and use is predominantly street level and indoor modeling, the model would also be applicable to nadir and oblique aerial imagery [26,27].

Sensors—The model does not depend on a specific data acquisition technique as long as it ensures the spatial and temporal coherence of the RGB and D information. This temporal coherence is particularly important in urban environments with numerous moving objects such as cars, trams or pedestrians (see Section 4.4). In outdoor urban environments, the requirement of dense and temporally-coherent RGB-D values with a sufficiently high spatial resolution can currently only be met by stereo or trinocular camera setups [6,7,14,20] and a subsequent depth extraction using dense image matching (see Section 6). Once active range imaging sensors will reliably work in bright daylight and provide a sufficiently high spatial resolution, they might become the sensor technology of choice for acquiring urban 3D image spaces. Hybrid sensor configurations, e.g., combinations of monoscopic cameras and LiDAR sensors [8], are in wide-spread use in commercial mobile mapping systems. However, they are of limited use in typical urban environments, as long as the imagery and range information for a real-world object are not collected simultaneously and with the same viewing geometry.

Acquisition Strategy—The goal in generating 3D image spaces is to capture every object of interest within multiple images, ideally from multiple perspectives and with a spatial resolution, which allows the reliable identification of any object to be mapped. In case of street level urban mapping, multi-view imagery is typically acquired in dense sequences along the vehicle trajectory. These dense imaging patterns not only allow for a smooth navigation; they can also be exploited to increase the georeferencing accuracy [21,22] or the robustness of depth extraction, as will be shown in the following sections.

Exploitation—The interaction with our proposed 3D image spaces is relatively simple but powerful and is similar to well-known street level services such as Google Street View. Due to the dense image acquisition patterns, frame-to-frame navigation within the model is almost video-like. While the model is constraining the viewing position to the original acquisition position, users can still freely pan and

zoom through the original images without ever noticing the third dimension hidden behind the familiar (2D) imagery and without having to revert to an actual 3D visualization of the RGB-D image.

3.2. Discussion

3D images spaces differ from monoscopic, typically panoramic 2D image spaces such as offered by Cyclomedia's Cycloramas [23], whose main measuring principle is based on interactive forward intersections within panoramas captured from multiple positions and at different epochs. 3D image spaces have numerous similarities to the model behind Earthmine's commercial solution [6,28]. Their published approach, however, sacrifices some of the accuracy by projecting and mosaicking the original multi-head panoramic imagery and depth information onto a cylindrical panorama [28]. A recently published and similar approach uses collections of spherical RGB-D panoramas of large-scale urban environments for autonomous navigation and real-time localization [14]. The spherical RGB-D panoramas, referred to as "augmented spherical key-frames", are also generated by warping and mosaicking multiple images onto a sphere. Dense depth maps for the sphere are obtained by applying different dense image matching algorithms, such as semi-global matching (SGM) [29] and Efficient LArge-scale Stereo (ELAS) [30] to the rectified spherical panoramas with a subsequent triangulation of the depth information. Image data for both models are acquired using multi-head camera systems. By considering all sensors of a panoramic head to have a unique center of projection [14], both approaches reduce the relative and absolute 3D measurement accuracy from a potential cm-level to the dm-level or even lower. Other approaches such as the one described for the Stereopolis II system [7], preserve the original perspective image geometry and are thus not affected by the accuracy degradation.

In conclusion, the proposed native urban model of 3D image spaces preserves the original image geometry and ensures the spatial and temporal coherence of the radiometric and depth information. Thus, in contrast to parametric 3D models or hybrid image-LiDAR based urban models, 3D image spaces ensure the principle of WYSIWYG ("What You See Is What You Get"). This is particularly important in a smart city context, where urban models are not only to be used by experts but also by a potentially large number of the citizens. Last but not least, by preserving the original image geometry, the urban model itself can be used for very accurate integrated or exclusively image-based georeferencing. Thus, the accuracy of the urban model could be increased as and when needed—possibly even months or years after the original data acquisition.

4. Implementation and Test Environment

Following a series of research projects on vision-based mobile mapping and urban modeling at the University of Applied Sciences and Arts Northwestern Switzerland FHNW, *3D image spaces* have been the main focus of the most recent project named infraVIS (Sustainable Infrastructure Management based on Versatile Intelligent 3D Image Spaces). The main goal of the project was to demonstrate the feasibility of 3D image spaces and to evaluate their capabilities in real-world scenarios covering entire cities or states. The project thus addresses the entire process chain from the mobile acquisition of multi-view stereo and panoramic imagery, the image-based extraction of dense and accurate depth maps and the fully cloud-based processing and web-based exploitation of such large scale image spaces. An overview of the architecture and workflow for 3D image spaces is given in Figure 2. In the following sub-sections

we introduce the data acquisition system (Section 4.1), the processing pipeline (Section 4.2), the exploitation system (Section 4.3), and the test environment (Section 4.4) used in the experiments.

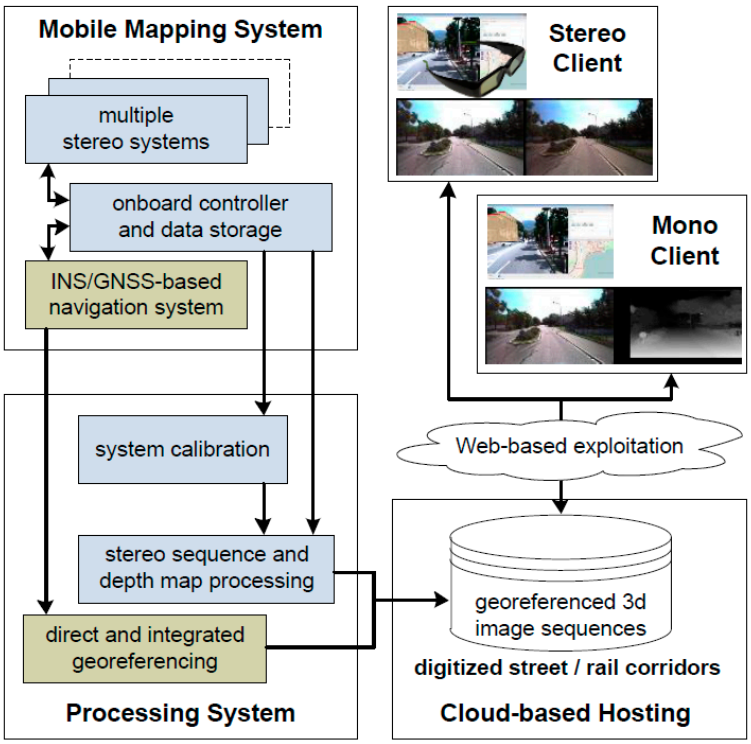
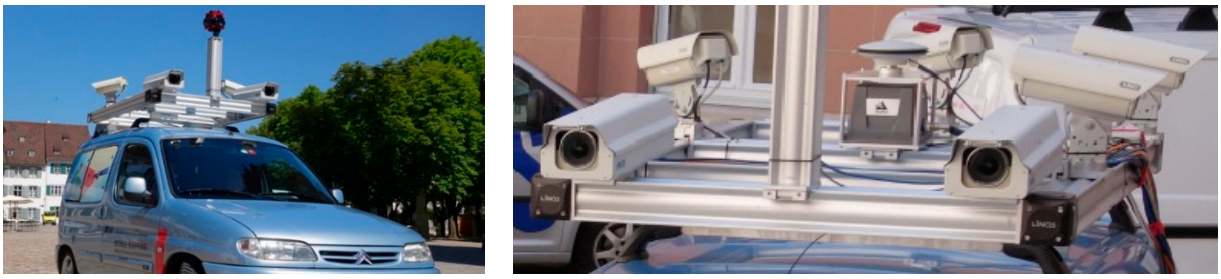


Figure 2. System architecture and workflow of the infraVIS project.

4.1. Data Acquisition System

Data for the subsequent tests was acquired using the multi-stereovision mobile mapping research platform of the Institute of Geomatics Engineering (IVGI) at FHNW [20]. The current acquisition system shown in Figure 3 has the following features [31]:

- A NovAtel SPAN inertial navigation system with a tactical grade UIMU-LCI inertial measuring unit (IMU) featuring fiber-optics gyros and with a L1/L2 GNSS kinematic antenna
- Up to five stereo camera systems with either 11 MP or Full HD resolution, a typical radiometric resolution of 12 bits and max. data capturing rates of 5 fps or 30 fps respectively
- The stereo systems are mounted on a rigid frame with typical stereo baselines of approx. 1 m
- Typical configurations consist of a main stereo system facing forward and additional stereo systems facing aft, sideways or even pointing downwards at the road surface
- Recent additions include up to two Ladybug 5 multi-head panoramic cameras
- All sensors are synchronized using hardware trigger signals from a custom-built trigger box which also supports distance-based triggering to ensure uniform image sequences even in busy or congested traffic
- Typical data acquisition speeds range from 30 to 80 km/h and max. acquired data volumes are in the order of up to 1 TB per hour of operation, depending on the acquisition parameters



(**a**) (**b**)

Figure 3. (**a**) Multiview multi-sensor stereovision IVGI mobile mapping system; and (**b**) detail view showing the three stereo camera systems and the GNSS/IMU positioning system.

4.2. Processing Pipeline

Our processing pipeline for creating geospatial 3D image spaces from multi-view stereo imagery includes the steps of system calibration, georeferencing (see Section 5), generation of normalized and distortion-free images, the subsequent depth map generation (see Section 6), and finally the generation of multi-resolution tiles for the imagery and the depth maps. The processing framework uses Python as a wrapper language and high-level languages for computationally-intensive tasks. The framework features a multi-platform support and can be operated on individual workstations, on high-performance compute clusters (HPC), and in highly-scalable cloud computing environments, e.g., Amazon's AWS Cloud Computing Services.

The system calibration plays a key role in ensuring the postulated high relative and absolute accuracies that can be achieved with the resulting urban model. Calibration of a multi-sensor mobile mapping platform such as ours is documented in [10,20,32–34]. It includes the determination of the interior orientations of the 20+ sensor heads; the determination of relative orientation parameters among all sensors heads; and, finally, the calibration of lever arm and misalignment between imaging sensors and the IMU body frame. For a description of the calibration procedure and an evaluation of the results, we refer the reader to [20].

For depth map extraction, we are using a number of dense image matching algorithms and implementations. These include OpenCV StereoSGBM [35], a simplified variant of the SGM algorithm [29], SURE by nFrames [36], and Agisoft PhotoScan [37]. An important element of our depth map generation is the calculation of a matching quality indicator, which is stored with every pixel of the depth map.

4.3. Cloud-Based Management and Web-Based Exploitation System

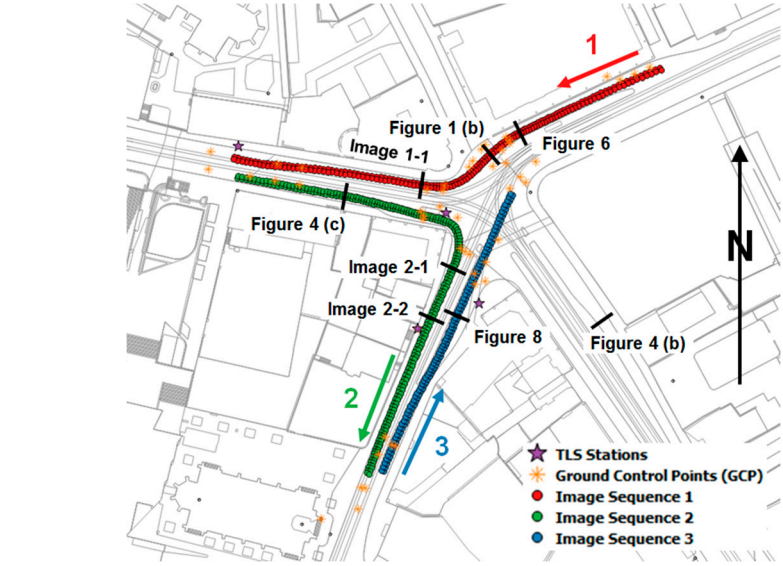
During the entire processing workflow a comprehensive meta-database, representing all aspects of a 3D images space is populated. It includes information on sensor calibration, road network topology, vehicle trajectories and image sequences, interior and exterior orientation of every image within the 3D image space, plus an abundance of additional metadata. The meta-database, among many other things, ensures a highly-efficient spatial and temporal access to image sequences and to individual 3D image frames.

For the web-based exploitation, a dedicated 3D engine and an SDK for the georeferenced 3D image spaces were implemented. They provide access to the cloud-based 3D imagery and metadata. They also incorporate several data streaming concepts, such as tiled image loading, caching, or spatial preloading.

The SDK also includes numerous features for intelligent 3D measurements and for augmenting the 3D imagery with other geospatial content (see Section 7). The engine and the SDK are entirely based on modern web technologies such as HTML5 and WebGL in order to ensure cross-platform access from any desktop or mobile device.

4.4. Study Area and Data

For the following investigations, we chose a relatively small but demanding test site depicted in Figure 4a. The site is located at a very busy junction between five roads in the city center of Basel, Switzerland. It includes three tramway stops resulting in many overhead wires and is surrounded by rather tall commercial properties (Figure 4b,c). This creates a very challenging environment for GNSS positioning. Furthermore, construction work, as well as a large number of moving objects in the form of pedestrians, cars, and tramways, were present during data acquisition.



(**a**)

(**b**)

(**c**)

Figure 4. (**a**) Base map of the test area with overlaid projection centers of the selected image sequences, ground control points (GCPs), terrestrial laser scanning (TLS) stations, and locations of figures of this paper (Source: Geodaten Kanton Basel-Stadt); forward-looking mobile mapping imagery illustrating typical challenges; and (**b**) GNSS shadowing and numerous pedestrians; (**c**) heavy traffic with multiple trams, cars and cyclists.

Mobile mapping data was acquired in July 2014 as part of a complete survey of the city-state of Basel in cooperation with our research partner iNovitas AG. While a total of eight image sequences were

available in the test area, we selected forward imagery from the three sequences depicted in Figure 4a for further investigations. Two roads were mapped in both directions and one road in one direction only. The selected test data was acquired at different dates and daytimes, which reflects typical real-world situations. The nominal along-track distance between successive image exposures was around 1 m.

In order to assess the performance and quality of different matching strategies, independent and highly accurate reference data was acquired. Four 360° terrestrial laser scans (TLS) recording XYZ point geometry and intensity were obtained using a Leica ScanStation P20 on 19.03.2015. By registering the point clouds onto several cadastral reference points, an absolute 3D TLS accuracy of 1–2 cm was obtained. In addition, coordinates of 70 ground control points (GCP) were determined using a Leica Nova MS50 total station with a resulting 3D accuracy of better than 1 cm. All measurements were performed in the Swiss reference frame LV95 and with orthometric heights in the LN02 reference frame.

5. High Accuracy Georeferencing—Strategies and Results

5.1. Motivation and Challenges

The goal of our high-definition urban model is to ensure relative 3D measurements accuracies within individual or neighboring 3D image frames at the cm or even sub-cm level and to allow for absolute measurement accuracies, *i.e.* 3D coordinate determination accuracies, at the sub-dm level down to the cm level. Thus, the targeted relative and absolute geospatial accuracies are at a similar level as the resolution of the imagery, which is in line with the WYSIWYG goal postulated earlier. These accuracy goals are very ambitious considering the following challenges:

- A kinematic acquisition with typical speeds between 30 and 80 km/h
- In challenging urban environments with generally poor GNSS coverage
- With the need to also create such models in GNSS-denied areas such as in tunnels or buildings,
- The requirement to tie the urban model, *i.e.* the 3D imagery, to local control points,
- The use of multi-sensor systems with typically more than 10 sensor heads.

5.2. Direct Georeferencing

The strength of mobile mapping systems is their ability to directly georeference their mapping sensors in relation to a mapping coordinate frame [10]. While there are also online calibration approaches for image-based mobile mapping systems [32], we subsequently assume an accurate off-line calibration of the entire multi-sensor system as described earlier. In airborne photogrammetry, where we can generally expect a good GNSS coverage, the direct georeferencing accuracy largely depends on the angular measurement quality of the inertial measuring unit (IMU) [38]. In street level mobile mapping of urban environments, by contrast, GNSS coverage is often poor to insufficient.

Earlier experiments with our mobile mapping system [20] with average to good GNSS coverage, demonstrated that absolute 3D point measurement accuracies of 3–4 cm horizontally and 2–3 cm vertically (1 sigma) can be achieved. In the same study [20] it was shown that the 11 MP stereo system is capable to deliver relative measurements within a single stereo frame or between points in neighboring frames of the image sequence with an accuracy better than 1 cm.

5.3. Integrated and Image-Based Georeferencing

In contrast to direct georeferencing, which exclusively relies on the position and attitude information provided by the inertial navigation system, integrated georeferencing—often also referred to as integrated sensor orientation (ISO)—uses other sensor observations for georeferencing the mobile mapping sensors. While integrated georeferencing is also applied to MLS, multi-view image-based mobile mapping systems offer a particularly great potential for exploiting accurate and often highly redundant image-based measurements.

There are a number of approaches for integrated georeferencing of image-based mobile mapping systems. One approach [21] uses image-based measurements to natural control points for position and attitude updates to the original, directly georeferenced vehicle trajectory. The authors apply an optimized least squares multi-ray matching algorithm [22] to stereo image sequences for the efficient semiautomatic measurement of control and tie points. Subsequently, a constrained stereo bundle adjustment is used for the independent estimation of exterior orientation parameters of a sequence of stereo frames [22]. With their approach of vision-based trajectory updates [21], the authors demonstrated a consistent improvement of the absolute 3D positioning accuracy—and subsequent absolute 3D measurement accuracy in the 3D imagery—from originally several dm to a level of 2–5 cm horizontally and vertically. Since the establishment of ground control measurements is often more time consuming and costly than the mobile mapping campaign itself, we earlier on proposed the fusion of ground-based imagery from mobile mapping systems with aerial imagery from airborne photogrammetric surveys [27]. The rationale for doing so is that airborne surveys are much less affected by the GNSS degradations experienced by ground-based mobile mapping systems. As such, the airborne imagery provides a quite homogeneous 'datum', to which the street level imagery can be referenced. In first experiments, horizontal accuracies in the order of 5 cm, equivalent to the ground sampling distance (GSD) of the aerial imagery, and vertical accuracies of approx. 10 cm were demonstrated.

5.4. Experiments and Results

The goals of the following experiments were (a) to assess the quality of directly georeferenced sensor orientations in a challenging urban environment and (b) to improve the sensor orientation quality using automated image-based georeferencing techniques. These improved sensor orientations were required for the subsequent evaluation of the extracted depth maps (Section 6) and their comparison with reference TLS data.

The directly georeferenced sensor orientations for the trajectories shown in Figure 4a were obtained by tightly coupled GNSS/INS post-processing using NovAtel Inertial Explorer. Subsequently, bundle adjustments using Agisoft PhotoScan were performed with the image projection center coordinates obtained from direct georeferencing as initial values. Image sequence 1 (Figures 4a and 5), for example, comprises 123 stereo image pairs of 11 MP resolution captured with the forward pointing stereo system on the 24.07.2014 at 10:20 over a length of 164 m. In the bundle adjustment 13,072 tie points and 91,239 projections were computed leading to an overall projection error of 0.67 px. 133 measurements on 27 GCP resulted in an overall error of 14 mm in the east direction, 11 mm in the north direction, and 4 mm in height, which corresponds to a 3D error of 18 mm or 0.34 px respectively.

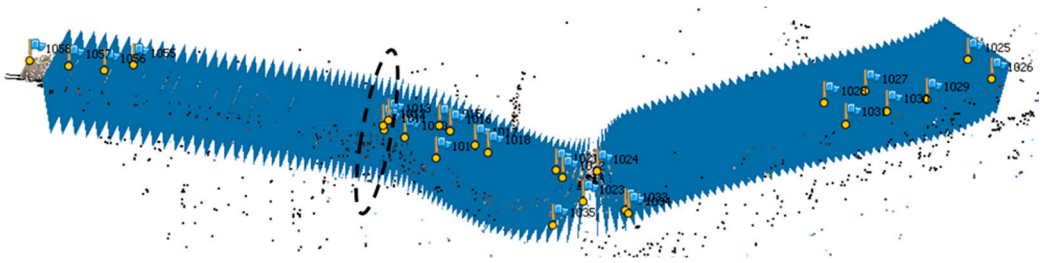


Figure 5. Perspective view of image sequence 1 with stereo frames (blue rectangles) and GCPs (yellow flags) following a bundle adjustment in PhotoScan. The location of frame 80 with a trajectory jump is marked with a dashed ellipse.

Overall deviations between direct georeferencing and bundle adjustment are 301 mm in the east direction, 38 mm in the north direction, and 423 mm in height as shown in Figure 6. The figure also shows a significant trajectory discontinuity at the location of image number 80 (see dashed ellipse in Figure 5 and dashed line in Figure 6) with a coordinate jump of 53 mm eastward, −22 mm northward, −40 mm in height, and 70 mm in 3D space. The coordinate jump occurred between the consecutive images depicted in Figure 6 when the mobile mapping vehicle had to stop for several seconds in front of a crosswalk. While the overall 3D deviation of direct georeferencing and bundle adjustment is 520 mm for image sequence one, the deviations of 93 mm and 81 mm for image sequences two and three are significantly lower. Further values are provided in [31].

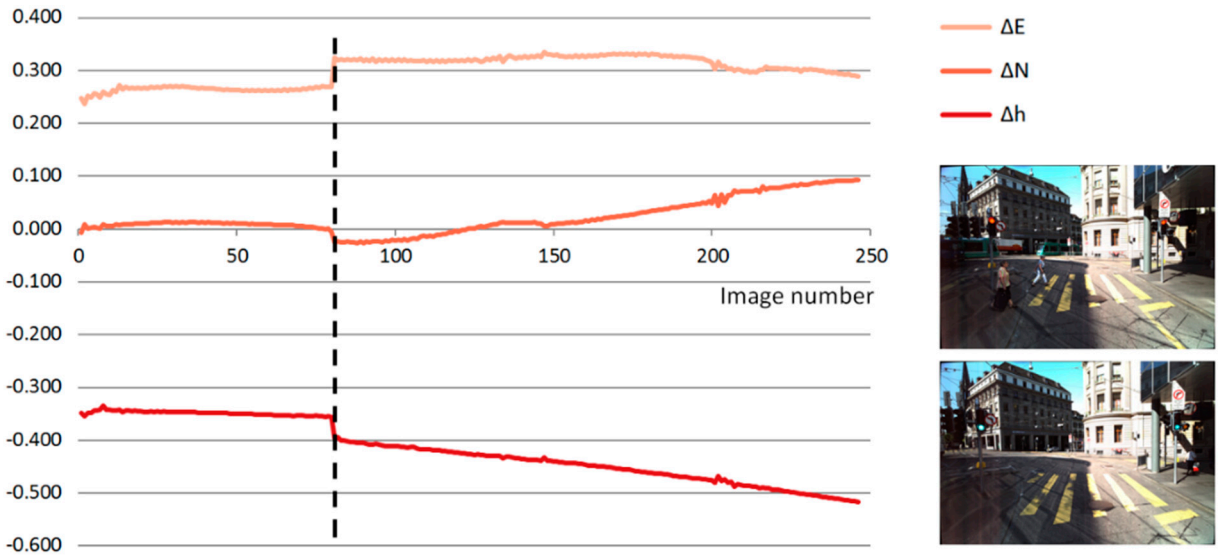


Figure 6. Differences in [m] between directly georeferenced sensor orientations (projection center coordinates of right and left stereo cameras) and image-based georeferencing from bundle adjustment for image sequence one. Trajectory discontinuity between image frames #80 and #82 are marked with a dashed line and the corresponding images are shown to the right.

5.5. Discussion

In open areas with good GNSS coverage, direct georeferencing of image-based mobile mapping systems in combination with state-of-the art calibration procedures is capable of delivering absolute 3D measuring accuracies better than 5 cm horizontally and vertically. However, in built-up urban environments with extended areas of poor GNSS coverage, direct georeferencing accuracies are typically in the order of (sub-) meter, even with expensive, high-grade inertial navigation equipment. Integrated georeferencing approaches [21,27], using image-based observations for transforming directly georeferenced trajectory segments onto GCPs, efficiently remove the potentially large offset and drift errors shown in Figure 6. However, they are not suitable to also detect and compensate discontinuities in the trajectories also shown in Figure 6. For urban modeling applications and for subsequent dense multi-image matching requiring relative and absolute accuracies at the cm level, the demonstrated image-based georeferencing approach, employing automated bundle adjustment, and offers an efficient and reliable solution.

6. Dense Image Matching for Depth Map Extraction—Strategies and Results

3D image spaces rely on good depth values for every pixel of each image. Thus, the accurate, robust, and complete extraction of depth information from dense image matching is an important goal of our research. In this section, we investigate the effect of different stereo and image sequence matching strategies on the quality of the extracted depth maps.

6.1. Matching Approaches and Configurations

For the following investigations image sequences from the forward-looking camera system shown in Figure 3b were used. Furthermore, the four matching Configurations c1 to c4 depicted in Figure 7 were selected. Configuration c1 represents standard stereo matching with one base image and one match image for which a great number of algorithms exists [39]. Configuration 2 represents the case in which only mono imagery would be available. It is limited to the sequential matching of the base image with the previous and the following image. This case puts high demands on providing sufficient relative orientation accuracy but does not require a synchronization between multiple cameras. With Configurations 3 and 4 we introduce and investigate two new multi-view stereo approaches, for which better results can be expected. In case of Configuration 4, the base image is matched with all five neighboring images. Omitting the two match images of Configuration 2 from Configuration 4 leads to Configuration 3.

In case of Configuration 1, stereo imagery captured at the same epoch is used for the matching process, which is a standard procedure. The other configurations include imagery acquired at different epochs and with strong motion and scale differences in viewing direction, which are typical for mobile mapping scenarios. The predominant motion in viewing direction between neighboring images results in stereo epipoles located either inside or close to the stereo partner. This requires advanced rectification approaches such as polar rectification, which can deal with all possible stereo geometries. In case of the software SURE [36] which was used for the following investigations, the polar rectification proposed in [40] (Figure 8) is used. Implementation details and first results are described in [31].

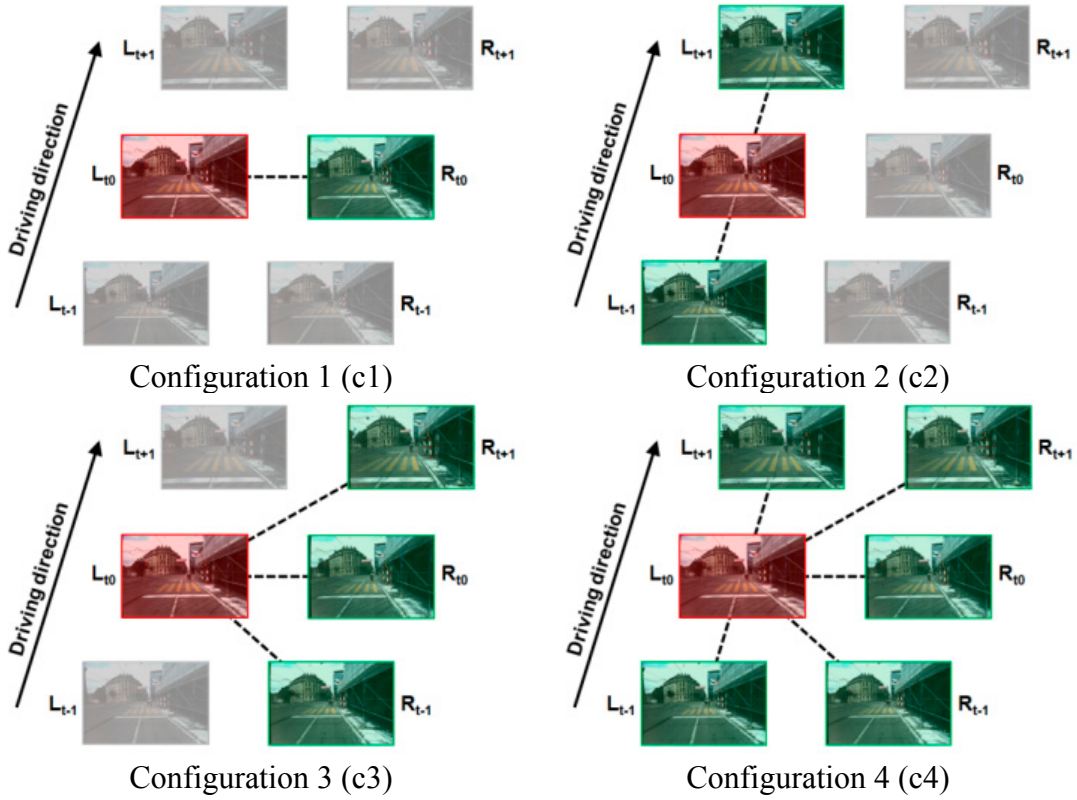


Figure 7. Selected image matching configurations, red: base image, green: match images.

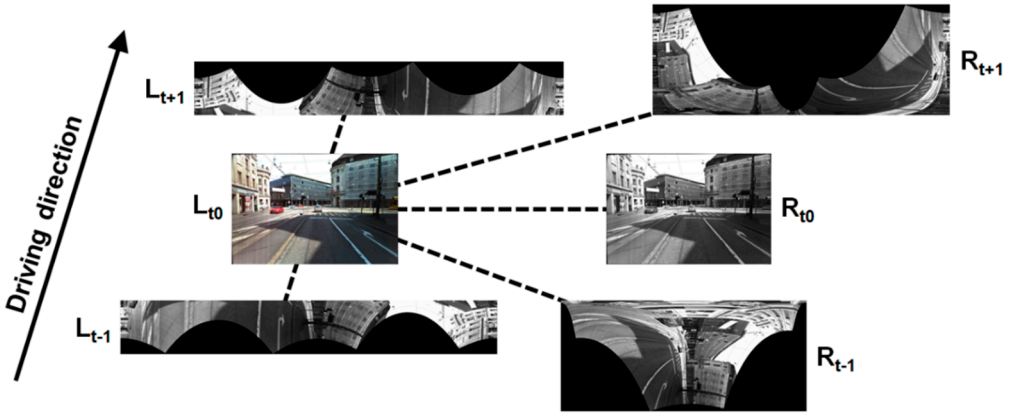


Figure 8. Base image (L_{t0}) and its neighboring images rectified by SURE using polar rectification.

6.2. Experiments and Results

Accurate depth maps are fundamental for the urban model of 3D image spaces and in particular for reliable and accurate 3D monoplotting applications. Therefore, comparisons of depth maps were performed, which were either generated directly by the SURE triangulation module or obtained by back-projecting point clouds to the viewing geometry of a base image. Similar to the methodology of [41], who did not interpolate ground truth disparity maps in order to avoid artificial errors, we also did not interpolate the depth maps. This allowed the evaluation of the raw depth values and to cope with missing parts of depth maps. Depth deviations were only computed for pixels holding values for both depth maps

and only deviations smaller than 50 cm were considered for RMSE and mean. The locations of the different image sub-sequences (e.g., 1-1 or 2-1) used in the following experiments are indicated in Figure 4a.

In a first series of tests, relative depth comparisons in image space were carried out with the depth map of configuration 4 (c4) as a reference (Figure 9). For all extracted 3D base images, c1–c4 delivered the lowest RMSE values. The highest RMSE, as well as mean values, were computed for c2–c4. While RMSE values for c1–c4 and c3–c4 are in the range of 36 mm to 56 mm, the range for c2–c4 is from 57 mm to 72 mm. c3–c4 delivered the most façade points and c2–c4 shows an opposite behavior compared to the two other configurations with inverted depth differences. In c2–c4 and c3–c4 the region close to the epipole, where depth estimations are not accurate and thus eliminated, is clearly visible. This effect results in a considerable number of road surface points not being mapped.

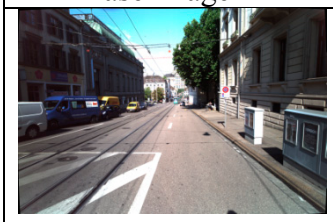
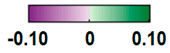
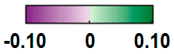
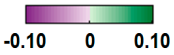
Matching configurations: Base image	c1–c4	c2–c4	c3–c4
 	 RMSE: 53 mm Mean: −7 mm	 RMSE: 72 mm Mean: 9 mm	 RMSE: 56 mm Mean: 4 mm
 	 RMSE: 36 mm Mean: −10 mm	 RMSE: 57 mm Mean: 22 mm	 RMSE: 38 mm Mean: 1 mm
 	 RMSE: 53 mm Mean: −6 mm	 RMSE: 64 mm Mean: 19 mm	 RMSE: 54 mm Mean: 3 mm

Figure 9. Deviations of depth maps generated by the SURE triangulation module.

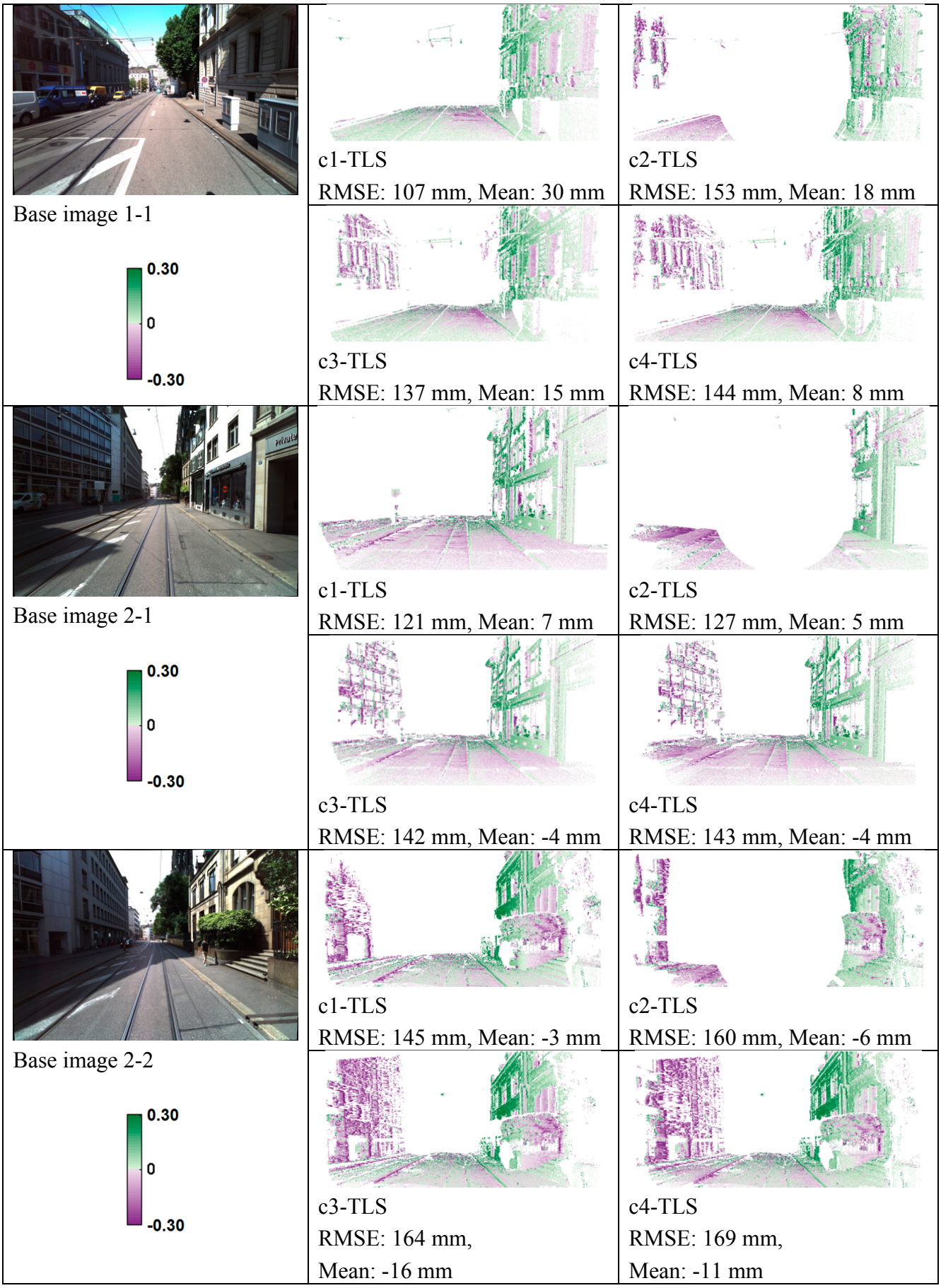


Figure 10. Depth deviations between point clouds from SURE triangulation and TLS.

In a second test series, terrestrial laser scanning points projected to image space were used as reference. These reference depth maps (TLS) were compared with dense image matching point clouds for all matching configurations (c1–c4), all generated using the SURE triangulation module. The highest mean values, *i.e.*, the highest depth offsets, were computed for base image 1-1 (see Figure 4a). The highest RMSE, *i.e.*, the highest depth noise values, were observed for base image 2-2, which were caused by a large shadow area and vegetation (see Figure 10). All RMSE values are in the range of 107 to 169 mm. While c1-TLS features the lowest RMSE values, c2-TLS shows the highest RMSE values for all but for base image 2-1, where significantly fewer points are mapped in comparison to the other configurations. RMSE values for c4-TLS are a little higher than for c3-TLS but c4-TLS also includes more depth observations.

6.3. Discussion

Investigations in image space showed similar results between the four selected configurations for all three base images. Due to a single large stereo base, the traditional stereo Configuration 1 provided high accuracies, which did not further improve with the additional use of images captured at different epochs. Configuration 2 with sequential matching of mono image sequences loses many points around the epipole. It also yields a limited accuracy since the base for image ray intersection is very small. The differences between Configurations 3 and 4 are not significant. However, especially compared to the standard stereo Configuration 1, the increasing number of match images available in Configuration 4 delivers significantly higher point densities. In summary, the traditional stereo matching Configuration 1 delivers depth maps with a medium completeness but with the highest accuracy and Configuration 4 yields depth maps with the highest completeness but slightly higher RMSE values. For a more detailed discussion of the multi-view configurations and the underlying epipolar rectification for in-sequence matching we refer the reader to our publication [31]. A further depth map improvement in terms of completeness and reliability can be expected from the future incorporation of the imagery from the back-right and left stereo systems as well as from the panorama camera.

7. Smart Exploitation of Cloud-Based 3D Image Spaces

In the introduction of 3D image spaces in Section 3, we postulated some key capabilities such as a high-fidelity WYSIWYG representation of the urban environment that shall be easy-to-use, support robust and accurate 3D measurements, as well as a simple and efficient augmentation of geospatial contents. In this section, we demonstrate some of these capabilities by presenting and discussing a selection of features offered by our cloud- and web-based software framework introduced in Section 4. Since the client application of the software framework is entirely web-based, running on every actual Internet browser, there is no need for a software installation on client computers. The same web application can even be deployed for different types of users within customer organizations. This could provide geospatial expert users with more advanced tools than, for example, users from different application domains or even public users. Expert-only functionality could include absolute 3D coordinate measurement, GIS editing tools or the augmentation of critical infrastructure not accessible to the public.

Key features and functionality offered by the urban model based on geospatial 3D image spaces include the following:

3D Monoplotting is the underlying core functionality, which enables accurate 3D measurements or the digitizing of points, lines or polygons (Figure 11a) simply by clicking on a location within the 2D imagery. Based on the exact georeference of each image and on its associated dense depth map, 3D world coordinates of every 2D cursor position are instantly calculated. With these 3D point coordinates, arbitrary relative measurements can be derived, e.g. distances, heights or areas as shown in Figure 11a,c,d. Monoplotting is a long-known photogrammetric procedure, which enables 3D digitizing and 3D feature extraction from single images where an underlying depth map representing the surface of the represented scene is available. The principle of digital 3D monoplotting dates back to the early 1970s and was originally applied to combinations of aerial imagery and digital elevation models (DEM). Later on it was extended to satellite imagery [42] and to the combination of close-range photogrammetry and TLS point clouds [43], where 3D monoplotting has become a standard measuring principle. Today, 3D monoplotting is also used in hybrid image- and LiDAR-based mobile mapping environments [7,8,11] allowing image-based monoscopic 3D measurements in combination with co-registered MLS point clouds. However, these hybrid solutions currently cannot guarantee the spatial and temporal coherence of the image and depth information, which would be required for both accurate and robust 3D measurements and which currently can only be provided by stereovision based mobile mapping.

Augmentation of Geospatial Contents—For the inspection and updating of existing geospatial content, such as infrastructure data, the web client provides functionality to load and store such data from or to a file or to interface with existing databases (Figure 11b). For data exchange, GIS standard geometry formats like GeoJSON or WFS services are supported. In the regular case, where existing infrastructure data is available in 2D only, *i.e.* without known height information, the original 2D geometry can be projected into 3D by means of a server-side process. This process again uses the dense depth maps. To increase the height accuracy, multiple 3D images are taken into account for this calculation. This process runs fully automatically and allows determining the missing height for huge existing 2D infrastructure databases for their accurate integration into the 3D image spaces.

Simple "One-Click" Measurements—In the context of street level urban environments and road management the height of objects above the terrain is an important measure. For accurately measuring such heights, a special "one click" height measurement tool was developed. As shown in Figure 11c, the user clicks on a point above the ground e.g. on a traffic light or on an overhanging tree branch. Based on this 3D position an exact vertical line in the map reference system is defined. The intersection of this vertical line with the depth map representing the underlying ground automatically determines the ground point for the vertical height measurement. The calculation is similar to the one used for assigning heights to an existing 2D infrastructure dataset, as described in the last section. The main difference is that the tool is running on client side. The same functionality is also used for automatically extracting longitudinal road profiles or cross-sections.

Figure 11. Overview of selected tools and features for interacting with 3D image spaces. (**a**) Measuring a polygonal area by 3D monoplotting; (**b**) superimposing existing infrastructure data e.g. water (blue) and waste water (red) pipes in an urban street scene; (**c**) single-click height measurement from a traffic light to the road surface; (**d**) measuring of a perpendicular distance (red line) from an orthogonal reference line (green) extracted from pavement border; (**e**) resulting images of a multi-view query looking for a 3D world coordinate (red square) under different viewing angles; and (**f**) mobile web application.

Intelligent Advanced Measurements—In urban infrastructure management it is often necessary to perform more sophisticated measurement tasks, such as determining exact orthogonal distances from a defined edge or axis (see Figure 11d). For this purpose, a special orthogonal measurement tool was created, which simultaneously exploits the radiometric and depth information of the 3D image. In a first step, the tool searches for edges at the user's current cursor position by using the radiometric image information. A subset of pixels around the cursor is analyzed by a restricted partial Hough transform [44]. The generated Hough space is then thresholded by a predefined limit to find clear Hough peaks. If no peak is found, then the threshold limit decreases iteratively, controlled by the duration a user presses the mouse button. This ensures that small intensity changes of edges are still recognized as a possible line. If a clear radiometric edge is found, a 3D edge is estimated using the co-registered depth information. The resulting 3D line is then set as reference line. By further clicking in the image, orthogonal distances to this reference line can be measured as shown in Figure 11d.

Multi-View Queries—Usually street level environments are captured in a multi-stereo configuration and in different driving directions. Often those streets are re-captured in a predefined interval, such as every year or every two years, to ensure actual data. This leads to a continuously growing huge 3D image database. To get the full benefit of these huge 3D image collections, intelligent image selection algorithms are necessary. One example is the efficient querying of all images containing a specific 3D world coordinate or object. This allows users to easily inspect an object from different viewing angles or to compare measurements in multiple frames. An implementation of such an algorithm in the multi-view tool is shown in Figure 11e. Future advanced multi-view queries on 3D image spaces could also be used for localization or augmentation tasks as shown in [45].

Mobile Measurements—The usage of a web application built with state of the art web technologies like HTML5/WebGL allows the software to run cross-platform on desktop or mobile devices (Figure 11f). Typical data streaming concepts were applied and performance in limited bandwidth environments was an important criterion during development in order to run the application also on mobile devices. As a result, users can access and exploit the 3D image spaces directly in the field, for example, for cross-check measurements, for checking where an existing underground pipe is located or even for staking-out tasks. The latter can be achieved by making cm-accurate 3D measurements between augmented infrastructure objects and natural features, such as road markings on the mobile client. These virtual measurements can then be staked out in the field, e.g. by using simple tape measures.

8. Conclusions and Future Work

In this paper we introduced the concept and an implementation of a new type of native urban model which we refer to as *geospatial 3D image spaces* and which is based on collections of georeferenced RGB-D imagery. A key requirement of the model is that the depth information (D) for each image shall be dense, as well as spatially and temporally coherent. This ensures that the urban model is WYSIWYG ("what you see is what you get"), *i.e.* that any visible object, including moving objects such as cars or pedestrians, can be correctly localized and measured in 3D. Multi-view stereo mobile mapping systems in combination with state-of-the-art dense image matching algorithms are capable of fulfilling these key requirements. With such systems, urban environments along road corridors, railway lines and even rivers can be captured with high fidelity and with a high level of accuracy. The urban model and the system

framework described in this paper are already operationally used by our spin-off company and research partner iNovitas AG for producing large-scale 3D image based urban models of entire cities and states.

In the paper we subsequently addressed three main research issues determining the performance and usefulness of 3D image spaces in real-world environments: (1) the obtainable georeferencing and subsequent absolute accuracy, (2) the density and relative measurement accuracy, and (3) specific exploitation features offered by the model.

First, we showed that the multi-view stereo imagery not only serves as metric 3D representation of the environment but that it can also be used for *significantly improving the georeferencing accuracy* in typical GNSS-degraded urban areas. In our experiments, the original direct georeferencing accuracy from high-grade inertial navigation equipment was in the order of one to several dm. With image-based georeferencing using a bundle adjustment of just the front stereo imagery and a number of GCPs, the *3D georeferencing accuracy* of the imaging sensors could be *improved by an order of magnitude* to less than 2 cm. It was also shown that largely automated image-based georeferencing is capable of compensating discontinuities in directly georeferenced trajectories, which were not addressed by earlier integrated georeferencing approaches.

Second, we presented a number of image matching strategies, which aim at obtaining *optimal depth maps* that are as dense and complete as possible and provide an optimal depth accuracy. As could be expected, the traditional single base stereo configuration provided the highest accuracy, but a limited completeness. A *new multi-image matching configuration*, which matches a base image with five spatially and temporally adjacent images using a modified polar rectification approach was introduced. It was shown that traditional stereo matching and the multi-stereo matching approaches yield far superior results to what can be obtained from matching monoscopic imagery, which is still the standard imagery with most commercial mobile mapping systems. This supports the idea of using stereo imagery for establishing high-quality urban 3D spaces. The evaluations further showed that the multi-stereo configurations yield depth maps with similar accuracies to traditional stereo matching but with a *significantly higher completeness* and robustness. These are valuable contributions towards creating *image-based urban models* not only with an unparalleled richness but also with a *reliable measuring capability at the cm accuracy level*—even in challenging urban environments.

Third, from a smart city perspective, the interesting aspects of 3D image spaces include their *intuitive interpretation* by geospatial experts and the general public alike as well as their ease of use by means of increasingly powerful web and mobile clients. A key element in the exploitation of 3D image spaces is the underlying *3D monoplotting* functionality. We demonstrated that the combination and extension of this well-known principle with accurately georeferenced high-resolution imagery, dense depth maps, and new algorithms are enabling a range of powerful, yet easy-to-handle new tools for interacting with 3D image spaces.

The potential for future work in image-based urban models in general and in 3D image spaces, in particular, is enormous. Our own work will focus on further improving the image-based georeferencing of 3D image spaces by incorporating imagery from all views into a new constrained bundle adjustment. Based on these results we will continue our research in optimal depth generation from multi-view stereo matching with the goal of significantly improving the density and accuracy of the depth information over what was presented in this paper. This work will, among other aspects, also include investigations on optimal along-track and cross-track baselines. Ongoing work in depth extraction also includes stereo

configurations with 360° coverage with the goal of preserving both the original image geometry and providing high depth extraction accuracies. Integrated and image-based georeferencing approaches, which are highly accurate and automated, will also be a prerequisite for creating accurate and large-scale indoor models based on 3D image spaces. In the longer term, multi-temporal large-scale 3D image spaces of entire cities or states will provide an ideal basis for urban change detection. However, due to the complexity and the vast number of rapidly changing objects in urban scenes, the automatic detection of 'slow' urban changes will remain a major scientific challenge for many years to come.

Acknowledgments

The authors would like to acknowledge the Swiss Commission for Technology and Innovation CTI (Project #158781. PFES-ES infraVIS) for their financial support of this research work. We would particularly like to thank our spin-off company and research partner iNovitas AG for their support of this project. We would also like to acknowledge the support by the additional team members Martin Christen, Robert Wüest, Kevin Hilfiker and Eric Matti for their contributions to the infraVIS project.

Author Contributions

Stephan Nebiker designed the overall research and wrote the majority of the paper. Stefan Cavegn designed, performed and analyzed the experiments on georeferencing and 3D extraction. Benjamin Loesch designed and contributed the section on the exploitation of 3D images spaces.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Albino, V.; Berardi, U.; Dangelico, R.M. Smart cities: Definitions, dimensions, performance, and initiatives. *J. Urban Technol.* **2015**, *22*, 3–21.
2. Hall, R.E.; Bowerman, B.; Braverman, J.; Taylor, J.; Todosow, H.; von Wimmersperg, U. The vision of a smart city. In Proceedings of the 2nd International Life Extension Technology Workshop, Paris, France, 28 Sepetember 2000.
3. Harrison, C.; Eckman, B.; Hamilton, R.; Hartswick, P.; Kalagnanam, J.; Paraszczak, J.; Williams, P. Foundations for smarter cities. *IBM J Res. Dev.* **2010**, *54*, 1–16.
4. Cretu, L.-G. Smart cities design using event-driven paradigm and semantic web. *Inform. Econ.* **2012**, *16*, 57–67.
5. Petrie, G. Mobile mapping systems—An introduction to the technology. *GeoInformatics* **2010**, *13*, 32–43.
6. NASA Mars Mapping Technology Brings Main Street to Life. Avaiable online: https://spinoff.nasa.gov/Spinoff2008/ct_9.html (Accessed on 28 July 2015).
7. Paparoditis, N.; Papelard, J.-P.; Cannelle, B.; Devaux, A.; Soheilian, B.; David, N.; Houzay, E. Stereopolis II: A multi-purpose and multi-sensor 3D mobile mapping system for street visualisation and 3D metrology. *Rev. Française Photogramm. Télédétection* **2012**, *200*, 69–79.

8. Anguelov, D.; Dulong, C.; Filip, D.; Frueh, C.; Lafon, S.; Lyon, R.; Ogale, A.; Vincent, L.; Weaver, J. Google street view: Capturing the world at street level. *Computer* **2010**, *43*, 32–38.
9. Lippman, A. Movie-maps: An application of the optical videodisc to computer graphics. In Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques, Seattle, WA, USA, 14–18 July 1980.
10. Ellum, C.; El-Sheimy, N. Land-based mobile mapping systems. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 13–17.
11. Puente, I.; González-Jorge, H.; Martínez-Sánchez, J.; Arias, P. Review of mobile mapping and surveying technologies. *Measurement* **2013**, *46*, 2127–2145.
12. Xiao, J.; Fang, T.; Zhao, P.; Lhuillier, M.; Quan, L. Image-based street-side city modeling. In Proceeding of the ACM SIGGRAPH Asia 2009, Yokohama, Japan, 16–19 Decemeber 2009.
13. Pollefeys, M.; Nistér, D.; Frahm, J.M.; Akbarzadeh, A.; Mordohai, P.; Clipp, B.; Engels, C.; Gallup, D.; Kim, S.J.; Merrell, P.; *et al*. Detailed real-time urban 3D reconstruction from video. *Int. J. Comput. Vis.* **2008**, *78*, 143–167.
14. Meilland, M.; Comport, A.I.; Rives, P. Dense omnidirectional RGB-D mapping of large-scale outdoor environments for real-time localization and autonomous navigation. *J. F. Robot.* **2015**, *32*, 474–503.
15. Nebiker, S.; Bleisch, S.; Christen, M. Rich point clouds in virtual globes—A new paradigm in city modeling? *Comput. Environ. Urban Syst.* **2010**, *34*, 508–517.
16. Musialski, P.; Wonka, P.; Aliaga, D.G.; Wimmer, M.; van Gool, L.; Purgathofer, W. A survey of urban reconstruction. *Comput. Graph. Forum* **2013**, *32*, 146–177.
17. Lafarge, F.; Mallet, C. Creating large-scale city models from 3D-point clouds: A robust approach with hybrid representation. *Int. J. Comput. Vis.* **2012**, *99*, 69–85.
18. Van Gool, L.; Martinovic, A.; Mathias, M. Towards semantic city models. In Proceedings of the 54th Photogrammetric Week, Stuttgart, Germany, 11–15 September 2013.
19. Grzeszczuk, R.; Kosecka, J.; Vedantham, R.; Hile, H. Creating compact architectural models by geo-registering image collections. In Proceedings of the 12th Computer Vision Workshops (ICCV Workshops), Kyoto, Japan, 27 September–4 October 2009.
20. Burkhard, J.; Cavegn, S.; Barmettler, A.; Nebiker, S. Stereovision mobile mapping: System design and performance evaluation. *Int. Arch. Photogram. Remote Sens. Spatial. Inform. Sci*. **2012**, *5*, 453–458.
21. Eugster, H.; Huber, F.; Nebiker, S.; Gisi, A. Integrated georeferencing of stereo image sequences captured with a stereovision mobile mapping system—Approaches and practical results. In Proceedings of the ISPRS—International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Melbourne, Australia, 25 August–1 September 2012.
22. Huber, F.; Nebiker, S.; Eugster, H. Image sequence processing in stereovision mobile mapping—Steps towards robust and accurate monoscopic 3D measurements and image-based georeferencing. In Proceeding of the Photogrammetric Image Analysis, ISPRS Conference, Munich, Germany, 5–7 October 2011.
23. Verbree, E.; Zlatanova, S.; Smit, K. Interactive navigation services through value-added CycloMedia panoramic images. In Proceedings of the 6th international conference on Electronic commerce, Delft, The Netherlands, 25–27 October 2004.

24. Swart, A.; Broere, J.; Veltkamp, R.; Tan, R. Refined non-rigid registration of a panoramic image Sequence to a Lidar point cloud. In Proceeding of the Photogrammetric Image Analysis, ISPRS Conference, Munich, Germany, 5–7 October 2011
25. Nebiker, S. Advances in imaging and photogrammetry. *Geospatial Today* **2012**, *11*, 12–16.
26. Cavegn, S.; Haala, N.; Nebiker, S.; Rothermel, M.; Tutzauer, P. Benchmarking high density image matching for oblique airborne imagery. *Int. Arch. Photogram. Remote Sens. Spatial. Inform. Sci*. **2014**, *3*, 45–52.
27. Nebiker, S.; Cavegn, S.; Eugster, H.; Laemmer, K.; Markram, J.; Wagner, R. Fusion of airborne and terrestrial image-based 3D modelling for road infrastructure management—Vision and first Experiments. *Int. Arch. Photogram. Remote Sens. Spatial. Inform. Sci*. **2012**, *4*, 79–84.
28. Chon, J.; Wang, J.; Ristevski, J.; Slankard, T. *High-Quality Seamless Panoramic Images*; InTech Open Access Publisher: Rijeka, Croatia, 2012.
29. Hirschmüller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341.
30. Geiger, A.; Roser, M.; Urtasun, R. Efficient large-scale stereo matching. In Proceedings of the 10th Asian Conference on Computer Vision, Computer Vision—ACCV, Queenstwon, New Zealand, 8–12 November 2010.
31. Evaluation of Matching Strategies for Image-based Mobile Mapping. Available online: http://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/II-3-W5/361/2015/isprsannals-II-3-W5-361-2015.pdf (accessed on 28 July 2015).
32. Cannelle, B.; Paparoditis, N.; Pierrot-Deseilligny, M.; Papelard, J.-P. Off-line *vs.* On-line calibration of a panoramic-based mobile mapping system. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *3*, 31–36.
33. Kersting, A.P.; Habib, A.F.; Rau, J.-Y. New method for the calibration or multi-camera mobile mapping systems. *Int. Arch. Photogram. Remote Sens. Spatial. Inform. Sci*. **2012**, *1*, 121–126.
34. Rau, J.-Y.; Habib, A.F.; Kersting, A.P.; Chiang, K.-W.; Bang, K.-I.; Tseng, Y.-H.; Li, Y.-H. Direct sensor orientation of a land-based mobile mapping system. *Sensors* **2011**, *11*, 7243–7261.
35. OpenCV Camera Calibration and 3D Reconstruction, StereoSGBM. Available online: http://docs.opencv.org/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html#stereosgbm (accessed on 28 July 2015).
36. Sure: Photogrammetric Surface Reconstruction from Imagery. Available online: www.ifp.uni-stuttgart.de/publications/2012/Rothermel_etal_lc3d.pdf (accessed on 28 July 2015).
37. Agisoft PhotoScan. Available online: http://www.agisoft.com/ (accessed on 28 July 2015).
38. Cramer, M.; Stallmann, D.; Haala, N. Direct georeferencing using GPS/inertial exterior orientations for photogrammetric applications. *Int. Arch. Photogramm. Remote Sens.* **2000**, *33*, 198–205.
39. Scharstein, D.; Szeliski, R.A Taxonomy and evaluation of dense two-frame stereo correspondence Algorithms. *Int. J. Comput. Vis.* **2002**, *47*, 7–42.
40. Pollefeys, M.; Koch, R.; van Gool, L. A simple and efficient rectification method for general motion. In Proceedings of the IEEE International Conference on Computer Vision, Kerkyra, 20–27 Sepetmber 1999.

41. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? The KITTI vision benchmark suite. In IEEE Conference on Computer Vision and Pattern Recognition, Providence, Rhode Island, 16–21 June 2012.
42. Single-Image High-Resolution Satellite Data for 3D Information Extraction. Available online: http://www.ipi.uni-hannover.de/fileadmin/institut/pdf/041-willneff.pdf (accessed on 28 July 2015).
43. Becker, R.; Benning, W.; Effkemann, C. 3D-monoplotting. kombinierte auswertung von laserscannerdaten und photogrammetrischen aufnahmen. *zfv Zeitschrift für Geodäsie, Geoinf. und Landmanag.* **2004**, *129*, 347–355.
44. Hough, P.V. Methods and Means for Recognizing Complex Patterns. U.S. Patent 3,069,654, 18 December 1962.
45. Zhang, J.; Hallquist, A.; Liang, E.; Zakhor, A. Location-based image retrieval for urban environments. In Proceedings of the Image Processing. (ICIP), 18th IEEE Internatinal. Conference, Brussels, Belgium, 11–14 September 2011.