Proc. of the Intl. Conf. on Advances in Civil, Structural and Mechanical Engineering --CSM 2013

ISBN: 978-981-07-7227-7 doi:10.3850/ 978-981-07-7227-7_22

Modeling Façade Structures Using Point Clouds From Dense Image Matching

Dieter Fritsch, Susanne Becker & Mathias Rothermel

Abstract— **The automated generation of 3D computer models for urban planning and disaster management is an ongoing activity since about two decades. For 10-15 years aerial LiDAR seemed to be a key technology to solve the data collection problem for this purpose. Today, aerial photographs with nadir and oblique viewing directions overcome many obstacles and can be processed by dense image matching algorithms do deliver very dense point clouds outnumbering LiDAR point clouds. For some applications terrestrial images seem to be a low-cost data collection method delivering as well very dense point clouds for further processing. This paper is focused on terrestrial image data collection and its processing using dense image matching algorithms. To solve the point cloud modeling problem a formal grammar processing pipeline delivers structured building information of Level-of-Detail 3 (LoD3). This information might contribute to refine existing FEM software of Civil Engineering to consider critical building parts, to be monitored during some underground activities. Furthermore, the LoD3 model will help to increase building comfort in cities by adavanced insulation and noise filter measures.**

Keywords—**3D Computer Models, Level-of-Details, Image Data Collection, Dense Image Matching, 3D Point Clouds, Formal Grammars, LoD3-Modelling** ***(key words)***

I. Introduction

The automated generation of three-dimensional computer models for urban planning and disaster management is a main issue in several engineering disciplines. Geodetic engineering, including photogrammetry, computer vision and geoinformatics, is developing, piloting and offering hardware, software and processing pipelines to deliver 3D models from aerial flyovers, several terrestrial positions and mobile platforms moving along streets and rails. Laser scanning (LiDAR) and digital imaging are key technologies which sometimes compete with each other, but are quite often complementary [5].

The development of LiDAR has arisen from the discovery of lasers in 1960 and the rapid development of these unique light sources [3]. Today, LiDAR systems offer rapid data collection from airborne, terrestrial and mobile platforms. One main advantage of LiDAR is, besides the direct delivery of georeferenced 3D point clouds, its capability to penetrate vegetation by registering multiple or full waveform echoes. Furthermore it resolves echoes of wires and street furnitures very well which makes it most versatile for powerline mapping and mobile mapping applications. The interpretation of terrestrial laser scans and mobile mapping point clouds using a formal grammar to deliver structured building information is investigated in [1].

Digital imaging has arisen in the 1980s when first CCD video cameras were used in machine vision applications [9]. Since 2000 digital airborne camera systems have replaced film-based photogrammetry. Although first image matching algorithms have been invented in the 1980s their results could not compete with LiDAR point clouds at all. Since 2011 photogrammetry is having a renaissance by using Semi Global Matching algorithms [7, 8]. Today, digital camera systems are used in airplanes, helicopters, UAVs, but also in terrestrial applications and mobile mapping systems. Off-the-shelf still video cameras allow for low-cost and rapid data collection to be processed by dense image matching software [12]. Computer vision offers Open Source packages such as VisualSfM, Bundler, PMVS, and others to process the collected photos for orientation and point cloud delivery. This makes photogrammetric data acquisition even more attractive: Besides nadir and oblique views acquired by one photo flight the point clouds are very dense (GSD of 5cm can deliver 400 points per sqm). They can be textured and refined in follow-up processes. For terrestrial applications the “one-image-per-step strategy” [17] allows for sufficient overlaps and therefore delivers dense 3D point clouds as well. Therefore we can state, that dense image matching is an emerging technology delivering point clouds outnumbering LiDAR point clouds in many respects.

In order to offer structured building information, which is useful for architectural and civil engineering applications, we are focusing on the Level-of-Detail 3 scale (LoD3). This standardized building representation contains besides the 3D silhouette and the true roof shape all details of the facades, e.g. doors, windows, stairs, balconies, etc. As we have learnt from some accidents in the past, for example the collapse of the Cologne City Archive in 2009, it is really helpful and necessary to refine the computer models integrating much more detailed building information as just using some block models.

Dieter Fritsch, Institute for Photogrammetry, Univ. Stuttgart *(Author 1)*
Address: Institute for Photogrammetry, University of Stuttgart, Geschwister-Scholl-Str. 24D, D-70174 Stuttgart
Germany
dieter.fritsch@ifp.uni-stuttgart.de

Susanne Becker, Institute for Photogrammetry, Univ. Stuttgart *(Author 2)*
Address: see D. Fritsch
susanne.becker@ifp.uni-stuttgart.de

Mathias Rothermel, Institute for Photogrammetry, Univ. Stuttgart *(Author 3)*
Address: see D. Fritsch
mathias.rothermel@ifp.uni-stuttgart.de

Proc. of the Intl. Conf. on Advances in Civil, Structural and Mechanical Engineering --CSM 2013
Copyright © Institute of Research Engineers and Doctors. All rights reserved.
ISBN: 978-981-07-7227-7 doi:10.3850/ 978-981-07-7227-7_22

The paper is about to demonstrate the potential of photogrammetric image data processing using dense image matching and point cloud interpretation by using formal grammars. Therefore the main contents are described as follows: First, a general overview is given using formal grammars based on Lindenmayer systems to solve the modeling problem of architecture. Secondly, image-based point cloud generation is outlined starting with image data collection using an off-the-shelf digital SLR camera. The data collection is carried out according to the "*one-image-per-step*" strategy. As we use imagery to derive high density point clouds the camera has to be calibrated and oriented, performed by the VisualSfM software package. The ifp software package SURE (SUrface REconstruction Using Imagery) [13] delivers dense point clouds of superb quality. Afterwards, a post-processing is carried out to prepare the point clouds for the grammar interpretation. Section IV describes the grammar-based façade modeling and concludes with the grammar application for the building chosen for this demonstration. A conclusion is summarizing the main steps and results presented in this paper.

II. Formal Grammars for the Modeling of Architecture

Usually formal grammars are applied during object reconstruction to ensure the plausibility and the topological correctness of the reconstructed elements. A famous example for formal grammars is given by Lindenmayer-systems (L-systems). Originally used to model the growth processes of plants, L-systems serve as a basis for the development of further grammars appropriate for the modeling of architecture. For instance, [11] produce detailed building shells without any sensor data by means of a shape grammar.

In our application, a formal grammar will be used for the generation of façade structures where only partially or no sensor data is available. In principle, formal grammars provide a vocabulary and a set of production or replacement rules. The vocabulary comprises symbols of various types. The symbols are called *non-terminals* if they can be replaced by other symbols, and *terminals* otherwise. The non-terminal symbol which defines the starting point for all replacements is the *axiom*. The grammar's properties mainly depend on the definition of its production rules. They can be, for example, deterministic or stochastic, parametric and context-sensitive. A common notation for productions which we will refer to in the following sections is given by

$$id: lc < pred > rc : cond \rightarrow succ : prob \,.$$

The production identified by the label *id* specifies the substitution of the predecessor *pred* for the successor *succ*. Since the predecessor considers its left and right context, *lc* and *rc*, the rule is context-sensitive. If the condition *cond* evaluates to true, the replacement is carried out with the probability *prob*. Based on these definitions and notations we develop a façade grammar which allows us to synthesize new façades of various extents and shapes. The axiom refers to the new façade to be modeled and, thus, holds information on the façade polygon. The sets of terminals and non-terminals, as well as the production rules are automatically inferred from the reconstructed façade as obtained by a data driven reconstruction process. Originally, our algorithm has been developed for the interpretation of 3D point clouds gathered by terrestrial laser scanning. However, as will be shown in the following sections, our approach is also appropriate to be applied to point clouds that stem from dense image matching.

Existing systems for grammar based reconstruction of building models which derive procedural rules from given images or models still resort to semi-automatic methods [4], [8] and [15]. In contrast, we propose an approach for the automatic inference of a façade grammar in the architectural style of the observed building façade.

III. Image-Based Point Cloud Generation

The image-based façade reconstruction approach presented in this chapter can be separated in three main steps. In a first step imagery of the façade has to be collected (section A). Thereby the object should be captured with high redundancy to guarantee the completeness of the reconstruction as well as good precision. Within the next step lens distortion and relative exterior and interior orientations of each image is computed using structure from motion techniques (section B). The relative exterior orientations represent position and rotation of the single image views in a common, arbitrary Euclidian coordinate frame. The interior orientation represents camera specific parameters as focal lengths and principle points. Based on this information, 3D façade points are densely extracted from the 2D image collection. In this last step 3D information of virtually each pixel in each of the incorporated images is computed (section C). This results in clouds providing immense density. However, for subsequent grammar related processing stages this amount of data is impracticable. Thus, some post-processing of the generated point clouds is required (section D).

A. *Data Collection*

For the image based façade reconstruction 95 images were captured from 33 camera stations using a Nikon D7000 SLR and 20mm fixed focal length lenses. To guarantee sufficient coverage at each viewpoint three images with slight variances in viewing directions were taken. Each image possesses a resolution of 16.2MP. Images of the façade (~50m wide, ~20m high) were collected in less than 15 minutes without using a tripod. A visualization of the captured façade is shown in Figure 1, sample images are displayed in Figure 2. If possible, images were collected in steps of ~1.5m at a distance of ~14.5m to the façade. This guarantees similar image content of images from proximate stations which facilitates the completeness of reconstructions. Assuming the image sensor being oriented parallel to the façade implies image overlaps up to 87% at object sampling distance less than 3.5mm.

Proc. of the Intl. Conf. on Advances in Civil, Structural and Mechanical Engineering --CSM 2013
Copyright © Institute of Research Engineers and Doctors. All rights reserved.
ISBN: 978-981-07-7227-7 doi:10.3850/ 978-981-07-7227-7_22

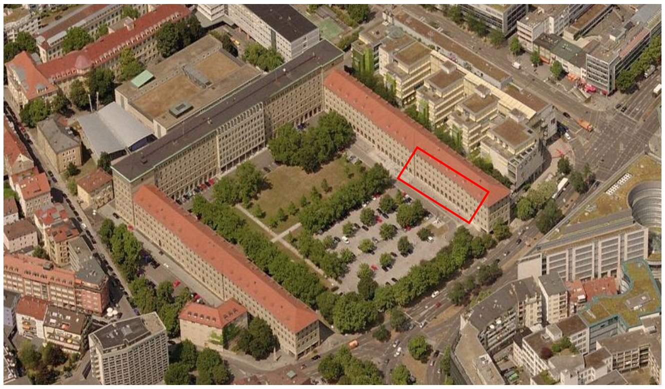


Figure 1. Birdseye view on the “Rotebühlbau, Stuttgart” (source: Bing Maps). Red rectangle marks the façade part for which images were collected and 3D reconstruction was performed.

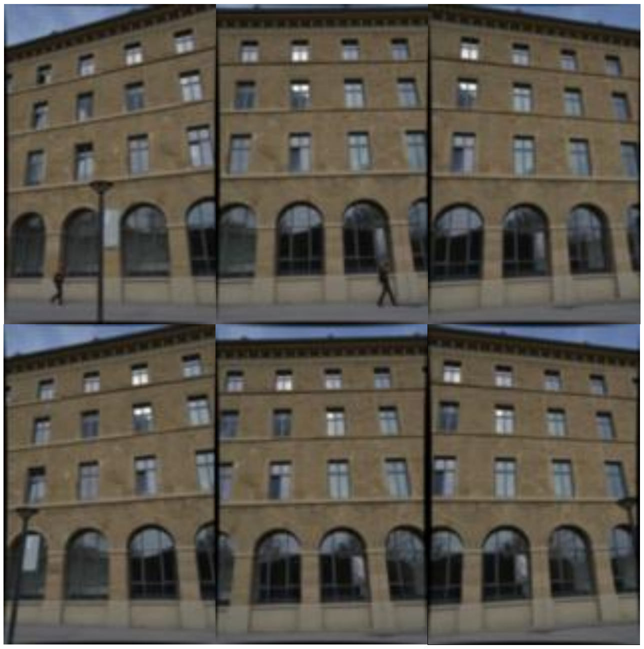


Figure 2. Images captured from proximate camera stations, at each station 3 images were collected with slight variances in viewing directions.

B. *Orientation Computation*

Lens distortions and orientations were computed using the Visual Structure from Motion (VSfM) software package [17]. It implements a structure from motion approach deriving information of image subsets which capture a common surface area (connectivity between the images). Thereby homologous image points across the views are derived and used in a subsequent and global bundle adjustment. Within this optimization step parameters modeling the radial distortion and all orientations are derived. As a side product of the bundle adjustment a sparse point cloud is obtained (Figure 3). Note that these results are arbitrary in scale. However, an explicit scale had not to be applied since the subsequent grammar extraction does not rely on metric coordinates at all.

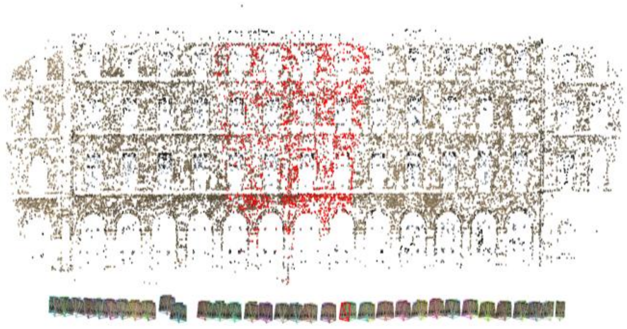


Figure 3. Sparse point cloud and camera poses generated by Visual Structure from Motion.

C. *Dense Matching and Multi-View Triangulation*

Using the before derived image orientations, dense 3D surface points were generated. Therefore the SURE [13] software package was utilized. It is based on *libtsgm* which implements a memory and time efficient modification of SGM [7] and subsequent multi-view structure computation.

SGM is a dense stereo image matching strategy which aims for finding corresponding pixels (representing the same object point) across two views. Ideally for each pixel in a reference image a correspondent pixel (representing the same world object) in a second view is found. The core idea of the method is minimizing a smart approximation of a global cost function, which represents the goodness of the alignment of all pixels across two views. This global cost function is composed of pixel-wise costs or intensity-based similarity measures as Census correlation, Mutual Information [16] or the Daisy descriptor [14]. Pixel-wise costs represent the similarity of two correspondent pixels or their surrounding pixel patches. Moreover, a penalty term is included solving for ambiguity of pixel-wise costs and forcing the smoothness of extracted surfaces. The key features of the algorithm are dense surface reconstructions providing preservation of depth steps at edges and moderate computation times. Within the classical

Copyright © Institute of Research Engineers and Doctors. All rights reserved.
ISBN: 978-981-07-7227-7 doi:10.3850/ 978-981-07-7227-7_22

approach the global cost function involves pixel-wise costs for each pixel in the reference image and a constant number of potential correspondences in the second view. SURE/libtsgm implements a hierarchical approach. Based on the surface estimations on low resolution imagery (for which processing is fast) the search of potential correspondences is limited to small ranges close to depth estimations from previous pyramid levels. This implies reduced times and memory consumption and enables matching imagery of scenes possessing large variances in depth. For the façade data set discussed in this paper 405 stereo models were calculated. Thus, each image was matched against more than 4 proximate images in average. The processing time amounted 2.75h (i7, 4 x 3.4 GHz) including input and output operations.

Since images are overlapping to a high degree, a specific façade area is mapped to multiple of the incorporated images. Within the dense matching stage images are pair-wise matched and for each of the pairs the 3D surface information is extracted. As a result multiple depth estimations of one specific surface patch is available. The redundancy is exploited by our multi-view structure computation implementation. This leads to increased precision of the generated surfaces [6][12]. Even more important erroneous depth estimations as propagated from dense matching module can be reliably detected and eliminated by checking for geometric consistency. As a result generated point clouds possess very low number of blunders. For the processed façade 405 stereo models were fused resulting in one point cloud for each of the 95 images. In total about 345 million surface points were reconstructed (Figure 4). The processing time amounted 1.33h (i7, 4 x 3.4GHz) including input and output operations.

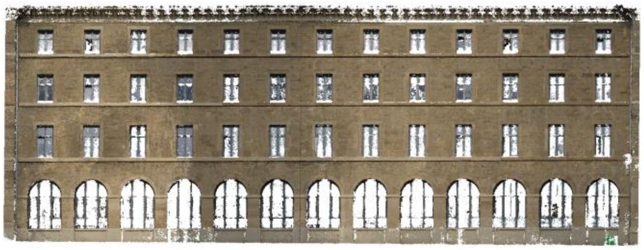


Figure 4. Visualization of the 95 dense point clouds generated by SURE.

D. *Post Processing Point Clouds*

The implemented algorithm for grammar extraction was optimized for façade points captured by LiDAR.

As will be discussed in chapter IV the algorithm bases on the assumption that window areas are represented by holes in the LiDAR point cloud. This assumption is justified by the fact that LiDAR pulses usually penetrate glass and, thus, do not lead to point measurements in window regions. In comparison to LiDAR point clouds, image-based results contain many points representing window panes. However, these points are located significantly behind the façade plane and can be easily identified and removed. For more sophisticated façade geometries alternative approaches would have to be utilized. Beside filters based on the radiometric information, elimination based on the local noise level are thinkable. Quality of dense matching heavily depends on the texture. Since windows provide weak texture, a comparable high noise level is expected. By eliminating the points measured in window areas, the assumption that window areas are represented by holes in the point cloud is fulfilled.

As discussed in section III.B the reference frame of generated point clouds so far possess arbitrary scale. Since the subsequent algorithm relies on metric information the correct scale has to be applied. Therefore dimensions of ground level window frames were measured on the test site. Since large parts of the window frames are available in the reconstructed points same window frame dimensions could be measured in the point clouds. Comparison leads to the correct scale factor which then was applied to all point coordinates.

A further post processing step concerns the density of the point clouds. In comparison to LiDAR data sets, the point clouds generated by dense image matching are rather dense. For the grammar extraction this density is not required and leads to increased processing time and memory demands. Therefore the results were spatially sub-sampled resulting in minimal point distances of 5cm. The resulting cloud, visualized in Figure 5, is now the starting point for the grammar based generation of a 3D façade model (see chapter IV).

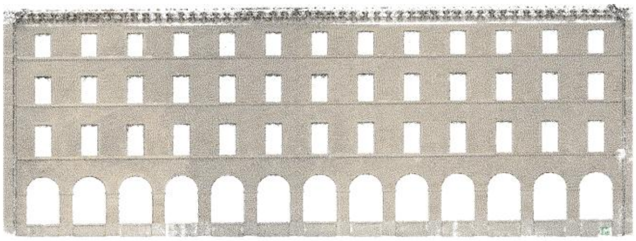


Figure 5. Post processed point cloud after eliminating window points and down sampling (5cm point distance).

IV. Grammar Based Facade Modeling

Our algorithm starts with the data driven extraction of façade structures from 3D point clouds (section A). Here, we apply point clouds generated by dense image matching (described in chapter III). After this interpretation step, the resulting reconstructed façade serves as a knowledge base for further processing (section B). Dominant or repetitive features and regularities as well as their hierarchical relationship are detected from the modeled façade elements. At the same time, production rules are automatically inferred. The rules together with the 3D representations of the modeled façade elements constitute a formal grammar which we will call *façade grammar*. It contains all the information which is necessary to reconstruct façades in the style of the respective building during knowledge based modeling (section C).

ISBN: 978-981-07-7227-7 doi:10.3850/ 978-981-07-7227-7_22

A. *Point Cloud Interpretation*

The approach for the data driven façade reconstruction aims at refining an existing coarse building model by adding 3D geometries to the planar façades [2]. Windows, doors and protrusions are extracted from the point cloud by searching for holes in those points that lie on the façade plane. The modeling process applies a 3D object representation by cell decomposition. The idea is to segment an existing coarse 3D building object with a flat front face into 3D cells. Each 3D cell represents either a homogeneous part of the façade or a window area. After classifying the 3D cells into window cells and façade cells, the window cells are eliminated while the remaining façade cells are glued together to generate the refined 3D building model. The difficulty is finding planar delimiters from the point cloud that generate a good working set of cells. Since our focus is on the reconstruction of windows, the delimiters have to be derived from 3D points at the window borders. For the exemplary dataset “Rotebühlbau, Stuttgart”, Figure 6 depicts the coarse building model with the aligned 3D points from dense image matching.

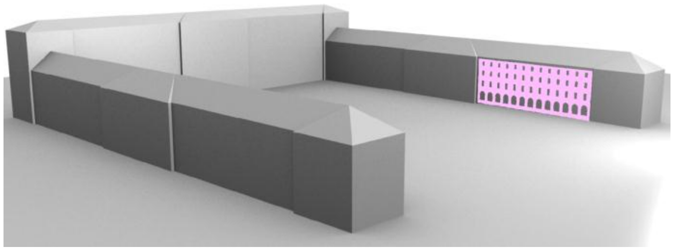
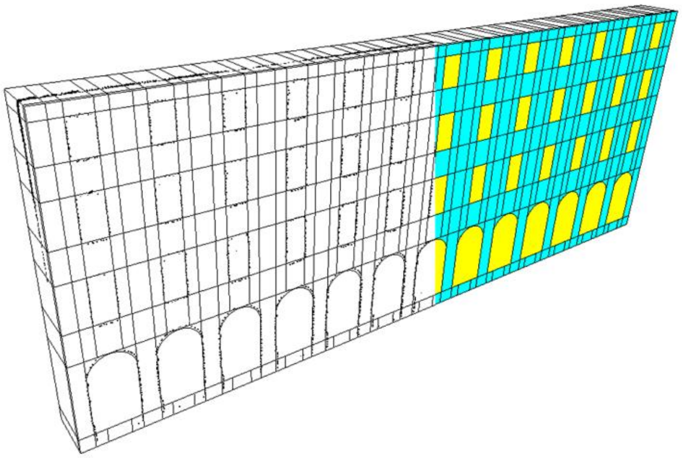


Figure 6. Rotebühlbau, Stuttgart: coarse building model and 3D point cloud from dense image matching.

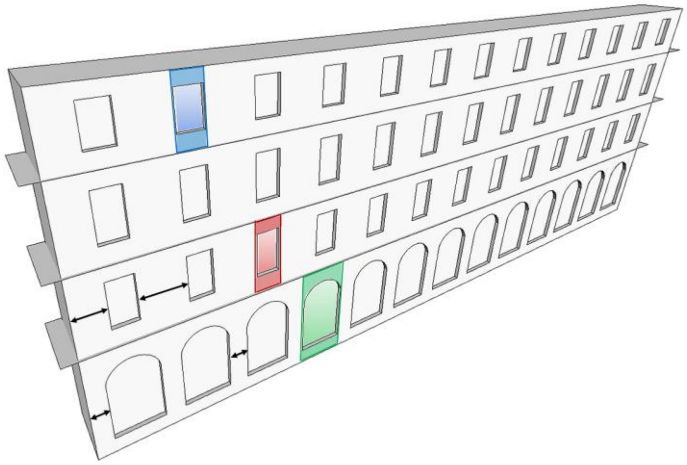
Point Cloud Segmentation. Usually, glass panes of windows are slightly set back from the façade plane. Based on this assumption, the 3D point cloud can be segmented by analyzing their distance from the façade plane (see section III.D). If only the points are considered as *façade points* that lie on or in front of the facade, the windows will describe holes within the points interpreted as façade points (see Figure 5). These no-data areas can be used for the point cloud segmentation which aims at the detection of window edges. For example, the edge points of a left window border are detected if no neighbor points to their right side can be found in a predefined search radius. In a next step, horizontal and vertical lines are estimated from non-isolated edge points. The left façade part in Figure 7(a) shows the extracted edge points at the window borders as well as the derived horizontal and vertical lines. Based on these window lines, planar delimiters can be generated for a subsequent spatial partitioning. Each boundary line defines a partition plane which is perpendicular to the facade. For the determination of the window depth, an additional partition plane can be estimated from 3D points representing the window panes. These points are detected by searching a plane parallel to the facade, which is shifted in its normal direction. The set of partition planes provides the structural information for the cell decomposition process. It is used to intersect the existing building model producing a set of small non-overlapping 3D cells.

Classification. In order to classify the 3D cells into facade and window cells, a point-availability-map is generated. It is a binary image with low resolution where each pixel defines a grid element on the facade. The optimal grid size is a value a little higher than the point sampling distance on the facade. Grid elements on the façade, where 3D points are available produce facade pixels; grid elements that belong to no-data areas are represented by non-facade pixels. The classification is implemented by computing the ratio of facade to non-facade pixels for each 3D cell. Cells including more than 70% facade pixels are defined as facade solids, whereas 3D cells with less than 10% facade pixels are assumed to be window solids. While most of the 3D cells can be classified reliably, the result is uncertain especially at window borders or in areas with little point coverage. However, the integration of neighborhood relationships and constraints concerning the simplicity of the resulting window objects allows for a final classification of such uncertain cells. The right façade part in Figure 7(a) shows the classified 3D cells: facade cells (cyan) and window cells (yellow). While the windows – as a result of plane partitioning – naturally are represented by polyhedral cells, also curved primitives can be included in the reconstruction process. This is demonstrated by the round-headed windows in the ground floor. Details about this refinement process can be found in [1].



(a)

Proc. of the Intl. Conf. on Advances in Civil, Structural and Mechanical Engineering --CSM 2013
Copyright © Institute of Research Engineers and Doctors. All rights reserved.
ISBN: 978-981-07-7227-7 doi:10.3850/ 978-981-07-7227-7_22



(b)

Figure 7. Data driven reconstruction for a façade of the Rotebühlbau, Stuttgart: (a) left part: detected edge points and window lines, right part: classified window and façade cells; (b) 3D façade model.

Modeling step. Within a subsequent modeling process, the window cells are cut out from the existing coarse building model. Thus, windows appear as indentations in the building facade which is depicted in Figure 7(b). Moreover, the reconstruction approach is not limited to indentations. Details can also be added as protrusions to the façade [2].

B. *Automatic inference of a façade grammar*

Based on the data driven reconstruction result, the façade grammar is automatically derived by searching for terminals, their interrelationship, and production rules.

Searching for terminals. In order to yield a meaningful set of terminals for the façade grammar, the building façade is broken down into some set of elementary parts, which are regarded as indivisible and therefore serve as terminals. For this purpose, a spatial partitioning process is applied which segments the façade into floors and each floor into tiles. Tiles are created by splitting the floors along the vertical delimiters of geometries. A geometry describes a basic object on the façade that has been generated during the data driven reconstruction process. It represents either an indentation like a window or a protrusion like a balcony or an oriel.

Two main types of tiles can be distinguished: *wall tiles*, which represent blank wall elements, and *geometry tiles*, which include structures like windows and doors. All these tiles are used as terminals within our façade grammar. In the remaining sections of the paper, wall tiles will be denoted by the symbols W for non-terminals and w_i for terminals. Geometry tiles will be referred to as G and g_i in case of non-terminals and terminals, respectively.

The set of terminals automatically derived from the reconstructed façade of the Rotebühlbau consists of three types of geometry tile (g_0, g_1, g_2) and four types of wall tile (w_0, w_1, w_2, w_3). To illustrate, in Figure 7(b) exemplary instances of each geometry tile are marked in different colors. Instances of the wall tiles are indicated by arrows.

Interrelationship between terminals. Having distinguished elementary parts of the façade we now aim at giving further structure to the perceived basic tiles by grouping them into higher-order structures. This is done fully automatically by identifying hierarchical structures in sequences of discrete symbols. The structural inference reveals hierarchical interrelationships between the symbols in terms of rewrite rules. These rules identify phrases that occur more than once in the string. Thus, redundancy due to repetition can be detected and eliminated. Considering the reconstructed façade of the data set “Rotebühlbau”, Table 1 depicts the corresponding tile string in its original version, the compressed string, and the extracted structures S_i .

TABLE I. GRAMMAR-BASED DESCRIPTION OF THE RECONSTRUCTED FAÇADE. UPPER ROW: ORIGINAL TILE STRING; LOWER ROW: COMPRESSED TILE STRING (LEFT) AND EXTRACTED STRUCTURES (RIGHT).

floor3 → w2 g2 w3 g2 w3 g2 w3 g2 w3 g2 w3 g2 w3 g2 w3 g2 w3 g2 w3 g2 w3 g2 w3 g2 w3 g2 w2 *floor2 → w2 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w2* *floor1 → w2 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w3 g1 w2* *floor0 → w0 g0 w0 g0 w1 g0 w0 g0 w0 g0 w0 g0 w0 g0 w0 g0 w0 g0 w0 g0 w0 g0 w0 g0 w0 g0 w0*	
floor3 → w2 S8 w3 S8 w3 S8 w3 g2 w2 *floor2 → w2 S6 w2* *floor1 → w2 S6 w2* *floor0 → w0 S0 w1 S1 w0 S1 w0 S0 w0 g0 w0*	*S0 → g0 w0 g0* *S1 → S0 w0 S0* *S2 → g1 w3 g1* *S3 → S2 w3 S2* *S4 → S3 w3 S3* *S5 → S4 w3 S3* *S6 → S5 w3 g1* *S7 → g2 w3 g2* *S8 → S7 w3 S7*

Inference of production rules. Based on the sets of terminals $T=\{w_0, w_1, \ldots , g_0, g_1, \ldots\}$ and non-terminals $N=\{W, G, \ldots , S_0, S_1, \ldots\}$, which have been set up as described above, the production rules for our façade grammar can be inferred. Following types of production rules are obtained during the inference process:

p_1: $F \rightarrow W+$

p_2: $W : cond \rightarrow W\ G\ W$

p_3: $G : cond \rightarrow S_i : P(\mathbf{x}|p_3)$

p_4: $G : cond \rightarrow g_i : P(\mathbf{x}|p_4)$

p_5: $lc < W > rc : cond \rightarrow w_i : P(\mathbf{x}|p_5)$

The production rules p_1 and p_2 stem from the spatial partitioning of the façade. p_1 corresponds to the horizontal segmentation of the façade into a set of floors. The vertical partitioning into tiles is reflected in rule p_2. A wall tile, which in the first instance can stand for a whole floor, is replaced by the sequence *wall tile, geometry tile, wall tile*. Each detected

ISBN: 978-981-07-7227-7 doi:10.3850/ 978-981-07-7227-7_22

structure gives rise to a particular production rule in the form of p_3. This rule type states the substitution of a geometry tile for a structure S_i. In addition, all terminal symbols generate production rules denoted by p_4 and p_5 in the case of geometry terminals g_i and wall terminals w_i, respectively.

C. *Grammar Application*

The façade grammar implies information on the architectural configuration of the observed façade concerning its basic façade elements and their interrelationships. This knowledge is applied in three ways. First, the façade model generated during the data driven reconstruction process can be verified and made more robust against inaccuracies and false reconstructions due to imperfect data. Second, façades which are only partially covered by sensor data are completed. Third, totally unobserved façades are synthesized by a production process.

The production process starts with an arbitrary façade, called the axiom, and proceeds as follows: (1) Select a non-terminal in the current string, (2) choose a production rule with this non-terminal as predecessor, (3) replace the non-terminal with the rule's successor, (4) terminate the production process if all non-terminals are substituted, otherwise continue with step (1).

During the production, non-terminals are successively rewritten by the application of appropriate production rules. When more than one production rule is possible for the replacement of the current non-terminal, the rule with the highest probability value is chosen. As soon as the façade string contains only terminals, the production is completed and the string can be transferred into a 3D representation. The grammar based façade reconstruction for the "Rotebühlbau, Stuttgart" is shown in Figure 8.

V. Conclusions and Outlook

This paper has introduced with some general remarks about the developments of LiDAR and photogrammetric imaging. The real breakthrough in photogrammetric image processing was the invention of Semi-Global matching providing point clouds of superb density and quality. Point cloud processing by means of formal grammars is an ongoing research issue in computer vision and photogrammetry and was proven to be very powerful. First results are very promising and have been applied here to point clouds provided by terrestrial photos. Those photos have been collected by an off-the-shelf SLR camera, which is of low-cost and provides imagery of good quality. The images are calibrated and oriented using Structure-from-Motion approaches, as implemented by the Open Source software package VisualSfM (VSfM). Dense image matching finally has provided a huge number of points (about 145 Mio.) for the selected test site in Stuttgart. The follow-up processing with our grammar-based approach could interpret the sample point clouds and finally automatically inference the complete set of the façade grammar, also to those regions of the test building where no data have been collected. The results are very encouraging. It seems that the LoD3 problem can also be overcome by an integrated approach of terrestrial photos, dense image matching and façade grammars. One more example to prove the following statement: "Façade reconstructions from point clouds using formal grammars are powerful generic methods, which combine data-driven and knowledge inference in one integrated approach".

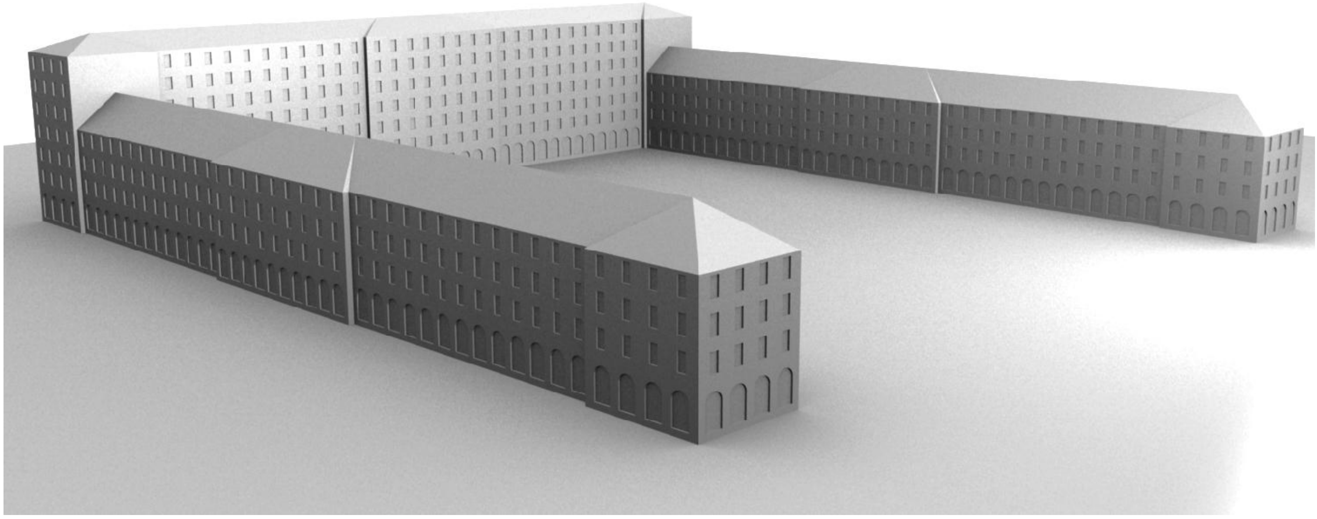


Figure 8. Grammar based façade reconstruction for the "Rotebühlbau, Stuttgart".

References

[1] S. Becker, “Automatische Ableitung und Anwendung von Regeln für die Rekonstruktion von Fassaden aus heterogenen Sensordaten, ” Deutsche Geodätische Kommission, C, No. 658, 156 pages, 2011.

[2] S. Becker, and N. Haala, “Refinement of Building Façades by Integrated Processing of LIDAR and Image Data,” IAPRS & SIS Vol 36 (3/W49A), pp. 7-12, 2007.

[3] A. Carswell, “Lidar Imagery – From Simple Snapshots to Mobile 3D Panoramas,” in Photogrammetric Week’11, Ed. D. Fritsch, Wichmann, VDE Verlag, Berlin and Offenbach, pp. 3-14, 2011.

[4] D. Bekins, and D. Aliaga, “Build-by-number: Rearranging the real world to visualize novel architectural spaces,” in IEEE Visualization, pp. 143-150, 2005.

[5] D. Fritsch (Ed.), “Photogrammetric Week ’11,” Wichmann, VDE Verlag, Berlin and Offenbach, 330 pages, 2011.

[6] N. Haala, and M. Rothermel, “Dense Multi-Stereo Matching for High Quality Digital Elevati-on Models,” in PFG Photogrammetrie, Fernerkundung, Geoinformation, 2012(4), pp. 331-343, 2012.

[7] H. Hirschmüller, ”Stereo Processing by Semiglobal Matching and Mutual Information,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 328-341, 2008.

[8] H. Hirschmüller, “Semi-Global Matching – Motivation, Developments and Applications”, In: Photogrammetric Week ’11, Ed. D. Fritsch, Wichmann, VDE Verlag, Berlin and Offenbach, pp. 173-184, 2011

[9] R. Lenz, and D. Fritsch, ”On the Accuracy of Videometry with CCD Sensors,” Int. Journal Photogrammetry & Remote Sensing (IJPRS), 45, pp. 90-110, 1990.

[10] P. Müller, G. Zeng, P. Wonka, and L. van Gool, “Image-based Procedural Modeling of Façades,” ACM Trans. Graph. Vol. 26 (3), article 85, 9 pages, 2007.

[11] P. Müller, P. Wonka, S. Haegler, A. Ulmer, and L. van Gool, “Procedural Modeling of Buildings,” ACM Transactions on Graphics (TOG) 25 (3), pp. 614-623, 2006.

[12] M. Rothermel, and N. Haala, “Potential of Dense Matching for the Generation of High Quality Digital Elevation models,” in Proceedings of ISPRS Hannover Workshop High-Resoultion Earth Imaging for Geospatial Information, p. 331 - 343 , 2011.

[13] M. Rothermel, K. Wenzel, D. Fritsch, and N. Haala, “SURE: Photogrammetric Surface Reconstruction from Imagery,” Online Proceedings LC3D Workshop, 2012.

[14] E. Tola, V. Lepetit, and P. Fua, “A Fast Local Descriptor for Dense Matching,” Conference on Computer Vision and Pattern Recognition, pp 1 – 8, 2008.

[15] L. van Gool, G. Zeng, F. van den Borre, and P. Müller, “Towards mass-produced building models,” in IAPRS & SIS, Vol. 36 (3/W49A), pp. 209–220, 2007.

[16] P. Viola, A. William, and M. Wells, “Alignment by Maximization of Mutual Information,” Int. J. Comput. Vision, pp. 137 - 154 ,1997.

[17] K. Wenzel, M. Rothermel, D. Fritsch, and N. Haala, “Image Acquisition and Model Selection for Multi-View Stereo”, Proceedings 3DArch Conference, Trento, Italy, 2013.

[18] C. Wu, VisualSFM, “A Visual Structure from Motion System,” 2011, http://www.cs.washington.edu/homes/ccwu/vsfm/.

Dieter Fritsch:

“Dense image matching is delivering point clouds outnumbering LiDAR point clouds in many respects.” “Façade reconstructions from point clouds using formal grammars are powerful generic methods, which combine data-driven analysis and knowledge inference in one integrated approach.”

Susanne Becker:

“The façade grammar implies information on the architectural configuration of the observed façade concerning its basic façade elements and their interrelationships.” “In our application, a formal grammar will be used for the generation of façade structure where only partially or no sensor data is available.”

Mathias Rothermel:

“The redundancy is exploited by our multi-view structure computation implementation. This leads to increased precision of the generated surfaces. Even more important erroneous depth estimations as propagated from dense matching module can be reliably detected and eliminated by checking for geometric consistency.”