Terrestrial LiDAR in Urban Data Acquisition

JAN BÖHM, Stuttgart

ABSTRACT

Terrestrial LiDAR plays an essential role in the acquisition of complete three-dimensional data for urban modeling. Especially the growing demand for detailed façade models drives the developments in acquisition and processing of terrestrial data. This paper reviews the past efforts in terrestrial data acquisition, which were mainly image based methods and gives a overview of the current state-of-the-art methods involving LiDAR data. Processing methods range from instantaneous visualization to complex modeling. Modeling approaches are described against the background of current efforts for standardized model categories. An outlook to future topics includes automated indoor modeling and latest sensor developments.

1. INTRODUCTION

Light Detection and Ranging (LiDAR) is a well-established technology for three-dimensional measurement of object surfaces. Aerial LiDAR has been used for over a decade to acquire highly reliable and accurate measurements of the earth's surface (Baltsavias, 1999). Today aerial LiDAR is used on a regular basis for the production of digital terrain models and the derivation of urban models. Terrestrial LiDAR systems entered the commercial domain some years later. However, the market has evolved dynamically over the past few years and systems are available from a number of vendors now. In a recent study by (Lemmens, 2007) nineteen systems from ten different manufacturers were listed. The number of available systems has grown since then.

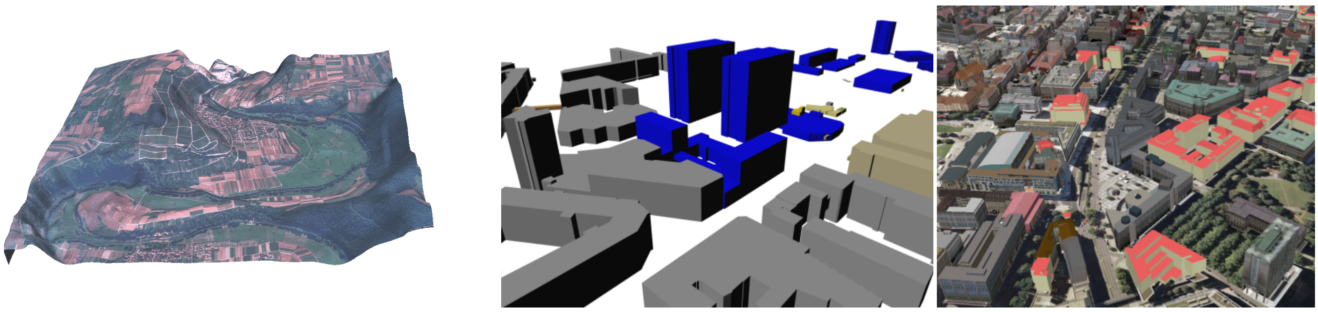
The growing interest in terrestrial data for urban data acquisition when compared to aerial data comes for one from the increased resolution due to a reduced standoff distance. However, a second aspect, which is probably even more important for the success of terrestrial data, is the different perspective of street-level data. When we look at Fig. 1 (a) we see the typical view of aerial data acquisition, which can be described as a bird's-eye-view. In many applications, when urban data is visualized, a completely different perspective is used. One example is given in Fig. 1 (b), where a planned architectural structure is visualized in 3D using existing urban data to give an impression of the interaction of the planned building to its surrounding. The building is clearly visualized from a pedestrian's viewpoint, not from a bird's-eye-view. This is not only true for typical applications in city planning, but can also be observed in many other applications such as modern navigation system using landmarks in 3D city models, augmented reality systems used for pedestrian navigation and many more. In order to create a realistic and detailed visualization from a pedestrian's view aerial data is not enough - street-level data clearly has to be integrated.

While data acquisition in terrestrial LiDAR has matured as mentioned above and can be used fully automated on a day-to-day basis, this level of automation is not maintained throughout the production pipeline. The full processing chain besides acquisition consists of registration, geo-referencing and modeling. For the registration and geo-referencing automation has been suggested in research. An overview of automated registration technologies is given by (Pfeifer and Böhm, 2008). While not all of these technologies are currently available from commercial vendors and ready for production purposes, we can easily foresee that full automation of registration will be at hand. The true bottleneck of LiDAR processing is in the modeling stage. This situation has already been pointed out in various scenarios, e.g. in (Böhm et al., 2006). Since modeling is highly application specific, we want to review in the following specifically the modeling problem using

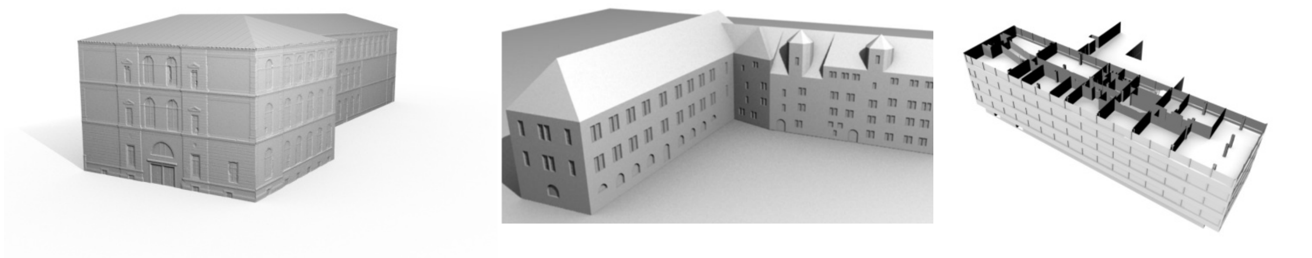
(a) (b)

Fig. 1: (a) A DMC image as a typical representative for the bird's-eye view of aerial data acquisition. Image is courtesy of Intergraph. (b) The visualization from a pedestrian view of an architectural design combining existing urban data and planned data (Burkert, 2000).

terrestrial data in the context of urban data. We will look at past efforts using terrestrial image data, summarize current suggestion using terrestrial LiDAR and have an outlook at future scenarios.
The attempt to generate realistic urban models has to be looked at against the background of current efforts for standardized model categories as described in the Open Geospatial Consortium (OGC) CityGML standard (Gröger et al., 2008). The model proposed by OGC supports five different levels of detail (LoDs) that provide a hierarchical description of building entities. Examples for the five levels are shown in Fig. 2. While the CityGML definition describes these levels according to their increasing geometric accuracy and detailedness, the five levels can also be seen as a historic development. In the context of this paper, we will specifically look at LoD2, which was a topic of past efforts to include terrestrial imagery for façade textures, LoD3 that contains three-dimensional façade structures and LoD4, which adds geometric information on the building's interior.



(a) (b) (c)



(d) (e)

Fig. 2: Examples for the five levels of detailed proposed by OGC's CityGML, from (a) LoD0 to (e) LoD4. Two possible examples are shown for (d) LoD3.

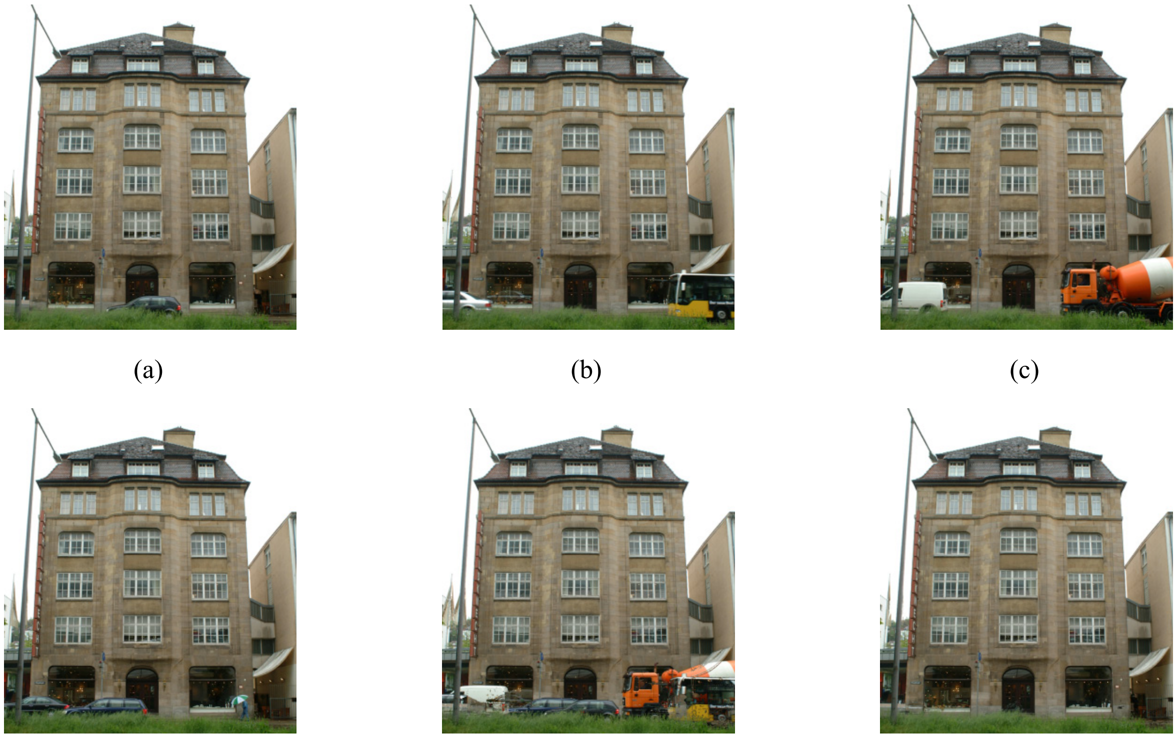
2. TERRESTRIAL IMAGERY – THE PAST

In the past, façade texturing has received some attention in both the photogrammetric and the computer vision community. Approaches for manual, semi-automatic and fully automated texture extraction have been presented. The simplest approach is to use a single image and warp this image onto the planar façade using the perspective transformation. Four points in both the planar coordinate system of the façade and in the image have to be determined. Assuming that all façades are bound by a rectangular curve, a common situation in LoD2 models, the approach can be further simplified. Since the target shape is a rectangle, the image can be transformed onto a unit square. This gives a proper perspective transform, but the width and height of the image are incorrect. The correct scaling will be applied to the cropped image at the rendering stage, when the image is mapped onto the corresponding façade. With this simplification, only the four corner points of the façade have to be identified in the image and no control information is required. The disadvantage of using only a single image is that any object occluding the façade will be mapped as well and thereby disturbing the texture image. To avoid these occlusions manual stitching of images is required. Despite these disadvantages, the approach has shown to be quite successful for LoD2 models and several hundred buildings have been textured using several thousand terrestrial images.

Many efforts in research and industry have been made to automate the removal of occluding objects, which can be either moving objects such as pedestrians and cars or static objects such as trees and street signs. Most of the approaches use the information from more than one image to distinguish disturbances from real texture. However, alternative approaches rely on the self-redundancy of architectural objects to obtain the same distinction. In the following, we will briefly review one approach that has been developed by the author, which is motivated by classical background detection from computer vision.

Considering the simplest case of moving objects, single-view multi-image fusion is appropriate to remove occluding objects completely. The proposed method works on a small number of images (e.g. four images) to keep the extra effort in data acquisition to a minimum. Since the number of images is small, we can process the full set of images at a time iterating over each pixel. We can think of the image sequence as a stack of per-pixel registered images, where each pixel exactly corresponds to the pixel of a different image at the same location. Thus, we can create a pixel stack for each pixel location. Using the basic assumption of background estimation that the background dominates over the altering foreground objects, we have to identify the subset of pixels within each pixel stack, which resembles the background. In other words, we have to identify the set of pixels, which forms the best consensus within each pixel stack. We can think of the problem as a clustering task, where we have to find the largest cluster of compatible pixels and disregard outliers. Fig. 3 shows an example for a single-view multi-image fusion. Four images from an input sequence show a façade observed from across a busy street, which is partly occluded by cars and pedestrians. The last two images show two synthetic images, which show the collection of all occluding objects and the occlusion-free facade, respectively. The method is extended to a multi-view fusion approach, when all the images of a stack are perspectively transformed onto the façade's plane. In this case, occlusions from static objects are removed as well. An in-depth explanation of the approach and further examples are presented in (Böhm, 2004).

As mentioned before, there exist alternative methods that solve the situation of replacing defective or occluded image parts from a single image. Prominent solution include inpainting, a technique, which is used to fill small, gaps, typically by propagating linear structures from the border into the defect area. A second approach is texture synthesis, which tries to copy repetitive texture to fill an occluded image area. There are many variations and also combinations of these methods, see for example (Criminisi et al., 2004). The exploitation of self-similarity, repetitive structures and symmetry of buildings and facades in particular have recently attracted great attention in both computer graphics and photogrammetry (Müller et al., 2006; Ripperda and Brenner, 2006). A



(a) (b) (c)

(d) (e) (f)

Fig. 3: (a)-(d) Four images from a fixed viewpoint of a façade imaged from across a busy street. Traffic and pedestrians partially occlude the lower portion of the façade. Synthetic image (e) shows all the occlusions combined. Synthetic image (f) shows the final result of the algorithm, where all occlusions are removed.

number of successful applications were developed, which exploit these properties (Müller et al., 2007). These approaches use grammars to store the repetitive pattern of elements. They have shown to be successful for synthesizing complete façades and for creating variations of facades.

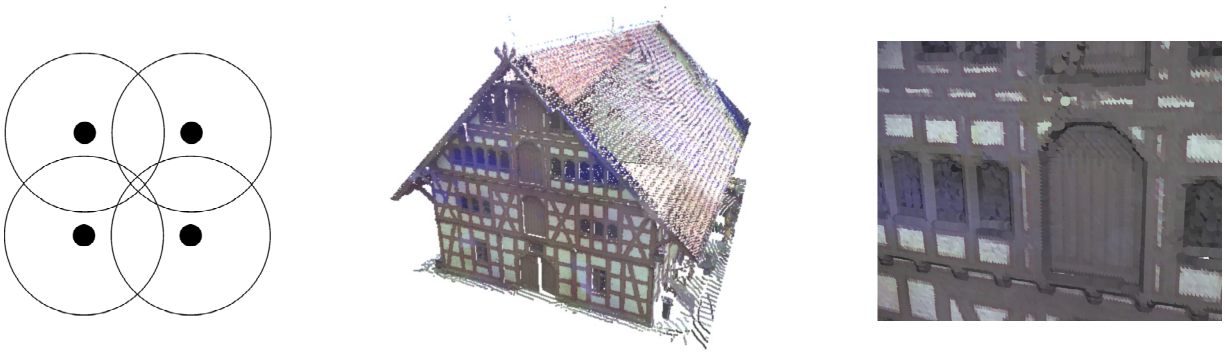
3. TERRESTRIAL LIDAR – THE PRESENT

Terrestrial laser scanners deliver a dense point-wise sampling of an object's surface. Depending on the instruments capabilities and the standoff distance, typically a sampling distance of a few centimeters down to a few millimeters is achieved. Such a point cloud, possibly colored when this information is available, can be directly visualized, e.g. to inspect its completeness and overall quality. Fig. 4 shows an example of such visualization. While the building displays fine from a far away viewpoint, we can observe that the impression of a solid front wall of the building cannot be maintained at a close viewing distance. It becomes clear that for a proper visualization the gaps in-between the point samples have to be closed. A typical approach to achieve this is mesh generation. However, for terrestrial data this process is not without problems and often requires manual intervention or post-processing. We therefore want to introduce an alternative method in the following section, which purely uses a point-based representation and is therefore ideally suited for point clouds.

Fig. 4: A colored point cloud acquired with terrestrial LiDAR. While the point cloud gives a good impression of the object from a far away view point, when we zoom in the impression of a solid surface collapses.

3.1. Direct Visualization using Point Splatting

In their survey on point-based techniques in computer graphics (Kobbelt and Botsch, 2004) raise the argument that while triangular meshes have proven to be a flexible and effective surface representation they become inefficient when the number of vertices becomes very large. When the number of triangles exceeds the number of pixels on the screen, most triangles cover less than a pixel and the rasterization of the triangles becomes extremely expensive. For the rendering of a point-based representation (Pfister et al., 2000) proposed surfels as rendering primitives also referred to as point splats. Each point is associated with a disk perpendicular to the normal vector and with a radius just large enough to cover the space to the neighboring points. This idea is illustrated in Fig. 5 (a). A point-based representation does not explicitly store the neighborhood relation of the points, but attempts to compute the neighborhood dynamically, typically using a k-nearest neighbor algorithm. This neighborhood relation is necessary for example to compute the surface normal in case it is not given and to compute the radius of each disk. These computations can be carried out extremely fast. Processing schemes exist for almost linear run-time behavior with respect to the number of input points. The method is thus capable of quickly processing large point clouds. The results are obtained almost instantaneous and without manual interaction. The results are ideally suited for quick visualization in the field, e.g. to decide on further data acquisition. Fig. 5 (b) shows the same dataset as in Fig. 4, but rendered as point splats. The difference becomes apparent when we zoom into the dataset. In Fig. 5 (c), we can see that the impression of a solid surface is maintained even at a very close viewing distance. While point splatting is an interesting approach for direct visualization, it is difficult to categorize the resulting model into the OGC's LoD hierarchy. Both LoD0 and LoD3 could be possible categories, yet both would be inappropriate at the same time. One might even argue if point splats form a model at all. To achieve proper LoD3



(a) (b) (c)

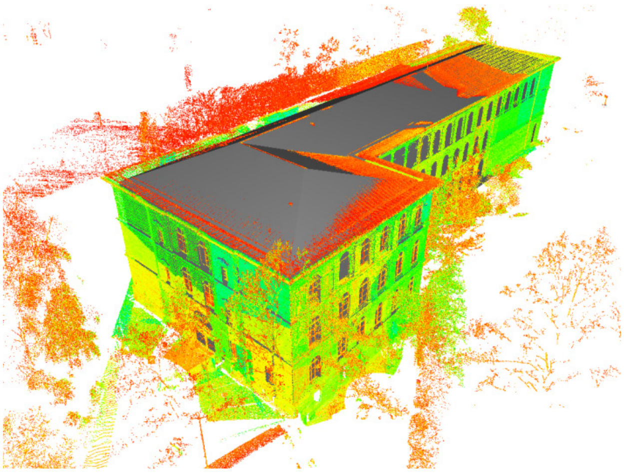
Fig. 5: (a) Each measured point is expanded into a disk large enough to cover the gap to the next surface element. (b)-(c) Point splatting can give the impression of a solid surface even at close viewing distance.

models the current standard procedure is to take a LoD2 model and augment it with terrestrial data and model the façade details on top of the existing model, which is described in the next section.

3.2. Façade Modeling Using LASERMAPs

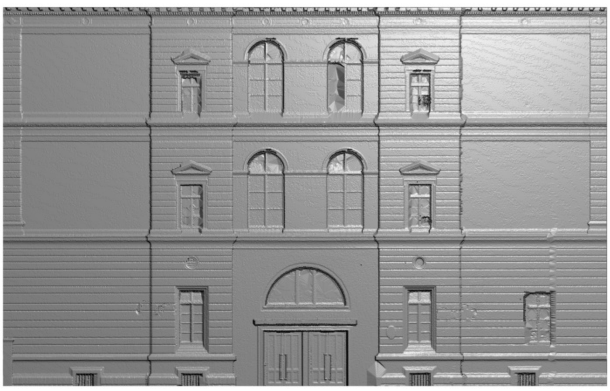
In order to be able to combine ground-based laser data with pre-existing building models, the data has to be registered. For the rest of this section we assume that the registration has been computed and the range data is given with respect to the same coordinate frame as the building model. There are many possibilities to compute the registration, ranging from direct georeferencing, to manual and automatic alignment, see for example (Schuhmacher and Boehm, 2005). This initial situation of a point cloud registered to a building model is shown in Fig. 6 (a). The dataset depicts the president's office at the Universität Stuttgart. The point cloud was acquired with a terrestrial laser scanner, a Leica HDS 3000, from more than 15 stations. The data covers the façades of the building at a point density better than 20 mm. The large number of stations was necessary to minimize shadowing of occluding objects.

Our method of modeling façades is motivated by concepts developed in computer graphics. In computer graphics, the duality of coarse over-all geometry and fine details has long been noted. The separation of the two is a fundamental modeling principle. Starting with the observations of (Blinn, 1978) that the effect of fine surface details on the perceived intensity is "primarily due to their

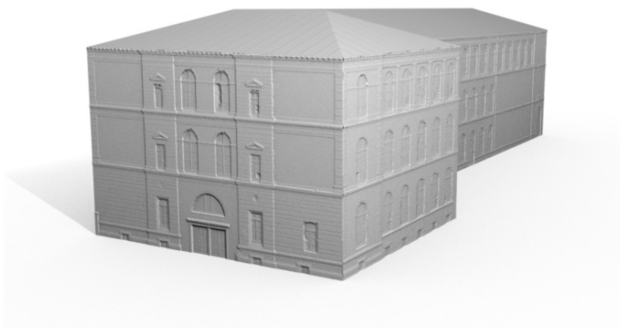


(a)

(b)



(c)



(d)

Fig. 6: (a) A prismatic building model with detailed roof structure and a registered point cloud acquired by terrestrial LIDAR. (b) A LASERMAP of a single facade derived from the point cloud and (b) the corresponding rendering as a normal map. (d) Rendering of the complete building model with displacement maps.

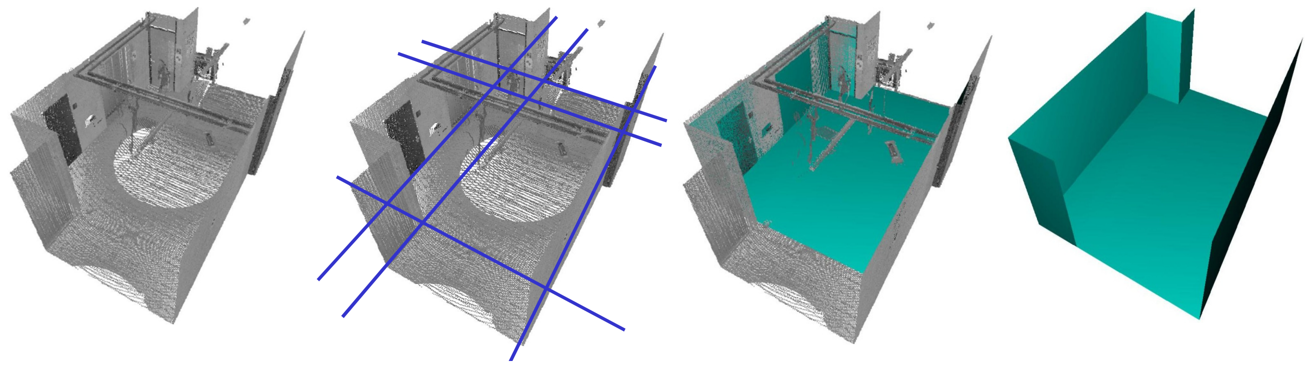
effect on the direction of the surface normal … rather than their effect on the position of the surface”, modeling concepts were developed, which keep fine surface details separate as a perturbation of the normal direction or a displacement to the underlying coarser geometry.
For a prismatic building model, the situation is rather simple, since the façades are typically planar polygons. We split the point cloud into groups with respect to the façades of the building using a simple buffer operation for each façade polygon. Each subset of the point cloud that is assigned to a particular façade is then interpolated into a regular raster. This procedure is very similar to digital elevation models derived from aerial LIDAR. We refer to such a re-interpolated point cloud as a LASERMAP (Böhm, 2005). The term is composed from two terms describing the source of the data, a laser scanner, and the use of the data as a source for 2D mapping. Fig. 6 (b) shows a LASERMAP of the front façade of the aforementioned building. The gray values correspond to offsets relative to the plane of the façade. The map was computed at a resolution of 10 mm to preserve details and each pixel stores the offset in 16 bits. Using an appropriate rendering algorithm a LASERMAP can be used to display the fine details of a façade. Fig. 6 (c) shows the rendering of a single façade using a normal map, which stores the perturbations of the normal vector at each pixel to model the variation of the surface rather than rendering the true 3D geometry. Fig. 6 (d) shows a rendering of the complete building model using displacement mapping, which uses true 3D geometry.

4. FUTURE TOPICS

While current efforts in the geomatics industry and in research focus on the generation of LoD3 models, it is obvious that the next step ahead is the production of LoD4 models. This requires the modeling of building interiors. The interest in accurate 3D representation of building interiors is twofold. For one there are the “classical” customers of geodetic products. The real estate industry requests indoor models for computer aided facility management. Public authorities can use accurate 3D models of building interiors to monitor critical structures. This includes disaster management, risk assessment and civil protection for example to plan and to monitor emergency routes and evacuation strategies. A second momentum is provided by the same driving forces that surround the success of the digital globes. Bill gates formulated a vision at the Internet Advertising Bureau’s Engage conference in 2005, where he would “be walking around in downtown London and be able to see the shops, the stores …” (Gibson, 2005). The idea to give stores a virtual presence in a digital globe opens a new advertisement market. This is one of the current business models behind the efforts to create digital globes. Potentially, digital globe providers could also participate in the revenue created by the virtual stores or their on-line counter parts.

4.1. Automated Indoor Modelling

The same arguments that were made for the automation of building exterior models can be brought forward for indoor models, perhaps with even stronger emphasis. Indoor scenes typically have an even higher complexity than façades, exhibiting more self-occlusion and clutter. This makes the task of creating an indoor model especially difficult. If a wide-spread dissemination of LoD4 models is demanded, automation is crucial, since it is clear that in such a complex case manual modeling is too time consuming and expensive. In the following we will review one approach for fully automated indoor modeling from terrestrial LiDAR that has been developed in cooperation with the author (Budroni and Böhm, 2009).



(a) (b) (c) (d)

Fig. 7: (a)-(d) Four stages of an fully automated approach to indoor modelling from terrestrial LiDAR. The ceiling is not displayed for better visibility.

The approach does not require any prior knowledge of the ground plan of the scene to be reconstructed. Instead, it just makes very basic and general assumptions on the LiDAR data and the scene, such as floor and ceiling being horizontal and walls vertical. A simple LiDAR data set of an indoor scene is shown in Fig. 7 (a). Typically, a large part of the work in manual modeling is spent on manual segmentation. In contrast, the proposed approach tries to decompose the point cloud automatically into meaningful sections. The first step is to identify floor and ceiling. As mentioned above we assume that the floor and ceiling are horizontal with respect to the terrestrial scanner's coordinate system. We use an algorithm known as plane sweeping to detect the floor and ceiling. Plane sweeping can be understood as a hypothesis and verification scheme. We assume the planes to be found are horizontal but we do not know at which height. Thus, we generate hypotheses for a plane at each discrete sample from a meaningful range of height values. This iterative generation of hypothesis planes can be interpreted as a sweeping motion and gave the approach its name. Each hypothesis is tested by checking how many points of the point cloud agree with it. This is done by checking how many points are closer than a certain distance to the plane. The two planes, which receive the largest consensus, are the floor and the ceiling.

Plane sweeping can obviously be used to detect all planar structures in the scene, so it can be used to detect the walls as well. The problem is, while we knew the normal vector for floor and ceiling (the vertical axis) and therefore the sweeping direction, we do not know the normal vector for the walls. Thus, before a linear plane sweep can be performed for the walls, we have to detect the dominant direction of the walls. This can be achieved with a variant of the algorithm, which we refer to as rotational sweep. The idea is very similar, as we create a hypothesis for each possible orientation of a vertical wall from discrete samples in an interval from zero to π. This is performed at a large number of randomly selected points. When only perpendicular walls are expected, the two dominant directions across all tests are chosen as the two major directions of walls. Now a linear plane sweep in these two horizontal directions can be performed to detect the walls.

The so computed positions for the walls are used as cut planes to decompose the space of the scene into cells. This situation is depicted in Fig. 7 (b), where the cut planes are represented as lines on the floor. In a further processing step, each of these cells has to be tested, whether it is part of the interior room or not. This can be done by checking if there are points on the ground or the ceiling within that cell. If there are no points, the cell is discarded. All remaining cells are collected and their bounding faces form the model of the room. The overlay of the bounding faces with the point cloud is shown in Fig. 7 (c) the pure model in Fig. 7 (d).

Fig. 8: A colored point cloud acquired with a V-Line terrestrial laser scanner from Riegl is shown on the left. The multi-pulse capability of the scanner facilitates 3D edge detection shown to the right, by selecting only first return pulses. LiDAR data is courtesy of RIEGL Laser Measurement Systems GmbH.

4.2. New Sensor Developments

So far, we have presented the developments in terrestrial LiDAR mainly from a modeling point of view. Obviously, there are also developments on the sensor side. The current trends are increased speed of acquisition, better accuracy, longer range and better handling of the sensors. While these trends are desirable features for practitioners, they do not have a major impact on the general use of terrestrial LiDAR for urban data acquisition. One feature that might drastically change the way we use terrestrial LiDAR in the future is multi target capability and full waveform analysis. While this feature has already been available in aerial LiDAR for a few years, it is relatively new in terrestrial applications. One possible way to exploit this feature might be automated edge extraction, which can be a major aid for modeling. Fig. 8 shows a colored point cloud of an indoor scene acquired with the new V-Line terrestrial laser scanner from Riegl, which provides online waveform processing (Pfennigbauer and Ullrich, 2009). When we display only those points that result from a multi target measurement and select only the first return pulse, we receive an automated detection of edge pixels without any additional data processing. These edges could be the starting point for refined modeling of building details both in- and outdoors.

5. SUMMARY

This paper gave an overview of the current use of terrestrial LiDAR in urban data acquisition. We have reviewed past efforts based on terrestrial imagery for LoD2 models. We have presented current technology using terrestrial LiDAR for LoD3 models. Automated indoor modeling for LoD4 models is identified as the next step for urban data acquisition. New developments in terrestrial LiDAR sensors might aid this new trend.

6. REFERENCES

Baltsavias, E. (1999). Airborne laser scanning: existing systems and firms and other resources. ISPR Journal of Photogrammetry & Remote Sensing **54**(2-3): 164-198.

Blinn, J. F. (1978). Simulation of wrinkled surfaces. Proceedings of the 5th annual conference on Computer graphics and interactive techniques, ACM Press.

Böhm, J. (2004). Multi-Image Fusion for Occlusion-Free Façade Texturing. IAPRS Vol. XXXV-B5, Istanbul, Turkey.

Böhm, J. (2005). Terrestrial Laser Scanning - A Supplementary Approach for 3D Documentation and Animation. Photogrammetric Week '05, Heidelberg, Wichmann.

Böhm, J., et al. (2006). Automation in laser scanning for cultural heritage applications. Recording, Modeling and Visualization of Cultural Heritage. Baltsavias, Gruen, Van Gool and Pateraki, Taylor & Franzis Group**:** 443-450.

Budroni, A. and Böhm, J. (2009). Toward automatic reconstruction of interiors from laser data. 3D-ARCH 2009: 3D Virtual Reconstruction and Visualization of Complex Architectures.

Burkert, G. (2000). city pool frankfurt, Entwurf am Institut für Grundlagen der Planung.

Criminisi, A., et al. (2004). Region filling and object removal by exemplar-based image inpainting. Image Processing, IEEE Transactions on **13**(9): 1200-1212.

Gibson, O. (2005). Gates unveils his vision of a future made of silicon. The Guardian.

Gröger, G., et al. (2008). OpenGIS City Geography Markup Language (CityGML) Encoding Standard, Open Geospatial Consortium.

Kobbelt, L. and Botsch, M. (2004). A Survey of Point-Based Techniques in Computer Graphics. Computers and Graphics **28**(6): 801-814.

Lemmens, M. (2007). Terrestrial Laser Scanners. GIM International **21**: 52-55.

Müller, P., et al. (2006). Procedural modeling of buildings. ACM Transactions on Graphics.

Müller, P., et al. (2007). Image-based procedural modeling of facades. SIGGRAPH '07: ACM SIGGRAPH 2007 papers, New York, NY, USA, ACM Press.

Pfeifer, N. and Böhm, J. (2008). Advances in Photogrammetry, Remote Sensing and Spatial Information Sciences: 2008 ISPRS Congress Book. Z. Li, J. Chen and E. Baltsavias, CRC Press. **7:** 169-184.

Pfennigbauer, M. and Ullrich, A. (2009). The New Riegl V-Line Laser Scanner's Higher Data Quality Using New Measurement Principles. 9th Conference on Optical 3-D Measurement Techniques.

Pfister, H., et al. (2000). Surfels: Surface Elements as Rendering Primitives. Siggraph 2000, Computer Graphics Proceedings, ACM Press / ACM SIGGRAPH / Addison Wesley Longman.

Ripperda, N. and Brenner, C. (2006). Reconstruction of Facade Structures Using a Formal Grammar and RjMCMC. DAGM-Symposium.

Schuhmacher, S. and Boehm, J. (2005). Georeferencing of terrestrial laser scanner data for applications in architectural modeling. 3D-ARCH 2005: Virtual Reconstruction and Visualization of Complex Architectures.