A Survey of Urban Reconstruction

Przemyslaw Musialski^{1,2,3}, Peter Wonka^{2,4}, Daniel G. Aliaga^{5,6}, Michael Wimmer^{1}, Luc van Gool^{6,7}, Werner Purgathofer^{1,3}

^{1}Vienna University of Technology
^{2}Arizona State University
^{3}VRVis Research Center
^{4}King Abdullah University of Science and Technology
^{5}Purdue University
^{6}Swiss Federal Institute of Technology Zurich
^{7}Katholieke Universiteit Leuven

Abstract

This paper provides a comprehensive overview of urban reconstruction. While there exists a considerable body of literature, this topic is still under very active research. The work reviewed in this survey stems from the following three research communities: computer graphics, computer vision, and photogrammetry and remote sensing. Our goal is to provide a survey that will help researchers to better position their own work in the context of existing solutions, and to help newcomers and practitioners in computer graphics to quickly gain an overview of this vast field. Further, we would like to bring the mentioned research communities to even more interdisciplinary work, since the reconstruction problem itself is by far not solved.

Categories and Subject Descriptors (according to ACM CCS): Computer Graphics [I.3.5]: Computational Geometry and Object Modeling—; Image Processing And Computer Vision [I.4.6]: Segmentation—; Image Processing And Computer Vision [I.4.8]: Scene Analysis—;

1. Introduction

The documentation of the cultural heritage of our world is a vivid task of many research areas. Also in the field of computational sciences, the reconstruction of cities has obtained a significant attention in recent years. Urban reconstruction is an exciting area of research with several potential applications. Despite the high volume of previous work, there are many unsolved problems, especially when it comes to the development of fully automatic algorithms.

Urban reconstruction is a wide spread domain. Practical fields that benefit from reconstructed three-dimensional urban models are multiple as well:

- In the entertainment industry, the storyline of several movies and computer games takes place in real cities. In order to make these cities believable at least some part of the models are obtained by urban reconstruction.
- Digital mapping for mobile devices, cars, and desktop computers requires two-dimensional and three-dimensional urban models. Examples of such applications are Google Earth and Microsoft Bing Maps.
- Urban planning in a broad sense relies on urban reconstruction to obtain the current state of the urban environment. This forms the basis for developing future plans or to judge new plans in the context of the existing environment.
- Applications such as emergency management, civil protection, disaster control, and security training benefit from virtual urban worlds.

From the economical standpoint, there is an enormous benefit of being able to quickly generate high-quality digital worlds in the growing virtual consumption market.

1.1. Scope

Urban habitats consist of many objects, such as people, cars, streets, parks, traffic signs, vegetation, and buildings. In this paper we focus on urban reconstruction, which we consider as the creation of 3d geometric models of urban areas, individual buildings, façades, and even their further details.

Most papers discussed in this survey were published in computer graphics, computer vision, and photogrammetry and remote sensing. There are multiple other fields that contain interesting publications relevant to urban reconstruction, e.g. machine learning, computer aided design, geo-sciences, mobile-technology, architecture, civil engineering, and electrical engineering. Our emphasis is the geometric reconstruction and we do not discuss aspects, like the construction of hardware and sensors, details of data acquisition processes, and particular applications of urban models.

We also exclude *procedural modeling*, which has been covered in a recent survey by Vanegas et al. [VAW*10]. Procedural modeling is an elegant and fast way to generate huge, complex and realistically looking urban sites, but due to its generative nature it is not well suited for exact reconstruction of existing architecture. It can also be referred to as *forward procedural modeling*. Nevertheless, in this survey we do address its counterpart, called *inverse procedural modeling* (Section 3.3), in addition to other urban reconstruction topics.

We also omit *manual modeling*, even if it is probably still the most widely applied form of reconstruction in many architectural and engineering bureaus. From a scientific point of view, the manual modeling pipeline is well researched. An interesting overview of methods for the generation of polygonal 3d models from CAD-plans has been recently presented by Yin et al. [YWR09].

In order to allow unexperienced computer graphics researchers to step in into the field of 3d reconstruction, we provide a little more detailed description of the fundamentals of *stereo vision* in Section 2. We omit concepts like *trifocal tensor* or the details of *multiview vision*. Instead, we refer more computer vision-versed readers to the referenced papers and textbooks, e.g., by Hartley and Zisserman [HZ04], Moons et al. [MvGV09], and recently by Szeliski [Sze11]. Due to the enormous range of the literature, our report is designed to provide a broad overview rather than a tutorial.

1.2. Input Data

There are various types of possible input data that is suitable as a source for urban reconstruction algorithms. In this survey, we focus on methods which utilize imagery and LiDAR scans (Light Detection And Ranging).

Imagery is perhaps the most obvious input source. Common images acquired from the ground have the advantage of being very easy to obtain, to store, and to exchange. Nowadays, estimated tens of billions of photos are taken worldwide each year, which results in hundreds of petabytes of data. Many are uploaded and exchanged over the Internet, and furthermore, many of them depict urban sites. In various projects this information has been recognized as a valuable source for large scale urban reconstruction [SSS06, IZB07, ASSS10, FFGG*10]. Aerial and satellite imagery, on the other hand, for many years was restricted to the professional sector of photogrammetry and remote sensing community. Only in the recent decade, this kind of input data has become more available, especially due to the advances of Web-mapping projects, like Google Maps and Bing Maps, and was successfully utilized for reconstruction [VAW*10].

Another type of input that is excellently suitable for urban reconstruction is LiDAR data. It typically utilizes laser light which is projected on surfaces and its reflected backscattering is captured, where structure is determined trough the *time-of-flight* principle [CW11]. It delivers semi-dense 3d point-clouds which are very precise, especially for long distance acquisition. Although scanning devices are expensive and still not available for mass markets, scanning technology is frequently used by land surveying offices or civil engineering bureaus for documentation purposes, making LiDAR data especially available for urban reconstruction tasks. Many modern algorithms rely on input from LiDAR, both terrestrial and aerial.

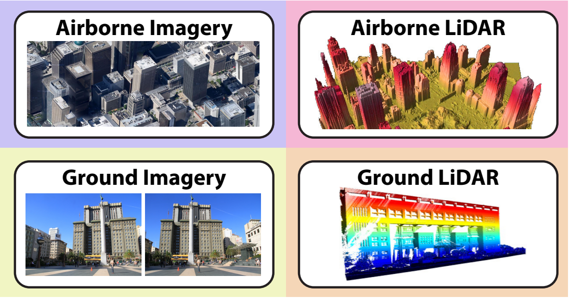

Figure 1: *Input data types. We review interactive and automatic reconstruction methods which use imagery or LiDAR-scans acquired either from the ground or from the air.*

Furthermore, some approaches incorporate both data types in order to combine their complementary strengths: imagery is inherently a 2d source of extremely high resolution and density, but view depended and lacking depth information. Laser-scan is inherently a 3d source of semi-regular and semi-dense structure, but not solid, and often incomplete and noisy. Combining both inputs promises to introduce more insights into the reconstruction process [LCOZ*11].

Finally, both types can be acquired from the ground or from the air (cf. Figure 1), providing a source for varying levels of detail (LOD). The photogrammetry community proposes a predefined standard (OpenGIS) for urban reinstruction LODs [GKCN08]. According to this scheme, airborne data is more suitable for coarse building models reconstruction (LOD1, Section 5), ground based data is more useful for individual buildings (LOD2, Section 3), and facade details (LOD3, Section 4).

1.3. Challenges

Full Automation. The ultimate goal of most computer-based reconstruction approaches is to provide as solutions that are as automatic as possible. In practice, full automation turns out to be hard to achieve. The related vision problems quickly result in huge optimization tasks, where global processes are based on local circumstances, and local processes often depend on global estimates. In other words, the detection of regions of interest is both context dependent (top

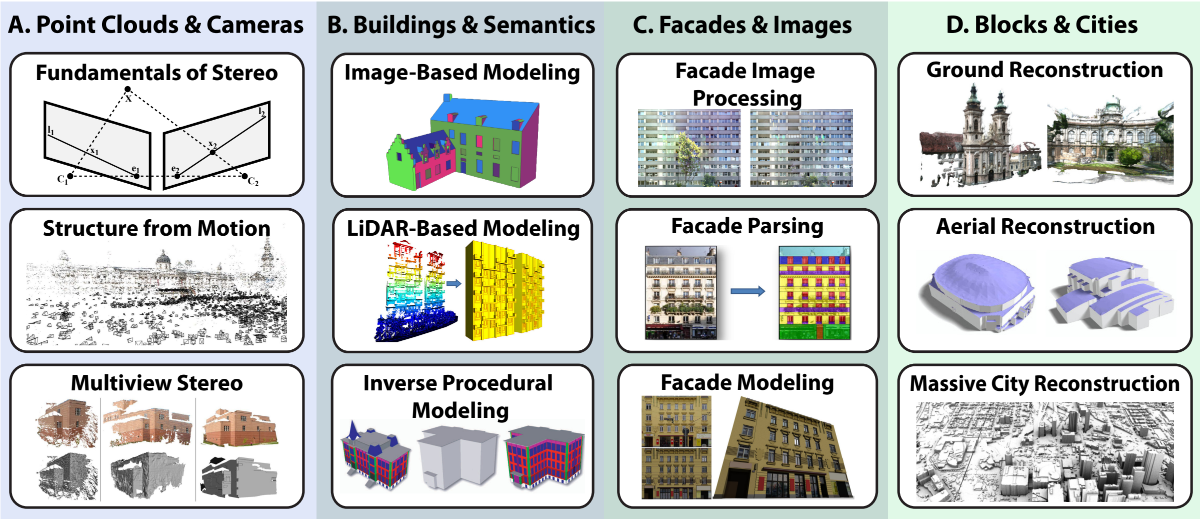

Figure 2: *Overview of urban reconstruction approaches. We attempt to roughly group the methods according to their outcome. We report about interactive methods using both user input and automatic algorithms as well as about fully automatic methods. Note that this is a schematic illustration, and in practice many solutions cannot be strictly classified into a particular bin.*

down), since we expect a well-defined, underlying object, and context free (bottom-up), since we do not know the underlying object and want to estimate a model from the data. In fact, this is a paradox and these dependencies can be generally compared to the "chicken or egg" dilemma.

There is no unique solution to this fundamental problem of automatic systems. Most approaches try to find a balance between these constraints, for instance, they try to combine two or more passes over the data, or eventually to incorporate the human user in order to provide some necessary cues.

Quality and Scalability. An additional price to pay for automation is often the loss of quality. From the point of view of interactive computer graphics, the quality of solutions of pure computer vision algorithms is quite low, while especially for high-quality productions like the movie industry, the expected standard of the models is very high. In such situations, the remedy is either pure manual modeling or at least manual quality control over the data. The downside of this approach is its poor scalability: human interaction does not scale well with huge amounts of input data.

For these reasons, many recent approaches employ compromise solutions that cast the problem in such a way that both the user and the machine can focus on tasks which are easy to solve for each of them. Simplified user interaction that can be performed even by unskilled users often provides the quantum of knowledge that is needed to break out from the mentioned dilemma.

Acquisition Constraints. Other problems that occur in practice are due to the limitations given during the data acquisition process.

For example, it is often difficult to acquire coherent and complete data of urban environments. Buildings are often located in narrow streets surrounded by other buildings and other obstructions, thus photographs, videos or scans from certain positions may be impossible to obtain, neither from the ground nor from the air. The second common handicap is the problem of unwanted objects in front of the buildings, such as vegetation, street signs, vehicles and pedestrians. Finally, there are obstacles like glass surfaces which are problematic to acquire with laser-scans. Photographs of glass are also difficult to process due to many reflections. Lighting conditions, e.g., direct sunshine or shadows, influence the acquisition as well, thus, recovery of visual information that has been lost through such obstructions is also one of the challenges.

A common remedy is to make multiple overlapping acquisition passes and to combine or to compare them. However, in any case post-processing is required.

1.4. Overview

It is a difficult task to classify all the existing reconstruction approaches, since they can be differentiated by several properties, such as input data type, level of detail, amount of automatism, or output data. Some methods are data-driven (bottom-up), some are model-driven (bottom-up), and some combine both approaches.

In this report we propose an output-based ordering of the presented approaches. This ordering helps us to sequentially explain important concepts of the field, building one on top of another; but note that this is not always strictly possible, since many approaches combine multiple methodologies and data types.

Another advantage of this ordering is that we can specify the expected representation of the actual outcome for each section. Figure 2 depicts the main categories that we handle. In this paper, the term *modeling* is generally used for interactive methods, and the term *reconstruction* for automatic ones.

A. Point Clouds & Cameras. Image-based stereo systems have reached a rather mature state in recent times and often serve as preprocessing stages for many other methods since they provide quite accurate camera parameters. Many other methods, even the interactive ones which we present in later sections, rely on this module as a starting point for further computations. For this reason we first introduce the **Fundamentals of Stereo Vision** in Section 2.1. Then, in Section 2.2, we provide the key concepts of image-based automatic **Structure from Motion** methodology, and in Section 2.3, we discuss **Multiview Stereo** approaches.

B. Buildings & Semantics. In this section we introduce a number of concepts that aim at the reconstruction of individual buildings. We start in Section 3.1 with **Image-Based Modeling** approaches. Here we present a variety of concepts based on *photogrammetry* and adapted for automatic as well as for interactive use. In Section 3.2, we introduce concepts of interactive **LiDAR-Based Modeling** aiming at reconstruction of buildings from laser-scan point clouds. In Section 3.3, we describe the concept of **Inverse Procedural Modeling**, which has recently received significant attention due to its ability to compute a compact and editable representation.

C. Façades & Images. We handle the façade topic explicitly because it is of particular importance in our domain of modeling urban areas. In Section 4.1, we handle traditional **Façade Image Processing**, like panoramas and textures. In Section 4.2, we introduce automatic **Façade Parsing** concepts that aim at segmentation, detection of symmetry and repetitive elements, and higher-order model fitting. In Section 4.3, we introduce concepts which aim at interactive **Façade Modeling**, such as subdivision into sub-elements (e.g., floors, windows, and other domain-specific features).

D. Blocks & Cities. In this section we discuss automatic reconstruction of models of large areas or whole cities. Such systems often use multiple input data types, like aerial images and LiDAR. We first mention methods performing **Ground Reconstruction** in Section 5.1. In Section 5.2, we focus on **Aerial Reconstruction** from aerial imagery, LiDAR or hybrids, and finally, in Section 5.3, we discuss methods which aim at automatic **Massive City Reconstruction** of large urban areas.

In the remainder of this article we review those categories.

2. Point Clouds & Cameras

Generally speaking, stereo vision is a method which allows restoring the third dimension from multiple (at least two) distinct two-dimensional images. The underlying paradigm is called *stereopsis*, which is also the way humans are able to perceive depth from two slightly dispaired images.

2.1. Fundamentals of Stereo Vision

In computer vision, the goal is to reconstruct 3d structure which lies in the 3d Euclidian space in front of multiple camera devices, where each of them projects the scene on a 2d plane. For the purpose of simplification and standardization, the established common model of a camera in computer vision is the *pinhole camera*. This model allows expressing the projection by means of a linear matrix equation using the *homogeneous coordinates*.

Camera Model. The operation we want to carry out is a linear *central projection*, thus the camera itself is defined by an *optical center* **C** which is also the origin of the local 3d coordinate frame. Typically, in computer vision, a right-handed coordinate system is used, where the "up-direction" is the Y-axis and the camera "looks" along the positive Z-axis, which is also called the *principal axis* as shown in Fig. 3. The scene in front of the camera is projected onto the *image plane*, which is perpendicular to the principal axis, and its distance to the optical center is the actual *focal length* f of the camera. The principal axis pierces the image plane at the *principal point* $\mathbf{p} = [p_x, p_y]^{\mathrm{T}}$ as depicted in Figure 3.

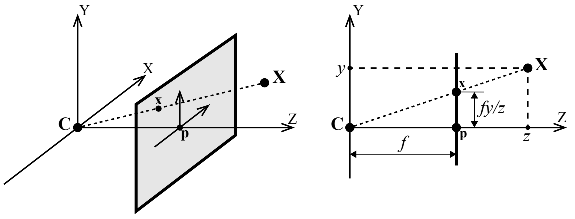

Figure 3: *Camera geometry: (left)* **C** *denoted the camera center and* **p** *the principal point. In a basic setup the center of the first camera is centered at the origin; (right) 2d cross section of the projection.*

In practice, lenses of common cameras are quite sophisticated optical devices whose projective properties are usually not strictly linear. In order to obtain the standardized camera from any arbitrary device, a process called *camera calibration* is carried out. In this process the internal camera parameters are determined and stored in the camera *intrinsic calibration matrix* **K**. The notation of the matrix varies throughout the literature, but a basic version can be described as:

$$\mathbf{K} = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

where f denotes the focal length, and the point $\mathbf{p} = [p_x, p_y]^{\mathrm{T}}$ is the principal point of the camera plane. This setup allows projecting a point $\mathbf{X} = [x,\ y,\ z]^{\mathrm{T}}$ from 3d space onto a point $\mathbf{x}$ on the image plane by a simple equation:

$$\mathbf{x} = \mathbf{K}\mathbf{X} \ \rightarrow\ [fx/z + p_x,\ fy/z + p_y]^{\mathrm{T}}\ . \tag{2}$$

Another aspect of camera calibration is its location in space, which is often called the *extrinsic camera parameters*. In single-view vision, it is sufficient to define the origin of the global space at the actual camera center without changing any of the mentioned equations. In multiview vision, this is not adequate anymore, since each camera requires its own local projective coordinate system. These cameras, as well as the objects in the scene, can be considered as lying in a common 3d space that can be denoted as the *world space*. The pose of each particular camera can be described by a rotation, expressed by a 3-by-3 matrix $\mathbf{R}$, and the position of its optical center $\mathbf{C}$, which is a vector in 3d world space. This leads to an extension of Equation 1 to a 3 × 4 matrix:

$$\mathbf{P} = \mathbf{K}\mathbf{R}\,[\mathbf{I}\,|-\mathbf{C}]\ , \tag{3}$$

where $\mathbf{P}$ is referred to as *homogeneous camera projection matrix*. Note that now the 3d space points have to be expressed in homogeneous coordinates $\mathbf{X} = [x, y, z, 1]^{\mathrm{T}}$. In this way, an arbitrary point $\mathbf{X}$ in world space can be easily projected onto the image plane by:

$$\mathbf{x} = \mathbf{K}\mathbf{R}\,[\mathbf{X} - \mathbf{C}] = \mathbf{P}\mathbf{X}\,. \tag{4}$$

Determining the extrinsic parameters is often referred to as *pose estimation* or as *extrinsic calibration*.

For a typical hand-held camera, the mentioned parameter sets are not known a priori. There are several ways to obtain the intrinsic camera calibration [LZ98, WSB05, JTC09], where one of them is to take photos of predefined patterns and to determine the parameters by minimizing the error between the known pattern and the obtained projection [MvGV09]. Extrinsic parameters are of more importance in a multi-camera setup, which can be obtained automatically from a set of overlapping images with common corresponding points [MvGV09].

Please note that the described camera model is a simplified version which does not take all aspects into account, like the radial distortion or the aspect ratio of typical CCD-pixels. We refer the reader to Hartley and Zisserman [HZ04] and to Moons et al. [MvGV09] for exhaustive discussions about calibration and self-calibration of multiview setups.

Epipolar Geometry. For a single camera, we are able to determine only two parameters of an arbitrary 3d point projected to the image plane. In fact, the point $\mathbf{X}$ lies on a projecting ray as depicted in Figure 4. Obviously, it is not possible to define the actual position of the point along the ray without further information. An additional image from a different position provides the needed information. Figure 4 depicts this relationship: The projective ray from the first camera trough a 2d image point $\mathbf{x}_1$ and a 3d point $\mathbf{X}$ appears as a line $\mathbf{l}_2$ in the second camera, which is referred to as an *epipolar line*. Consecutively, a corresponding point in the second image must lie on the line and is denoted as $\mathbf{x}_2$. Note that also the optical centers of each camera project onto the image planes of each other, as shown in Figure 4. These points are denoted as the *epipoles* $\mathbf{e}_1$ and $\mathbf{e}_2$, and the line connecting both camera centers is referred to as the *baseline*. The plane defined by both camera centers and the 3d point $\mathbf{X}$ is referred to as *epipolar plane*.

Stereo Correspondence and Triangulation. In a stereo setup, the relation of two views to each other is expressed in a 3-by-3 rank 2 matrix, referred to as the *fundamental matrix* $\mathbf{F}$, which satisfies:

$$\mathbf{x}_1^{\mathrm{T}}\mathbf{F}\,\mathbf{x}_2 = 0\,, \tag{5}$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are two corresponding points in both images. There exist well-known algorithms to determine the fundamental matrix from 8 (linear problem) or 7 (non-linear problem) point correspondences [MvGV09]. When working with known intrinsic camera settings, the relation is also often referred to as the *essential matrix* $\mathbf{E}$, which can be determined even from the correspondences of five points [Nis04].

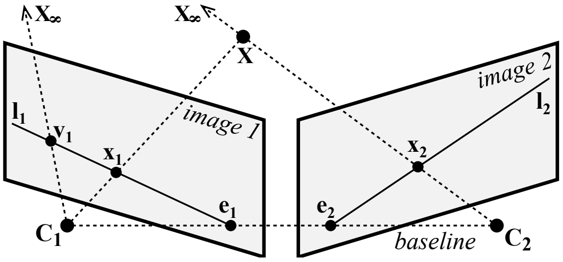

Figure 4: *Epipolar geometry in a nutshell: points* $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *are corresponding projections of the 3d point* $\mathbf{X}$. *In* image 1 *the point* $\mathbf{x}_1$ *lies on the epipolar line* $\mathbf{l}_1$. *The epipoles* $\mathbf{e}_1$ *and* $\mathbf{e}_2$ *indicate the positions where* $\mathbf{C}_1$ *and* $\mathbf{C}_2$ *project respectively. The point* $\mathbf{v}_1$ *in image 1 is the vanishing point of the projecting ray of* $\mathbf{x}_2$.

Assuming full camera calibration, the problem of 3d structure reconstruction from stereo can be reduced to two sub-problems: (1) the one-to- one correspondence problem across the images and (2) the intersection of the projective rays problem. The second operation is usually referred to as *structure triangulation* due to the triangle which is formed by the camera centers $\mathbf{C}_1$ and $\mathbf{C}_2$, and each consecutive point $\mathbf{X}$ in 3d space. Note, that this term has a different meaning then the triangulation of geometric domains, which is often used interchangeably to a tessellation into triangles in computer graphics literature.

One of the key inventions which advanced this research field are robust feature-point detection algorithms, like SIFT [Low04] and SURF [BTvG06, BETvG08]. These image processing methods allow for efficient detection of characteristic *feature points* which can be matched across multiple images. Both algorithms compute very robust descriptors which are mostly invariant to rotation and scale, at least to a certain degree as shown by Schweiger et al. [SZG*09]. Once the corresponding features have been established, the extrinsic (i.e., pose in 3d space) and, under certain circumstances, also the intrinsic (e.g., focal length) parameters of their cameras, as well as positions of the 3d space points can be determined in an iterative process often called *structure from motion*.

2.2. Structure from Motion

In practice, the stereo vision procedure described in the previous section can be used to register multiple images to one another, to orient and place their cameras, and to recover 3d structure. It is carried out incrementally in several passes, usually starting from an initial image pair and adding consecutive images to the system one by one. Mutual relations between the images are detected sequentially, new 3d points are extracted and triangulated, and the whole 3d point cloud is updated and optimized.

In a first stage, for each image a sparse set of feature-points is detected, which are than matched in a high-dimensional feature space in order to determine unique pairs of corresponding points across multiple images. This stage is usually approached with high-dimensional spatial nearest-neighbor search algorithms, like the kd-tree, vp-tree [KZN08] or the vocabulary-tree [NS06].

In order to improve the stability of the feature matching process, robust estimation algorithms (i.e,. RANSAC [FB81, RFP08]) are employed in order to minimize the number of wrong matches across images. By utilizing the already known parameters it is possible to "filter out" outliers which deviate too far from an estimated mapping.

Finally, advanced *bundle adjustment* solvers [TMHF99, LA09, ASSS10, WACS11] are used to compute highly accurate camera parameters and a sparse 3d point cloud. Bundle adjustment is a non-linear least-squares optimization process which is carried out after the addition of several new images to the system in order to suppress the propagation of an error. In addition, is is always performed at the end, after all images have been added, in order to optimize the whole network. In this process both the camera parameters (**K**, **R**, and **C**) as well as the positions of the 3d points **X** are optimized simultaneously, aiming at minimization of the re-projection error:

$$\sum_{j}\sum_{i\in j}\left\|\mathbf{x}_{ij}-\left(\mathbf{K}_j\mathbf{R}_j\left(\mathbf{X}_i-\mathbf{C}_j\right)\right)\right\|^2 \longrightarrow \min_{\mathbf{K}_j,\mathbf{R}_j,\mathbf{C}_j,\mathbf{X}_i}, \qquad (6)$$

where $i \in j$ indicates that the point $\mathbf{X}_i$ is visible in image j, and $\mathbf{x}_{ij}$ denotes the projection of 3d points $\mathbf{X}_i$ onto image j. Usually optimization is carried out using the non-linear Levenberg-Marquardt minimization algorithm [HZ04].

The entire process is typically called structure from motion (SfM) due to the fact that the 3d structure is recovered from a set of photographs which have been taken by a camera that was in motion. In fact, this methodology applies to video sequences as well [vGZ97], and it can also be performed with line-feature correspondences across images [TK95, SKD06], which is especially suitable to urban models.

The advantage of general SfM is its conceptual simplicity and robustness. Furthermore, since it is a bottom-up approach that makes only few assumptions about the input data, it is quite general.

2.3. Multiview Stereo

The described procedure of SfM delivers networks of images that are registered to each other, including their camera properties, as well as sparse point clouds of 3d structure. However, the point clouds are usually rather sparse and do not contain any solid geometry. The next step in order to obtain more dense structure is usually called *dense matching*. It is mostly used for image-based reconstruction of detailed surfaces as shown in Figure 6. In this context, dense means to try to capture information from all pixels in the input images – in contrast to sparse methods, where only selected feature points are considered.

In this report we mention several dense matching methods which have been utilized for urban reconstruction. For a more detailed overview, we refer the reader to Scharstein and Szeliski [SS02a] for *two-view stereo* methods, and to Seitz et al. [SCD*06] for *multiview stereo* methods (MVS).

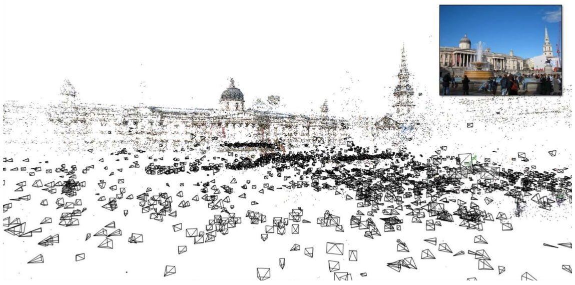

Figure 5: *A sparse point cloud generated from several thousands of unordered photographs, and one photo taken from the nearly the same viewpoint. Figure courtesy of Noah Snavely [SSG*10], ©2010 IEEE.*

Furthermore, many multiview stereo methods often utilize a concept called "plane-sweeping". This process, originally proposed by Collins [Col96], is approached with multiple to each other registered views. The main idea is to "sweep" a plane through the 3d space along one of the axes with rays shot from all pixels of all cameras onto the plane. According to epipolar geometry, intersections of the rays with each other at their hitpoints on the plane indicate 3d structure points. Collins showed how to utilize a series of homographies in order to efficiently accumulate these points and to generate reconstructions [Col96]. The main advantages of this idea are that (1) it works with an arbitrary number n

of images, (2) its complexity scales with $O(n)$, and (3) all images are treated in the same way. Thus, the method was called by the author as *true multi-image matching* approach. Plane sweeping has been successfully utilized for recovery of dense structure and consecutively extended in order to exploit with modern programmable hardware graphics accelerators [YP03] or multiple sweeping directions [GFM*07].

Both sparse and dense frameworks have been utilized in urban reconstruction and in this section we want to review the most important publications.

Sparse Reconstruction. There is a number of papers which utilize sparse SfM for exploration and reconstruction of urban environments. All these methods produce, either as the end-product or at least as an intermediate step, sparse 3d point clouds. In a series of publications, Snavely et al. [SSS06,SSS07,SGSS08,SSG*10] develop a system for navigation in urban environments which is mainly based on sparse points and structure from motion camera networks. In this system, called "Photo Tourism" it is possible to navigate through large collections of registered photographs. The density of photographs combined with sparse point clouds and smooth animations gives the user the impression of spatial coherence. These works contributed significantly to the maturity of the current state-of-the-art of SfM and to the use of unstructured collections of Internet images [LWZ*08].

Further methods introduced semi-dense (quasi-dense) SfM [LL02,LQ05] and aimed at improving the performance, scalability, and accuracy [ASS*09,FQ10,AFS*10,COSH11] in order to deal with arbitrarily high numbers of input photographs. Recent work of Agarwal et al. demonstrates impressively how to reconstruct architecture from over hundred thousand images in less than one day [AFS*11]. They cast the problem of matching of corresponding images as a graph estimation problem, where each of the images is a vertex and edges connect only images which depict the same object. They approach this problem using multiview clustering of scene objects [FCSS10].

Bauer et al. [BZB06] proposed a method based on plane-sweep in order to recover sparse point-clouds of buildings.

Dense Reconstruction. Dense structure of the surface is also computed by a multiview stereo matching algorithm proposed by Pollefeys [PvGV*04]. Vergauwen and Van Gool [VvG06] extended this method from regular sequences of video frames to still images by improved feature matching, additional internal quality checks and methods to estimate internal camera parameters. This approach was introduced as the free, public ARC3D web-service, allowing the public to take or collect images, upload them, and get the result as dense 3d data and camera calibration parameters [TvG11]. Images of buildings are among the most often uploaded data. Further extensions to this methodology were presented by Akbarzadeh et al. [AFM*06] and Pollefeys et al. [PNF*08].

Furukawa and Ponce [FP07, FP09] presented a different approach for multiview stereo reconstruction. Their method uses a structure from motion camera network as a preliminary solution, but further, it is based on matching small patches placed on the surface of the scene object which are back-projected onto the images. First, features like Harris corners [HS88] or DoG spots [Low04] are detected and matched across images, which, projected on the object, define the locations of the patches. These are defined in such a way that their re-projected footprints cover the actual images. They are then optimized such that a photometric discrepancy function across the re-projected patches is minimized. The results are semi-dense clouds of small patches which serve as a basis for denser structure triangulation and, finally, for polygonal surface extraction. To achieve this, they employ the Poisson surface reconstruction algorithm [KBH06], as well as an iteratively refined visual hull method [FP08]. Also this 3d reconstruction idea is very generic, but it has since been extended and applied to 3d urban reconstruction as well [FCSS09a, FCSS10].

Another approach for the reconstruction of dense structures is to perform pairwise dense matching [SS02a] of any two registered views and then to combine the computed depth maps with each other. Usually this approach is denoted as *depth map fusion*. There are several ideas how to perform this, such as from Goesele et al. [GCS06, GSC*07], Zach et al. [ZPB07, IZB07], Merrell et al. [MAW*07].

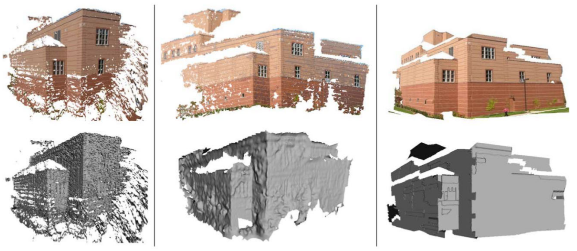

Figure 6: *Comparison of 3d models created by different methods. Left: Vergauwen and van Gool [VvG06], middle: Furukawa and Ponce [FP07], right: Micusik and Kosecka [MK10]. Figure courtesy of Branislav Micusik [MK10]. ©2010 Springer.*

A common problem of dense stereo methods is that the models exhibit a relatively hight amount of noise along flat surfaces. This is due to the nature of matching nearby points more or less independently from each other. This, in fact, is a major obstacle in urban reconstruction, where most models are composed of groups of planar surfaces. Several methods try to overcome this problem by including hierarchical models [LPK09], Manhattan-world assumptions [FCSS09a, FCSS09b], multi-layer depth maps [GPF10], or piece-wise planar priors [MK09, MK10, SSS09, CLP10, GFP10].

Generally, dense multiview approaches deliver quite impressive results, like the large scale system presented by

Frahm et al. [FFGG*10]: it deals with almost 3 million images, performs image clustering, SfM, and dense map fusion in one day on a single PC. On the downside, these systems usually provide dense polygonal meshes without any higher-level knowledge of the underlying scene, even though such information is very useful in complex architectural models. However, there exist other approaches which provide well-defined geometric shapes and often also some semantics. We cover such methods in Section 3.

3. Buildings & Semantics

Manually modeling architecture is a tedious and time-consuming task, but for a long time it was the only way to obtain 3d models of urban sites. However, in the past two decades there has been significant research in automating this process. In this section we turn our attention to approaches which aim at reconstructing whole buildings from various input sources, such as a set of photographs or laser-scanned points, typically by fitting some parameterized top-town building model.

3.1. Image-Based Modeling

In *image-based modeling*, a static 3d object is modeled from of or with the help of one or more images or videos. While this definition is very general, such methods are often also referred to as *photogrammetric modeling*, especially in the photogrammetry and remote sensing community. In this section we restrict our review to approaches which model single buildings mainly from ground-based or close-range photographs.

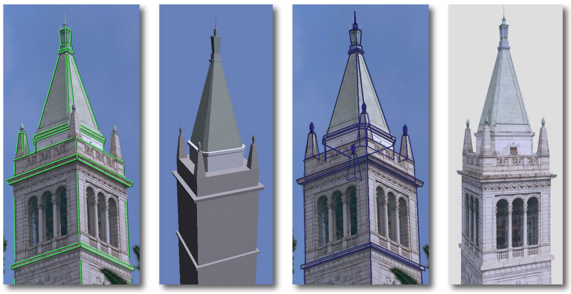

Figure 7: *Interactive image-based modeling: (1) input image with user-drawn edges shown in green, (2) shaded 3D solid model, (3) geometric primitives overlaid onto the input image, (4) final view-dependent, texture-mapped 3D model. Figure courtesy of Paul Debevec [DTM96] ©1996 ACM.*

Generally, in order to obtain true 3d properties of an object, the input must consist of at least two or more perspective images of the scene. There are also single-image methods which usually rely on user input or knowledge of the scene objects in order to compensate the missing information.

Nonetheless, also multiview methods make a number of assumptions about the underlying object in order to define a top-down architectural model which is successively completed from cues derived from the input imagery. The outcome usually consists of medium-detail geometric building models, in some cases enriched with finer detail, such as as windows. Some methods also deliver textures and more detailed façade geometry, but we omit discussion of these features in this section, and instead elaborate them in Sec. 4.

The degree of user interaction varies across the methods as well. Generally, the tradeoff is between quality and scalability. More user interaction leads to more accurate models and semantics, but such approaches do not scale well to huge amounts of data. Using fully automatic methods is an option, but they are more error prone and also depend more on the quality of the input.

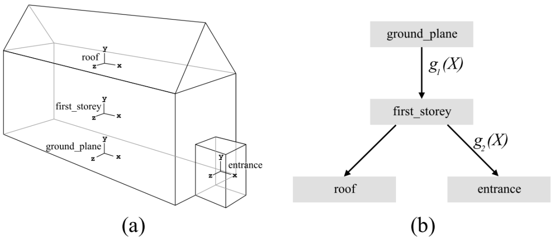

Figure 8: *A geometric model of a simple building (a); the model's hierarchical representation (b). The nodes in the tree represent parametric primitives while the links contain the spatial relationships between the blocks. Figure courtesy of Paul Debevec [DTM96] ©1996 ACM*

Interactive Multiview Modeling. A seminal paper in this field was the work of Debevec et al. [DTM96]. Their system, called "Façade", introduced a workflow for interactive multiview reconstruction.

The actual model is composed of parameterized primitive polyhedral shapes, called *blocks*, arranged in a hierarchical tree structure (cf. Figure 8). Debevec et al. based their modeling application on a number of observations [DTM96]:

- Most architectural scenes are well modeled by an arrangement of geometric primitives.
- Blocks implicitly contain common architectural elements such as parallel lines and right angles.
- Manipulating block primitives is convenient, since they are at a suitably high level of abstraction; individual features such as points and lines are less manageable.
- A surface model of the scene is readily obtained from the blocks, so there is no need to infer surfaces from discrete features.
- Modeling in terms of blocks and relationships greatly reduces the number of parameters that the reconstruction algorithm needs to recover.

Composing an architectural model from such blocks turned out to be quite a robust task which provides very good results (cf to Figure 8). During the modeling process, the user interactively selects a number of photographs of the same object and marks corresponding edges in each of them. The correspondences allow establishing epipolar-geometric relations between them, and the parameters of the 3d primitives can be fitted automatically using a non-linear optimization solver [TK95]. Because the number of views is kept quite low, and because many of the blocks can be constrained to each other – thus significantly reducing the parameter space – the optimization problem can be solved efficiently (e.g., up to a few minutes on the 1996 hardware).

The "Façade" system was one of the first of its kind. The observations made in this paper turned out to be quite appropriate for urban scenes. Furthermore, its additional advantage over other, mostly automatic approaches, was the high quality of the obtained results.

This encouraged other researchers to invest time in the development of interactive systems. For example, another image-based modeling framework called "Photobuilder" was presented by Cipolla and Robertson [CR99, CRB99]. Their work introduced an interactive system for recovering 3d models from few uncalibrated images of architectural scenes based on vanishing points and the constraints of projective geometry. Such constraints, like parallelism and orthogonality, were also exploited by Liebowitz et al. [LZ98, LCZ99], who presented a set of methods for creating 3d models of scenes from a limited numbers of images, i.e., one or two, for situations where no scene coordinate measurements are available.

Lee et al. introduced an interactive technique for block-model generation from aerial imagery [LHN00]. They extended the method further and introduced automatic integration of ground-based images with 3d models in order to obtain high-resolution façade textures [LJN02a, LJN02b, LJN02c]. They also proposed an interactive system which provides a hierarchical representation of the 3d building models [LN03]. In this system, information for different levels of detail can be acquired from aerial and ground images. The method requires less user interaction than the "Façade" system, since it uses more automatic image calibration. It also requires at most 3 clicks for creating a 3d model and 2 model-to-image point correspondences for the pose estimation. Finally, they also handled more detailed façade and window reconstruction [LN04] (cf. Section 4.3).

Also El-Hakim et al. [EhWGG05, EhWG05] proposed a semi-automatic system for image-based modeling of architecture. Their approach allows the user to model parameterized shapes which are stored in a database and can be reused for further modeling of similar objects.

The next important advance of interactive modeling was the combination of automatic sparse structure from motion methods with parameterized models and user interaction. SfM provides a network of registered cameras and a sparse point-cloud (cf. Section 2). The goal is to fit a parameterized model into this data.

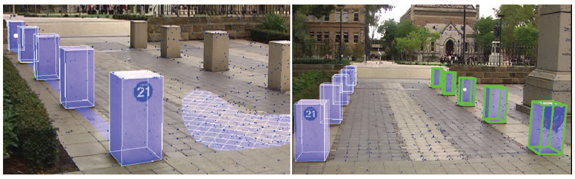

Figure 9: *Interactive modeling of geometry in video. Left: Replicating the bollard by dragging the mouse. Right: Replicating a row of bollards. Figure courtesy of Anton van den Hengel [vdHDT*07a] ©2007 ACM.*

A series of papers published by van den Hengel and colleagues describe building blocks of an image and video-based reconstruction framework. Their system [vdHDT*06] uses camera parameters and point clouds generated by a structure from motion process (cf. Section 2) as a starting point for developing a higher-level model of the scene. The system relies on the user to provide a small amount of structure information from which more complex geometry is extrapolated. The regularity typically present in man-made environments is used to reduce the interaction required, but also to improve the accuracy of fit. They extend their higher-level model [vdHDT*07a], such that the scene is represented as a hierarchical set of parameterized shapes, as already proposed by others [DTM96, LN03]. Relations between shapes, such as adjacency and alignment, are specified interactively, such that the user is asked to provide only high-level scene information and the remaining detail is provided through geometric analysis of the images (cf. Figure 9). In a follow-up work [vdHDT*07b], they present a video-trace system for interactive generation of 3d models using simple 2d sketches drawn by the user, which are constrained by 3d information already available.

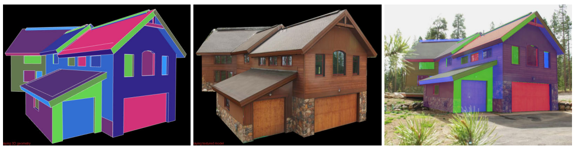

Figure 10: *Results of interactive image-based modeling method. Figure courtesy of Sudipta Sinha [SSS*08], ©2008 ACM.*

Sinha et al. [SSS*08] presented an interactive system for generating textured 3d models of architectural structures from unordered sets of photographs. It is also based on structure from motion as the initial step. This work introduced novel, simplified 2d interactions such as sketching of outlines overlaid on 2d photographs. The 3d structure is automatically computed by combining the 2d interaction with the multiview geometric information from structure from motion analysis. This system also utilizes vanishing point constraints [RC02], which are relatively easy to detect in architectural scenes (cf. Figure 10).

Recently, also Larsen and Moeslund [LM11b] proposed an interactive method for modeling buildings from sparse SfM point-clouds. It provides simple block-models and textures. The pipeline also includes an approach for automatic segmentation of façades. Arikan et al. [ASW*12] proposed a framework for generation of polyhedral models over semi-dense unstructured point-clouds from SfM. Their system automatically extracts planar polygons which are optimized in order to "snap" to each other to form an initial model. The user can refine it with simple interactions, like coarse 2d strokes. The output are accurate and well-defined polygonal objects.

Automatic Multiview Modeling. A number of image-based and photogrammetric approaches attempt fully automatic modeling. Buildings are especially suited to such methods because the model can be significantly constrained by cues typically present in architectural scenes, like parallelism and orthogonality. These attributes help to extract line-features and vanishing points from the images, which opens the door for compact algorithms [LZ98, Rot00, RC02, KZ02] that aim at both reliable camera recovery and consecutive reconstruction of 3d structure.

While the mentioned papers provided well-defined tools for multiview retrieval of general objects, others proposed model-based systems which aim more specifically at building reconstruction. An early project for reconstructing whole urban blocks was proposed by Teller [Tel98]. Coorg and Teller [CT99] detected vertical building planes using the space-sweep algorithm [Col96] and provided a projective texture for their façade, however, their system did not yet utilized any stronger top-down model of a building.

Werner and Zisserman [WZ02] proposed a fully automatic approach inspired by the work of Debevec et al. [DTM96]. Their method accepts a set of multiple short-range images and it attempts to fit quite generic polyhedral models in the first stage. In the second stage, the coarse model is used to guide the search for fitting more detailed polyhedral shapes, such as windows and doors. The system employs the plane-sweep approach [Col96] for polyhedral shape fitting, which was also used by Schindler and Bauer [BKS*03], who additionally introduced more specific templates for architectural elements.

The work of Dick et al. [DTC00, DTC04] aims also at an automatic acquisition of 3d architectural models from small image sequences. Their model is Bayesian, which means that it needs the formulation of a prior distribution. In other words, the model is composed of parameterized primitives (such as walls, doors or windows), each having assigned a certain probabilistic distribution. The prior of a wall layout, and the priors of the parameters of each primitive are partially learned from training data, and partially added manually according to the knowledge of expert architects. The model is reconstructed using a *Markov Chain Monte Carlo* (MCMC) machinery, which generates a range of possible solutions from which the user can select the best one when the structure recovery is ambiguous. In a way this method is loosely related to inverse procedural methods described later in Section 3.3 because it also delivers semantic descriptions of particular elements of the buildings.

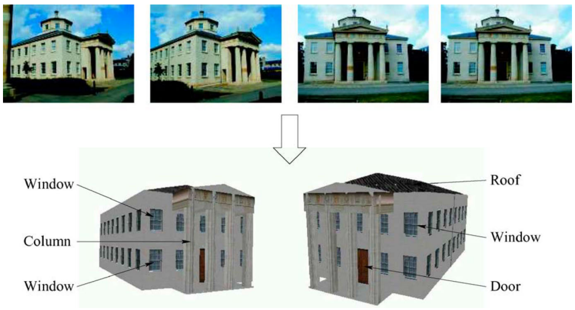

Figure 11: *Example of fully automatic modeling: A labeled 3d model is generated from several images of an architectural scene. Figure courtesy of Anthony Dick [DTC04], ©2004 Springer.*

More recently, Xiao et al. [XFZ*09] provided another automatic approach to generate 3d models from images captured along the streets at ground level. Since their method reconstructs a larger urban area than a single building, we discuss it in Section 5.1.

Interactive Single-view Modeling. Assuming some knowledge about the scene, it is often possible to reconstruct it from a single image. Horri et al. [HAA97] provided an interactive interface for adding perspective to a single photograph, which is then subsequently exploited in order to simulate the impression of depth. Shum and Szeliski [SHS98] introduced a system for interactive modeling of building interiors from a single panoramic image. Photogrammetric tools, e.g., a linear algorithm which computes plane rectification, plane orientation, and camera calibration from a single image [LCZ99], paved the way for further single-image approaches. For example, van den Heuvel [vdH01] introduced an interactive algorithm for extraction of buildings from a single image. Oh et al. [OCDD01] proposed a tool for interactive depth-map painting in a single photo, which is then utilized for rendering.

The most recent paper in this category was presented by Jiang et al. [JTC09], who introduced an algorithm to calibrate the camera from a single image, and proposed an interactive method which allows for recovery of 3d points driven by the symmetry of the scene objects. Its limitation is that it only works for highly symmetric objects because the epipolar constraints are derived from symmetries present in the scene.

Automatic Single-view Modeling. Some fully automatic methods have been attempted. Hoiem et al. [HEH05] proposed a method for creation of simplified "pop-up" 3d models from a single image, by using image segmentation and depth assignments based on vanishing points [RC02, KZ02]. Kosecka and Zhang [KZ05] introduced an approach for automatic extraction of dominant rectangular structures from a single image using a model with a high-level rectangular hypothesis.

To summarize image-based modeling, we must say that fully automatic modeling still suffers considerable quality loss compared to interactive approaches, and as of today, the best quality is still obtained by interactive multiview methods. For this reason, due to the current demand for high-quality models, most close-range reconstruction is approached with semi-automatic modeling.

3.2. LiDAR-Based Modeling

Another group of methods focusing on the reconstruction of buildings utilizes laser-scan data, also referred to as LiDAR-data (Light Detection and Ranging). Generally, there are two main types of this class of data: those acquired by ground-based devices (terrestrial LiDAR), and those captured from the air (aerial LiDAR).

Laser scanning is widely used in the photogrammetry and remote sensing community for measurement and documentation purposes. In this report, we omit those methods. Only in the recent years, the goal of further segmentation and fitting of parameterized high-level polyhedral models emerged, and we will focus on those approaches.

Interactive Modeling. Due to advances in laser-scanning technology, LiDAR data has become more accessible in recent time, but also the quality demands on the models has grown due to the larger bandwidth and higher resolution displays. While laser-scans are in general dense and relatively regular – thus perfectly suited for architectural reconstruction – on the other hand, the practical process of acquisition is difficult and the resulting data is often corrupted with noise, outliers and incomplete coverage. In order to overcome such problems, several methods propose to process the data with user interaction.

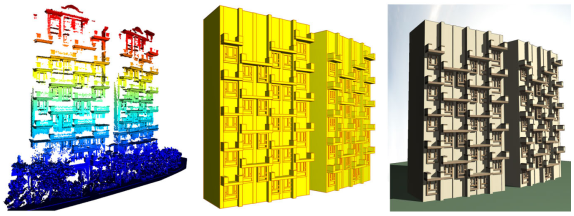

Figure 12: *Results of interactive fitting of "SmartBoxes" to uncomplete LiDAR data. Figure courtesy of Liangliang Nan [NSZ*10], ©2010 ACM.*

Böhm [Bö08] published a method for completion of terrestrial laser-scan point clouds, which is done by interactively utilizing the repetitive information typically present in urban buildings. Another approach aiming for a similar goal was introduced by Zheng et al. [ZSW*10]. It is also an interactive method for consolidation which completes holes in scans of building façades. This method exploits large-scale repetitions and self-similarities in order to consolidate the imperfect data, denoise it, and complete the missing parts.

Another interactive tool for assembling architectural models directly over 3d point clouds acquired from LiDAR data was introduced by Nan et al. [NSZ*10]. In this system, the user defines simple building blocks, so-called SmartBoxes, which snap to common architectural structures, like windows or balconies. They are assembled through a discrete optimization process which balances between fitting the point-cloud data [SWK07] and their mutual similarity. In combination with user interaction, the system can reconstruct complex buildings and façades from sparse and incomplete 3d point clouds (cf. to Figure 12).

Other approaches aim at the enhancement of LiDAR data by fusing it with optical imagery. Some work on registration and pose estimation of ground-images with laser-scan point clouds was done by Liu and Stamos [LS07]. The method aims at robust registration of the camera-parameters of the 2d images with the 3d point cloud. Recently, Li et al. [LZS*11] introduced an interactive system for fusing 3d point-clouds and 2d images in order to generate detailed, layered and textured polygonal building models. The results of this method are very impressive, of course again, at the cost of human labor and extended processing time.

Automatic Modeling. Similar as with image-based modeling, there also exist many approaches that aim at full automation. While such systems scale well with the data, they usually require the user to set up a number of parameters. This kind of parametrization is very common in fully automatic methods and it turns out to be also an often underestimated obstacle, since the search for proper parameters can be very time consuming. The benefit is that once good parameters are found for a dataset, it can be processed automatically irrespective its actual size.

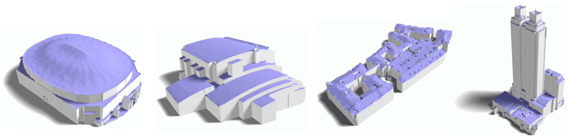

Figure 13: *Results of the automatic method which uses LiDAR segmentation. Figure courtesy of Qian-Yi Zhou [ZN10], ©2010 Springer.*

In earlier works, Stamos and Allen developed a system for reconstruction of buildings from sets of range scans combined with sets of unordered photographs [SA00b, SA00a, SA01, SA02]. Their method is based on fitting planar polygons into pre-clustered point-clouds. Bauer et al. [BKS*03] also proposed an approach for the detection and partition of planar structures in dense 3d point clouds of façades, like polygonal models with a considerably lower complexity than the original data.

Pu and Vosselman [PV09b] proposed a system for segmenting terrestrial LiDAR data in order to fit detailed polygonal façade models. Their method uses least-squares fitting of outline polygons, convex hulls, and concave polygons, and it combines a polyhedral building model with the extracted parts. The reconstruction method is automatic and it aims at detailed façade reconstruction (refer to Section 4.2).

Toshev et al. [TMT10] also presented a method for detecting and parsing of buildings from unorganized 3d point clouds. Their top-down model is a simple and generic grammar fitted by a dependency parsing algorithm, which also generates a semantic description. The output is a set of parse trees, such that each tree represents a semantic decomposition of a building. The method is very scalable and is able to parse entire cities.

Zhou and Neumann [ZN08] presented an approach for automatic reconstructing building models from airborne LiDAR data. This method features vegetation detection, boundary extraction and a data-driven algorithm which automatically learns the principal directions of roof boundaries. The output are polygonal building models. A further extension [ZN10, ZN11] produces polygonal 2.5d models composed of complex roofs and vertical walls. Their approach generates buildings with arbitrarily shaped roofs with high level of detail, which is comparable to that of interactively created models (cf. Figure 13).

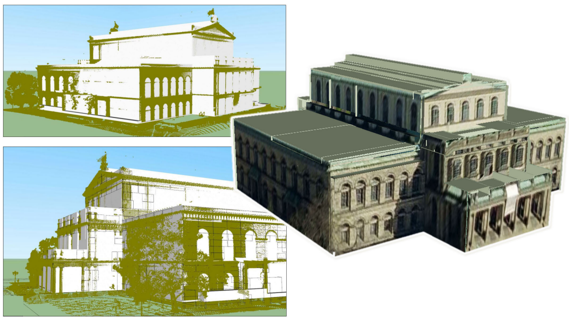

Figure 14: *Automatic reconstruction of a building with volumetric models. For purposes of visual evaluation, the reconstructed volume is superimposed over the original point set, including noise and obstacles (left), and textured with photographs of the buildings (right). Figure courtesy of Carlos Vanegas [VAB12], ©2012 IEEE.*

Recently, Vanegas et al. [VAB12] proposed an approach for the reconstruction of buildings from 3d point clouds with the assumption of Manhattan World building geometry. Their system detects and classifies features in the data and organizes them into a connected set of clusters from which a volumetric model description is extracted (cf. Figure 14). The Manhattan World assumption has been successfully used by several urban reconstruction approaches [FCSS09a, VAW*10], since it robustly allows to identify fundamental shapes of most buildings.

Recently, Korah et al. [KMO11] published a method for segmentation of aerial urban LiDAR scans in order to determine individual buildings, and Shen et al. [SHFH11] proposed a hierarchical façade segmentation method based on repetitions and symmetry detection in terrestrial LiDAR scans (cf. Section 4.2).

While LiDAR data is accessible for quite a while, and methods which aim at robust fitting of top-down models into it deliver good results, the whole potential of this combination is still not fully exhausted, thus, we may expect further interesting papers on this topic in the near future.

3.3. Inverse Procedural Modeling

A new and growing area is that of inverse procedural modeling (IPM), where the framework of grammar-driven model construction is not only used for synthesis, but also for the reconstruction of existing buildings. Traditional forward procedural urban modeling provides an elegant and fast way to generate huge, complex and realistic looking urban sites. A recent survey [VAW*10] presented this approach for the synthesis of urban environments. An inverse methodology is applicable to many types of procedural models, but such an exploration has been quite prolific with respect to building models. The most general form of the inverse procedural modeling problem is to discover both the parameterized grammar rules and the parameter values that, when applied in a particular sequence, yield a pre-specified output.

Discovering both the rules and the parameter values that result in a particular model effectively implies compressing a 3d model down to an extremely compact and parameterized form. Stava et al. proposed a technique to infer a compact grammar from arbitrary 2d vector content [SBM*10]. Bokeloh et al. [BWS10] exploited partial symmetry in existing 3d models to do inverse procedural modeling. Recently, Talton et al. [TLL*11] used a Metropolis-based approach to steer which rules (from a known large set) and parameter values to apply in order to obtain a 3d output resembling a pre-defined macroscopic shape. Benes et al. [BvMM11] defined guided procedural modeling as a method to spatially dividing the rules (and productions) into small guided procedural models that can communicate by parameter exchange in order to obtain a desired output.

Various methods have specialized the inverse framework to the application of building reconstruction, often by assuming that the rules are known – thus inferring only the parameter values. A very complete, yet manual solution to this problem was presented by Aliaga et al. [ARB07]. They interactively extract a repertoire of grammars from a set of photographs of a building and utilize this information in order to visualize a realistic and textured urban model. This approach allows for quick modifications of the architectural structures, like number of floors or windows in a floor. The disadvan-

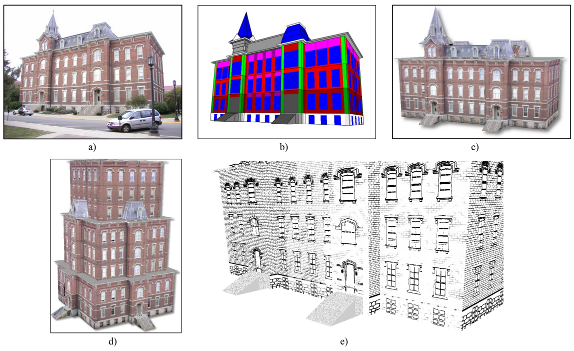

Figure 15: *Example of inverse procedural modeling of a building from a photograph (top) and the application of the grammar to generate novel building variations (bottom). Figure or [ARB07], ©2007 IEEE.*

tage of this approach is the quite labor-intensive grammar creation process.

Another grammar-driven method for automatic building generation from air-borne imagery was proposed by Vanegas et al. [VAB10]. Their method uses a simple grammar for building geometry that approximately follows the Manhattan World assumption. This means that it expects a predominance of the three mutually orthogonal directions. The grammar converts the reconstruction of a building into a sequential process of refining a coarse initial building model (e.g., a box), which they optimize using geometric and photometric matching across images. The system produces complete textures polygonal models of buildings (Figure 16).

Hohmann et al. [HKHF09, HHKF10] presented a modeling system which is a combination of procedural modeling with GML shape grammars [Hav05]. Their method is based on interactive modeling in a top-down manner, yet it contains high-level cues and aims at semantic enrichment of geometric models. Mathias et al. [MMWvG11] reconstruct complete buildings as procedural models using template shape grammars. In the reconstruction process, they let the grammar interpreter automatically decide on which step to take next. The process can be seen as instantiating the template by determining the correct grammar parameters. Another approach where a grammar is fitted from laser-scan data was published by Toshev et al. [TMT10].

Also in the photogrammetry community the idea of IPM has found a wide applicability in papers aiming at reconstruction of buildings and façades: Ripperda and Brenner introduced a predefined façade grammar which they automatically fit from images [BR06, Rip08] and laser scans [RB07, RB09] using the Reversible Jump Markov Chain Monte Carlo (RJMCMC). A similar approach was proposed by Becker and Haala [BH07,BH09,Bec09] but in this system they also propose to automaticaly derive a façade-grammar from the data in a bottom-up manner.

Other work aims on grammar-driven image segmentation. For example, Han and Zhu [HZ05, HZ09] presented a simple attribute graph grammar as a generative representation for made-made scenes and propose a top-down/bottom-up inference algorithm for parsing image content. Is simplifies the objects which can be detected to square boxes in order to limit the grammar space. Nevertheless, this approach provides a good starting point for inverse procedural image segmentation.

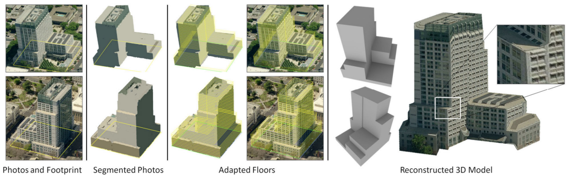

Figure 16: *Results of the automatic method which uses aerial imagery registered to maps and an inverse procedural grammar. Figure courtesy of Carlos Vanegas [VAB10], ©2010 IEEE.*

The field of inverse procedural modeling is relatively new and still not very well researched. For this reason, we expect more exciting papers on this topic in the near future.

4. Façades & Images

In this section we focus on approaches aiming at the reconstruction and representation of façades. In recent years, many different approaches for the extraction of façade texture, structure, façade elements, and façade geometry have been proposed.

First, we discuss façade image processing approaches which aim at an image-based representation of façades. Here we include panorama imaging and projective texturing. Second, we continue with façade-parsing methods. These methods aim at automatic subdivision of façades into their structural elements. Third, we address the topic of interactive façade modeling systems which aim at higher quality and level of detail.

4.1. Façade Image Processing

Imagery is essential in urban reconstruction as both a source of information as well as a source of realism in the final renderings. Additional advantages of imagery are its, in general, simple acquisition process, and also the fact, that there exists an enormous amount of knowledge about its processing. It has been the subject of very active research in the recent two decades. In this section we cover urban panorama imaging as well as texture generation approaches.

Panoramas and Image Stitching. Panoramas are traditionally generated for the purpose of visualizing wide landscapes or similar sights, but in the context of urban reconstruction, panoramas might already be seen as final results of virtual models on its own.

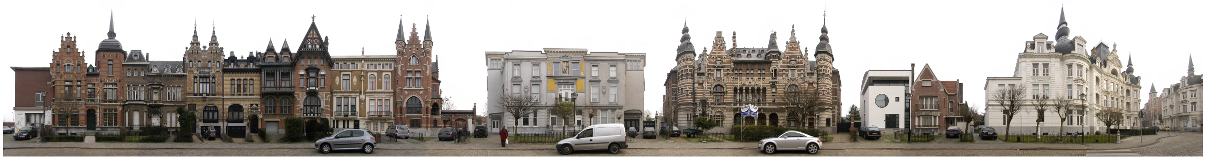

Figure 17: *A multi-viewpoint panorama of a street in Antwerp composed from 107 photographs taken about one meter apart with a hand-held camera. Figure courtesy of Aseem Agarwala [AAC*06], ©2006 ACM.*

In practice, panoramas are composed from several shots taken at approximately the same location [SS02b, Sze06]. For urban environments, often the composed image is generated along a path of camera movement, referred to as strip panorama. The goal of those methods is to generate views with more than one viewpoint in order to provide an approximation of an orthographic projection. Variants of those are pushbroom images, which are orthographic in the direction of motion and perspective in the orthogonal one [GH97, SK03], and the similar x-slit images presented by Zomet et al. [ZFPW03]. Similar approaches for the generation of strip-panoramic images was proposed also by Zheng [Zhe03] and Roman et al. [RGL04]. Agarwala et al. [AAC*06] aim at the creation of long multiview strip panoramas of street scenes, where each building is projected approximately orthogonal on a proxy plane (cf. Figure 17). Optimal source images for particular pixels are chosen using a constrained MRF-optimization process [GG84, KZ04].

Panoramas are usually generated by stitching image content from several sources, often also referred to as photomosaics. The stitching of two signals of different intensity usually causes a visible junction between them. An early solution to this problem were transition zones and multiresolution blending [BA83]. Pérez et al. [PGB03] introduced a powerful method for this purpose: image editing in the gradient domain. There is a number of further papers tackling, improving, accelerating and making use of this idea [PGB03, ADA*04, Aga07, MP08]. Zomet et al. presented an image stitching method for long images [ZLPW06]. The foundations behind the gradient domain image editing method are described in the aforementioned papers as well as in the ICCV 2007 Course-Notes [AR07].

Texture Generation. Another fundamental application of imagery is its necessity for texturing purposes. The particular problem of generating textures for the interactive rendering of 3d urban models can be addressed by *projective texturing* from perspective photographs. Most interactive modeling systems, like "Façade" [DTM96], allow sampling projective textures on the reconstructed buildings. Based on input from video [vdHDT*07c] or image collections [ARB07, SSS*08, XFT*08], they introduce projective texture sampling as part of their modeling pipeline and they rely on user interaction in order to improve the quality of the results.

Others also proposed tools for texturing of existing models, like an interactive approach by Georgiadis et al. [GSGA05], or an automatic by Grzeszczuk et al. [GKVH09]. There are further fully automatic attempts (most of them in the photogrammetry literature) which aim at projective texture generation for existing building models [CT99, WH01, WTT*02, Bö04, OR05, GKKP07, TL07, TKO08, KZZL10].

More tools dedicated to interactive enhancement and inpainting for architectural imagery were presented by Korah and Rasmussen [KR07b] who detected repetitive building parts to inpaint façades, Pavic et al. [PSK06] who proposed an interactive method for completion of building textures, and Musialski et al. [MWR*09] who used translational and reflective symmetry in façade-images to remove unwanted content (cf. Figure 19). Eisenacher et al. [ELS08] used example-based texture synthesis to generate realistically looking building walls.

Recently, some interesting tools for façade imagery processing have exploited the matrix factorization methodology. Matrix factorization allows for good approximation of low-rank matrices with a small number of certain basis functions [Str05]. Façade images are usually of low-rank due to many orthogonal and repetitive patterns. The approach presented by Ali et al. [AYRW09] utilizes factorization for a compression algorithm in order to overcome a memory transfer bottleneck and to render massive urban models directly from a compressed representation. Another method proposed by Liu et al. [LMWY09, LMWY12] aims at inpainting of missing image data. Their algorithm is built on studies about matrix completion using the trace norm and relaxation techniques. Façades are well suited for such algorithms due to many repetitions (cf. Figure 18).

While processing of urban imagery is basically a well researched topic, it still provides some challenges. Especially the issue of segmentation of façades is an active research direction, and we will elaborate on it in the next section.

4.2. Façade Parsing

The term *façade parsing* denotes methods which aim at automatic detection of structure in façade data (i.e., images or laser scans). While recent interactive algorithms, which we review in the next section, deliver very good results, automatic façade parsing is still an error-prone problem.

In the first step, façade imagery is usually processed with

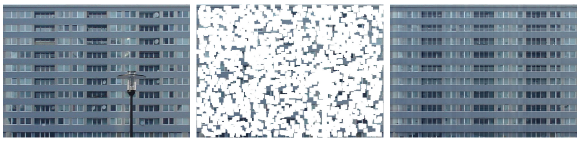

Figure 18: *Façade in-painting. The left image is the original image. Middle: the lamp and satellite dishes together with a large set of randomly positioned squares has been selected as missing parts (80% of the façade shown in white). The right image is the result of the tensor completion algorithm proposed in [LMWY09], ©2009 IEEE.*

classic image processing methods, like edge [Can86], corner [HS88] and feature [Low04, BETvG08] detection as basic tools to infer low-level structure. We omit low-level processing and for details we refer to textbooks, e.g., Gonzales and Woods [GW08], or Sonka et al. [SHB08].

The next step is to employ the low-level cues in order to infer more sophisticated structure, like floors or windows. Most earlier attempts were based on locally acting filtering and splitting heuristics, but it turned out that such segmentation ist not enough to reliably detect structure in complex façades. The necessity of higher-order structure has emerged, thus, many methods turned to symmetry detection, which is widely present in architecture. These approaches often combine the low-level cues with *unsupervised clustering* [HTF09], with searching and matching algorithms, as well as with Hough transforms. Another trend of current research is towards *machine learning* [Bis09, HTF09] in order to fit elements in databases, or to infer façade structure with predefined grammars or rules. In this section we provide an overview over these various approaches.

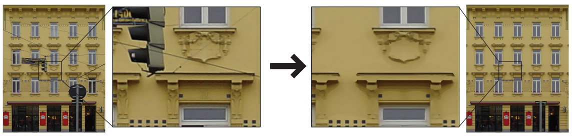

Figure 19: *The input image on the left contains a traffic light and several cables. To the right the unwanted objects have been successfully removed by utilizing the symmetry in the façade image [MWR*09].*

Heuristic Segmentation. Wang and Hanson [WH01] and Wang et al. [WTT*02] proposed a system which aims at the generation of textured models and the detection of windows. They introduced a façade texture based on the weighted average of several source images projected on a (previously registered) block model, which serves both for texturing and for detection of further detail (i.e. windows). They proposed a heuristic oriented region growing algorithm which iteratively enlarges and synchronizes small seed-boxes until they best fit the windows in the texture. Another use of local image segmentation and heuristics is presented by Tsai et al. [TLLH05], who calculate a "greenness index" to identify and suppress occlusions by vegetation on façade textures extracted from drive-by video sequences. They detect local mirror axes of façade parts in order to cover holes left after removing the occluding vegetation. On the cleaned textures they also apply oriented region growing in order to detect windows. Further extensions to their method, like processing of video input, are covered in [TLH06,TCLH06]. In both methods the used assumptions, e.g., that windows are darker than their surrounding façade, or the "greenness index", are, however, weak and often provide erroneous results.

Lee and Nevatia [LN04] proposed a segmentation method that uses only edges. They project the edges horizontally and vertically to get the marginal edge-pixel distributions and assume that these have peaks where window-frames are located. From the thresholded marginal distributions they construct a grid which approximates a subdivision of the façade. While the subdivisions are often quite good, on the downside, this approach depends very strongly on the parameters of the edge detector.

Symmetry and Pattern Detection. Symmetry abounds in typical architecture, which is mostly the result of economical manufacturing as well as for aesthetic reasons. In fact, symmetry is a topic that has inspired mankind from the beginning. Recent approaches try to detect the inherent symmetry of a façade in order to infer some information about its structure.

In image processing, early attempts include [RWY95], who introduced a continuous *symmetry transform* for images. Later, Schaffalitzky and Zisserman [SZ99] detected groups of repeated elements in perspective images, and Turina et al. [TTvG01, TTMvG01] detected repetitive patterns on planar surfaces, also under perspective skew, using Hough transforms. They demonstrated that their method works well on building façades. Further, a considerable amount of work on this topic has been done by Liu and collaborators [LCT04]. They detected crystallographic groups in repetitive image patterns using a dominant peak extraction method from the autocorrelation surface. Other image processing approaches utilized the detected symmetry of regular [HLEL06] and near-regular patterns [LLH04, LBHL08] in order to model new images.

Further approaches specialized on detecting affine symmetry groups in 2d images [LHXS05,LE06] and in 3d point clouds [MGP06, PSG*06]. Follow-ups of those methods introduced data-driven modeling frameworks for symmetrization [MGP07] and 3d lattice fitting (cf. Figure 20) in laser-scans of architecture [PMW*08, MBB10] .

The work finally boiled down to the insight that the repetitive nature of façade elements can be exploited to segment them. Berner et al. [BBW*08, BWM*11] and Bokeloh et al. [BBW*09] proposed a set of methods to detect symmetry in ground-based urban laser scans. A heuristic segmentation based on detection of symmetry and repetitions was proposed by Shen et al. [SHFH11]. Their method segments LiDAR scans of façades and detects concatenated grids. It automatically partitions the façade in an adaptive manner, such that a hierarchical representation is generated.

Detection of repeated structures in façade images was approached by Wenzel et al. [WDF08], and Musialski et al. [MRM*10], who proposed methods to detect rectilinear patterns in orthographic-rectified façade images. A similar method was also introduced by Zhau and Quan [ZQ11]. Other detect symmetry directly in perspective images. For example, Wu et al. [WFP10] proposed a method to detect grid-like symmetry in façade images under perspective skew, which they have used to reconstruct dense 3d structure in a follow-up work [WACS11]. Park et al. [PBCL10] introduced a method detect translational symmetry in order to determine façades. Recently, Nianjuan et al. [NTC11] proposed a method for detecting symmetry across multiview networks of urban imagery. A similar setup was used by Ceylan et al. [CML*12] in order to detect reliable symmetry across multiple registered images, which is utilized to recover missing structure of buildings.

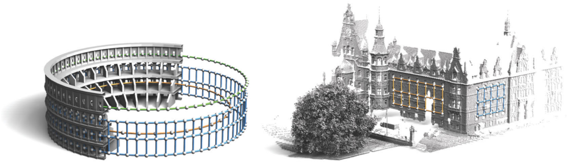

Figure 20: *This example shows automatic symmetry detection results performed on point-clouds of architectural objects. Figure courtesy of Mark Pauly [PMW*08], ©2008 ACM.*

Window Detection. Another group of methods specializes at the detection of windows and other pre-specified structural elements. Some rely on template matching, others try to detect more general shapes, like simple rectangles. The advantage of template matching is that the results look very realistic. However, the disadvantage is that the windows are in most cases not authentic because there is no template database that contains all possible shapes.

For example, Schindler and Bauer [SB03] matched shape templates against dense point clouds. Also Mayer and Reznik [MR07] matched template images from a manually constructed window image database against their façades. Müller et al. [MZWvG07] matched the appearance of their geometric 3d window models against façade image-tiles.

Some approaches combine template matching with machine learning, e.g., Ali et al. [ASJ*07], who proposed to train a classifier, or Drauschke et al. [DF08], who used Adaboost [SS99]. These systems identify a high percentage of windows even in images with perspective distortion.

Another approach, which is based on rectangles, is the window-pane detection algorithm by Cech and Sara [CS08], which identifies strictly axis-aligned rectangular pixel configurations in a MRF. Given the fact that the majority of windows and other façade elements are rectangular, a common approach to façade reconstruction is searching for rectangles or assuming that all windows are rectangular. Also Haugeard et al. [HPFP09] introduced an algorithm for inexact graph matching, which is able to extract rectangular window as a sub-graph of the graph of all contours of the façade image. This serves as an basis to retrieve similar windows from a database of images of façades.

Almost all methods discussed here somehow assume rectangular shapes in some stages of their algorithms, but do not solely rely on it.

Higher-Order Knowledge Models. Here we discuss a class of solutions that aim at knowledge-based object reconstruction, which means that they employ an a-priori top-down model that is supposed to be fitted by cues derived from the data. In fact, some methods utilize the concept of inverse procedural modeling presented in Section 3.3.

Quite a number of approaches proposed grammar-based models. For example, Alegre and Dellaert [AD04] introduced a set of rules from a stochastic context-free attribute grammar, and a Markov Chain Monte Carlo (MCMC) solution to optimize the parameters. Mayer and Reznik [MR05, MR06, MR07] and Reznik and Mayer [RM07] published a series of papers in which they present a system for façade reconstruction and window detection by fitting an implicit shape model [LLS04], again using MCMC optimization.

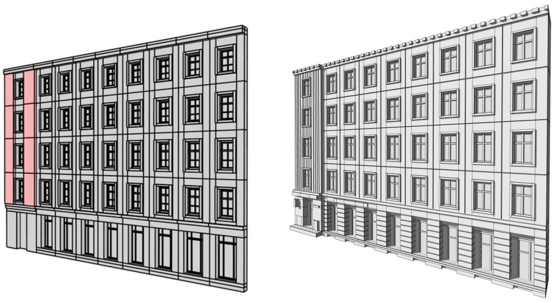

Figure 21: *Comparison of the results of the automatic method of [MZWvG07] (left, 409 shapes, excluding windows matched from a template library) to the interactive method of [MWW12] (right, 1,878 shapes). Left image courtesy of Pascal Müller [MZWvG07].*

A single-view approach for rule extraction from a segmentation of simple regular façades was published by Müller et al. [MZWvG07]. They cut the façade image into floors and tiles in a synchronized manner in order to reduce it to a so-called irreducible form, and subsequently fit CGA-rules into the detected subdivision. This method is limited to rectilinearly distributed façades (cf. Fig. 21). Van Gool et al. [vGZBM07] provided an extension which detects similarity chains in perspective images and a method to fit shape grammars to these.

Korah and Rasmussen introduced a method for automatic detection of grids [KR07a]. Also Tylecek and Sara [TS10] pursued a similar approach, where both systems detect grids of windows in ortho-rectified façade images using a weak prior and MCMC optimization. Brenner and Ripperda [BR06, RB07, Rip08, RB09] developed in a series of publications a system for detecting façade elements and especially windows from images and laser scans. In this work, a context-free grammar for façades is derived from a set of façade images and fitted to new models using the Reversible Jump Markov Chain Monte Carlo technique (RJMCMC). Becker and Haala [BH07, BHF08, Bec09, BH09] presented in a series of papers a system which attempts to automatically discover a formal grammar. This system was designed for reconstruction of façades from a combination of LiDAR and image data.

Pu and Vosselman proposed a higher-order knowledge-driven system which automatically reconstructs façade models from ground laser-scan data [PV09b]. In a further approach, they combine information from terrestrial laser point clouds and ground images. The system establishes the general structure of the façade using planar features from laser data in combination with strong lines in images [PV09a, PV09c].

This topic is also of wide interest in the computer vision community. In an automatic approach, Koutsourakis [KST*09] examines a rectified façade image in order to fit a hierarchical tree grammar. This task is formulated as a Markov Random Field [GG84], where the tree formulation of the façade image is converted into a shape grammar responsible for generating an inverse procedural model (cf. Section 3.3). Teboul et al. [TSKP10] extend this work by combining a bottom-up segmentation through superpixels with top-down consistency checks coming from style rules. The space of possible rules is explored efficiently. In a recent follow-up they improve their method by employing reinforcement learning [TKS*11].

Recently, Sunkel et al. [SJW*11] presented a user supervised technique that learns line features in geometrical models.

While recent approaches based on inverse procedural modeling provide quite stable results, the quality and the level of detail of these methods is still not good enough for current demands. In practice, the expected quality for production is much higher, therefore manual or interactive methods still have wide applicability.

4.3. Façade Modeling

The previous section presented an overview of automatic façade-subdivision approaches. All these methods share the property that they create models of low or intermediate level of detail and complexity. Interactive approaches, on the other hand, promise better quality and higher level of detail.

An interactive image-based approach to façade modeling was introduced by Xiao et al. [XFT*08]. It uses images captured along streets and also relies on structure from motion as a source for camera parameters and initial 3d data. It considers façades as flat rectangular planes or simple developable surfaces with an associated texture. Textures are composed from the input images by projective texturing. In the next step, the façades are automatically subdivided using a split heuristic based on local edge detection [LN04]. This subdivision is then followed by an interactive bottom-up merging process. The system also detects reflectional symmetry and repetitive patterns in order to improve the merging task. Nonetheless, the system requires a considerable amount of user interaction in order to correct misinterpretations of the automatic routines.

Hohmann et al. [HKHF09] proposed a system for modeling of façades based based on the GML shape grammar [Hav05]. Similar as in the work of Aliaga et al. [ARB07], grammar rules are determined manually on the façade imagery and can be used for procedural remodeling of similar buildings.

Another interactive method for the reconstruction of façades from terrestrial LiDAR data was proposed by Nan et al. [NSZ*10], which is based on semi-automatic snapping of small structural assemblies, called SmartBoxes. We mention the method also in Section 3.2.

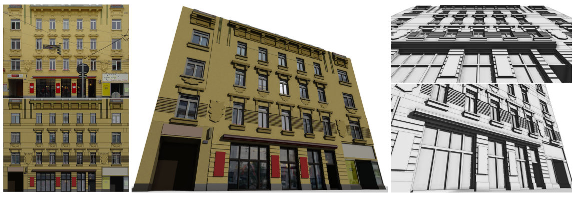

Figure 22: *Results of interactive modeling with the method of Musialski et al. [MWW12]. The façade image has been segmented into 1346 elements. ©2012 The Eurographics Association and Blackwell Publishing Ltd.*

Recently, Musialski et al. [MWW12] introduced a semi-automatic image-based façade modeling system. Their approach incorporates the notion of coherence, which means that façade elements that exhibit partial symmetries across the image can be grouped and edited in a synchronized manner. They also propose a modeling paradigm where the user is in control of the modeling workflow, but is supported by automatic modeling tools, where they utilize unsupervised clustering in order to robustly detect significant elements in orthographic façade images. Their method allows modeling high detail in competitive time (cf. Figure 22).

While interactive methods seem to be too slow and not well scalable, the advantage of the high-quality output is a considerable value (refer to Figure 21). For this reason, we believe that with the plethora of research in automatic computer vision algorithms, it will become equally important to study the efficient integration of automatic processing and user interaction in future.

5. Blocks & Cities

The problem of measuring and documenting the world is the objective of the photogrammetry and remote sensing community. In the last two decades this problem has been also extended to automatic reconstruction of large urban areas or even whole urban agglomerations. Additionally, also the computer vision and computer graphics communities started contributing to the solutions. In this section we want to mention several modern approaches which have been proposed in this vast research field.

The common property of large-scale approaches is the demand of minimal user interaction or, in the best case, no user interaction at all, which leads to the best possible scalability of the algorithms. There is quite a variety of methods, which either work with aerial or ground-level input data or both. It is difficult to compare these methods directly to each other since they have been developed in different contexts (types of input data, types of reconstructed buildings, level of interactivity, etc.). For this reason we do not attempt a comparison; we will merely review the mentionable approaches and state their main contributions and ideas.

In large scale reconstruction, there is a trend towards multiple input data types. Some publications involve aerial and ground-based input, some also combine LiDAR with imagery. Other methods introduce even more data sources, like a digital elevation model (DEM), a digital terrain model (DTM), or a digital surface model (DSM). Finally, some methods incorporate positioning systems, like the global positing system (GPS), or local inertial navigation systems (INS). We omit a detailed discussion on remote sensing concepts and refer to further literature [CW11]. A number of papers up to the year 2003 have been also reviewed in a survey by Hu et al. [HYN03].

5.1. Ground Reconstruction

One of the earlier approaches to reconstruct large urban areas was the work of Früh and Zakhor. They published a series of articles that aim at a fully automatic solution which combines imagery with LiDAR. First they proposed an approach for automated generation of textured 3d city models with both high detail at ground level and complete coverage for the bird's-eye view [FZ03]. A close-range façade model is acquired at the ground level by driving a vehicle equipped with laser scanners and a digital camera under normal traffic conditions on public roads. A far-range digital surface model (DSM), containing complementary roof and terrain shape, is created from airborne laser scans, then triangulated, and finally texture-mapped with aerial imagery. The façade models are first registered with respect to the DSM using Monte Carlo localization, and then merged with the DSM by removing redundant parts and filling gaps. In further work [FZ04], they improved their method for ground-based acquisition of large-scale 3d city models. Finally, they provided a comprehensive framework which features a set of data-processing algorithms for generating textured façade meshes of cities from a series of vertical 2d surface scans and camera images [FJZ05].

In the realm of image-based methods, Pollefeys et al. [PvGV*04] presented an automatic system to build visual models from images. This work is also one of the papers which pioneers fully automatic structure from motion of urban environments. The system deals with uncalibrated image sequences acquired with a hand-held camera and is based on features matched across multiple views. From these both the structure of the scene and the motion of the camera are retrieved (cf. Section 2.2). This approach was further extended by Akbarzadeh et al. [AFM*06] as well as Pollefeys et al. [PNF*08].

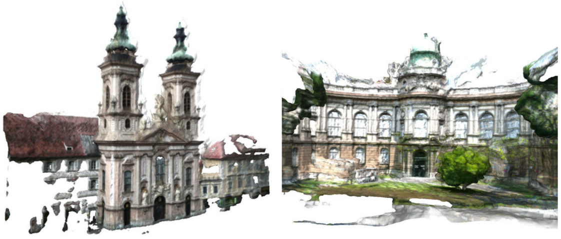

Figure 23: *Examples of dense reconstruction after depth map fusion. Figure courtesy of Arnold Irschara [IZB07], ©2007 IEEE.*

Another image-based approach ist the work of Irschara et al. [IZB07, IZKB12] which provides a combined sparse-dense method for city reconstruction from unstructured photo collections contributed by end users. Hence, the "wiki" principle, well known from textual knowledge databases, is transferred to the goal of incrementally building a virtual representation of a local habitat. Their approach aims at large scale reconstruction, using a vocabulary tree [NS06] to detect mutual correspondences among images, and combines sparse point clouds, camera networks, and dense matching in order to provide very detailed buildings (cf. Figure 23).

A ground-level city modeling framework which integrates reconstruction and object detection was presented by Cornelis et al. [CLCvG08]. It is based on a highly optimized 3d reconstruction pipeline that can run in real-time, hence offering the possibility of online processing while the survey vehicle is recording. A compact textured 3d model of the recorded path is already available when the survey vehicle returns to its home base (cf. Figure 24). The second component is an object detection pipeline, which detects static and moving cars and localizes them in the reconstructed world coordinate system.

Xiao et al. [XFZ*09] proposed to extend their previous method [XFT*08] in order to provide an automatic approach to generate street-side photo-realistic 3d models from images captured along the streets at ground level. They employ a multiview segmentation algorithm that recognizes and segments each image at pixel level into semantically meaningful classes, such as building, sky, ground, vegetation, etc. With a partitioning scheme the system separates buildings into independent blocks, and for each block, it analyzes the façade structure using priors of building regularity. The system produces visually compelling results, however it clearly suffers quality loss when compared to their previous, interactive approach [XFT*08].

Another system introduced by Grzeszczuk et al. [GKVH09] aims for fully automatic texturing of large urban aerials using existing models from GIS databases and unstructured ground-based photographs. It employs SfM to register the images to each other in the first step, and than the ICP algorithm [BM92] in order to align the SfM 3d point clouds with the polygonal geometry from GIS databases. In further steps their system automatically selects optimal images in order to provide projective textures to the building models.

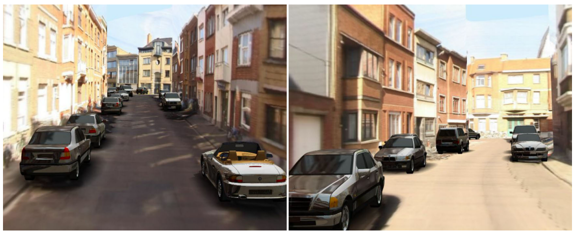

Figure 24: *A collection of rendered images from the final 3d city model taken from various vantage points. Figure courtesy of Nico Cornelis [CLCvG08], ©2008 Springer.*

In general, ground-based systems are usually limited to relatively small areas if compared to airborne approaches. In the other hand, these methods are the only ones to provide small-scale details, thus, the objective is often the combination of both acquisition methods.

5.2. Aerial Reconstruction

Aerial imagery is perhaps the most often used data source for reconstruction of urban environments, and has been explored in the photogrammetry and remote sensing community for many years. There has been a significant number of successful approaches in the past decade, like those of Baillard et al. [BZ99], the group of Nevatia et al. [NN01, NP02, KN04], or Jaynes et al. [JRH03].

Many approaches often combine imagery with other input data. In this section we review several systems developed in recent years.

Wang et al. [WYN07] combined both aerial and ground-based imagery in a semiautomatic approach. The framework stitches the ground-level images into panoramas in order to obtain a wide camera field of view. It also detects the footprints of buildings in orthographic aerial images automatically, and both sources are combined, where the system incorporates some amount of user interaction in order to correct wrong correspondences.

Another multi-input method was proposed by Zebedin et al. [ZBKB08]. This framework combines aerial imagery with additional information from DEMs. They introduced an algorithm for fully automatic building reconstruction, which combines sparse line features and dense depth data with a global optimization algorithm based on graph cuts [KZ04]. Their method also allows generating multiple LODs of the geometry. Also Karantzalos and Paragios [KP10] proposed a framework for automatic 3d building reconstruction by combining images and DEMs. They developed a generalized variational framework which addresses large-scale reconstruction by utilizing hierarchical grammar-based 3d building models as a prior. They use an optimization algorithm on the GPU to efficiently fit grammar-instance from the information extracted from images and the DEM.

A recent method of Mastin et al. [MKF09] introduced a method for fusion of 3d laser data and aerial imagery. Their work employs *mutual information* for registration of images with LiDAR point clouds, which exploits the statistical dependency in urban scenes. They utilize the downhill simplex optimization to infer camera pose parameters and propose three methods for measuring mutual information between LiDAR and optical imagery.

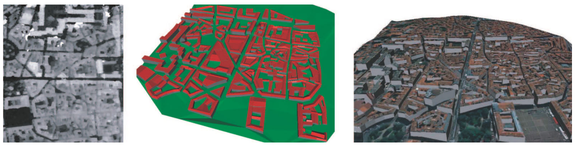

Figure 25: *Automatic urban area reconstruction results from a DSMs (left): without (middle) and with textures (right). Figure courtesy of Florent Lafarge [LDZPD10], ©IEEE 2010.*

Another source for large-scale reconstruction is a digital surface model (DSM), which can be obtained automatically from aerial and satellite imagery. Lafarge et al. [LDZPD10] proposed to use a DSM in order to extract individual building models. It treats each building as an assembly of simple 3d parametric blocks, which are placed on the DSM by 2d matching techniques, and then optimized using a MCMC solver. The method provides individual buildings models of urban areas (cf. Figure 25).

5.3. Massive City Reconstruction

In this section we mention several methods which employ fully automatic methodologies and also provide reconstructions of entre urban areas. One significant factor which allows for such vast reconstruction is the general technological progress in the data acquisition process, such as the easy access to huge collections of images on the internet, or the presence of many accurate and large LiDAR data sets. The second, perhaps more important, factor is the development of smart and scalable reconstruction algorithms. No hardware advantage will compensate for exponentially scaling approaches, thus development of such algorithms is still a challenge.

In the image-based reconstruction domain, an impressive system was recently presented by Frahm et al. [FFGG*10]. It is capable of delivering dense structure from unstructured Internet images within one day on a single PC. Their framework extends to the scale of millions of images, what they achieve by extending state-of-the-art methods for appearance-based clustering, robust estimation, and stereo fusion (cf. Section 2), and by parallelizing the tasks which can be efficiently processed on multi-core CPUs and modern graphics hardware.

Poullis and You introduced a method for massive automatic reconstruction from images and LiDAR [PY09a, PY09b, PY09c]. Their system automatically creates lightweight, watertight polygonal 3d models from airborne LiDAR. The technique is based on a statistical analysis of the geometric properties of the data and makes no particular assumptions about the input. It is able to reconstruct areas containing several thousand buildings, as shown in Figure 26. Recently they extended their method for texturing [PY11].

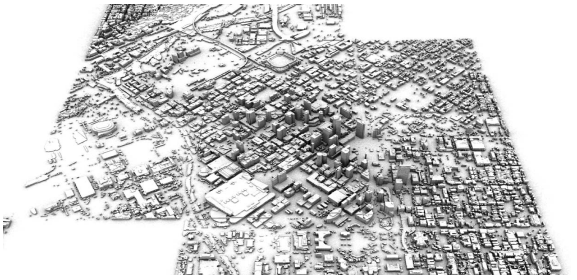

Figure 26: *Large scale reconstruction of Downtown Denver and surrounding areas. The model is a polygonal mesh generated from air-borne LiDAR data. Figure courtesy of Charalambos Poullis [PY09a]. ©2009 IEEE.*

Lafarge and Mallet [LM11a] published an approach which aims at even more complete modeling from aerial LiDAR. Its advantage is that it not only reconstructs building models, but also the inherent vegetation and complex grounds. Furthermore, it is also generalized such that it can deal with unspecified urban environments, e.g., with business districts as well as with small villages. Geometric 3d-primitives such as planes, cylinders, spheres or cones are used to describe regular roof sections, and are combined with mesh-patches to represent irregular components. The various geometric components interact through a non-convex optimization solver. Their system provides impressive large-scale results as shown in Figure 27.

Also Zhou and Neumann proposed a similar approach [ZN09, ZN11]. Generally, while the results of recent methods are very impressive, automatic large-scale reconstruction remains an open problem. With the goal of very detailed and dense virtual urban habitats, the problem still remains a very difficult one. The challenges lie in the management and processing of huge amounts of data, in the developments of robust automatic as well as fast and scalable algorithms, and finally, in the integration of many different types of data.

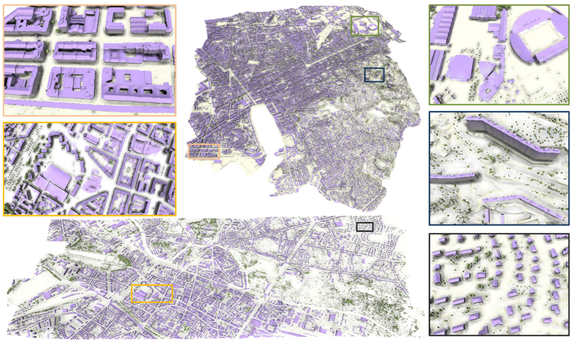

Figure 27: *Reconstruction of two large urban environments with closeup crops. Figure courtesy of Florent Lafarge [LM11a]. ©2011 IEEE.*

6. Conclusions

Despite many contributions to urban reconstruction, we see a number of important open problems remaining for future work.

Many automatic algorithms rely on some assumptions that are often not met in practice. We believe that the combination of interactive techniques and automatic algorithms can help to make significant progress towards the generation of even higher quality models.

In computer vision and data mining, there are several excellent examples of how the analysis of large photo collections can lead to impressive results. We believe that the investigation of how to efficiently combine the effort of many users for data collection or modeling is an area ripe for important future contributions.

Finally, if high quality urban models become available to more researchers, we believe that the analysis of these models as well as the investigation of novel applications will become more attractive to a wider audience.

References

[AAC*06] Agarwala A., Agrawala M., Cohen M., Salesin D., Szeliski R.: Photographing long scenes with multi-viewpoint panoramas. *ACM Transactions on Graphics 25*, 3 (July 2006), 853. 14

[AD04] Alegre F., Dellaert F.: A Probabilistic Approach to the Semantic Interpretation of Building Facades. In *International Workshop on Vision Techniques Applied to the Rehabilitation of City Centres, 2004* (2004). 16

[ADA*04] Agarwala A., Dontcheva M., Agrawala M., Drucker S., Colburn A., Curless B., Salesin D., Cohen M.: Interactive digital photomontage. *ACM Transactions on Graphics 23*, 3 (Aug. 2004), 294. 14

[AFM*06] Akbarzadeh A., Frahm J.-M., Mordohai P., Clipp B., Engels C., Gallup D., Merrell P., Phelps M., Sinha S. N., Talton B., Wang L., Yang Q., Stewenius H., Yang R., Welch G., Towles H., Nister D., Pollefeys M.: Towards Urban 3D Reconstruction from Video. In *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)* (June 2006), IEEE, pp. 1–8. 7, 18

[AFS*10] Agarwal S., Furukawa Y., Snavely N., Curless B., Seitz S. M., Szeliski R.: Reconstructing Rome. *Computer 43*, 6 (June 2010), 40–47. 7

[AFS*11] Agarwal S., Furukawa Y., Snavely N., Simon I., Curless B., Seitz S. M., Szeliski R.: Building Rome in a day. *Communications of the ACM 54*, 10 (Oct. 2011), 105. 7

[Aga07] Agarwala A.: Efficient gradient-domain compositing using quadtrees. *ACM Transactions on Graphics 26*, 3 (July 2007), 94. 14

[AR07] Agrawal A., Raskar R.: Gradient Domain Manipulation Techniques in Vision and Graphics, 2007. 14

[ARB07] Aliaga D. G., Rosen P. A., Bekins D. R.: Style grammars for interactive visualization of architecture. *IEEE Transactions on Visualization and Computer Graphics 13*, 4 (2007), 786–97. 12, 13, 14, 17

[ASJ*07] Ali H., Seifert C., Jindal N., Paletta L., Paar G.: Window Detection in Facades. In *14th International Conference on Image Analysis and Processing (ICIAP 2007)* (Sept. 2007), IEEE, pp. 837–842. 16

[ASS*09] Agarwal S., Snavely N., Simon I., Seitz S. M., Szeliski R.: Building Rome in a day. In *2009 IEEE 12th International Conference on Computer Vision* (Sept. 2009), IEEE, pp. 72–79. 7

[ASSS10] Agarwal S., Snavely N., Seitz S. M., Szeliski R.: Bundle adjustment in the large. In *ECCV 2010* (Sept. 2010), pp. 29–42. 2, 6

[ASW*12] Arikan M., Schwärzler M., Wimmer M., Flöry S., Maierhofer S.: O-Snap: Optimization-Based Snapping for Modeling Architecture. *ACM Transactions on Graphics* (2012), under review. 10

[AYRW09] Ali S., Ye J., Razdan A., Wonka P.: Compressed facade displacement maps. *IEEE Transactions on Visualization and Computer Graphics 15*, 2 (2009), 262–73. 14

[B04] Böhm J.: Multi-image fusion for occlusion-free facade texturing. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 35*, 5 (2004), 867–872. 14

[B08] Böhm J.: Facade Detail from Incomplete Range Data. In *ISPRS Congress Beijing 2008, Proceedings of Commission V* (2008). 11

[BA83] Burt P. J., Adelson E. H.: A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics 2*, 4 (Oct. 1983), 217–236. 14

[BBW*08] Berner A., Bokeloh M., Wand M., Schilling A., Seidel H.-P.: A Graph-Based Approach to Symmetry Detection. In *Volume Graphics* (2008), Hege H.-C., Laidlaw D. H., Pajarola R., Staadt O. G., (Eds.), Eurographics Association, pp. 1–8. 15

[BBW*09] Bokeloh M., Berner A., Wand M., Seidel H. P., Schilling A.: Symmetry Detection Using Feature Lines. *Computer Graphics Forum 28*, 2 (Apr. 2009), 697–706. 15

[Bec09] Becker S.: Generation and application of rules for quality dependent façade reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing 64*, 6 (Nov. 2009), 640–653. 13, 17

[BETvG08] Bay H., Ess A., Tuytelaars T., van Gool L.: Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding 110*, 3 (June 2008), 346–359. 5, 15

[BH07] Becker S., Haala N.: Refinement of Building Facades by Integrated Processing of LIDAR and Image Data. In *PIA07. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* (Sept. 2007), University of Stuttgart : Collaborative Research Center SFB 627 (Nexus: World Models for Mobile Context-Based Systems). 13, 17

[BH09] Becker S., Haala N.: Grammar supported facade reconstruction from mobile lidar mapping. In *ISPRS Workshop, CMRT09 - City Models, Roads and Traffic* (2009), vol. XXXVIII. 13, 17

[BHF08] Becker S., Haala N., Fritsch D.: Combined knowledge propagation for facade reconstruction. In *ISPRS Congress Beijing 2008, Proceedings of Commission V* (2008). 17

[Bis09] Bishop C. M.: *Pattern recognition and machine learning*. Springer, 2009. 15

[BKS*03] Bauer J., Karner K., Schindler K., Klaus A., Zach C.: Segmentation of building models from dense 3D point-clouds. In *27th Workshop of the Austrian Association for Pattern Recognition* (2003). 10, 11

[BM92] Besl P. J., McKay N. D.: A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence 14*, 2 (1992), 239–256. 19

[BR06] Brenner C., Ripperda N.: Extraction of Facades using rjMCMC and Constraint Equations. In *PCV '06, Photogrammetric Computer Vision* (Bonn, 2006), ISPRS Comm. III Symposium, IAPRS, pp. 155–160. 13, 17

[BTvG06] Bay H., Tuytelaars T., van Gool L.: SURF: Speeded Up Robust Features. In *Computer Vision - ECCV 2006* (2006), Leonardis A., Bischof H., Pinz A., (Eds.), vol. 3951 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 404–417. 5

[BvMM11] Beneš B., Šť'ava O., Měch R., Miller G.: Guided Procedural Modeling. *Computer Graphics Forum 30*, 2 (Apr. 2011), 325–334. 12

[BWM*11] Berner A., Wand M., Mitra N. J., Mewes D., Seidel H.-P.: Shape Analysis with Subspace Symmetries. *Computer Graphics Forum 30*, 2 (Apr. 2011), 277–286. 15

[BWS10] Bokeloh M., Wand M., Seidel H.-P.: A connection between partial symmetry and inverse procedural modeling. *ACM Transactions on Graphics 29*, 4 (July 2010), 1. 12

[BZ99] Baillard C., Zisserman A.: Automatic reconstruction of piecewise planar models from multiple views. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)* (1999), IEEE Comput. Soc, pp. 559–565. 19

[BZB06] Bauer J., Zach C., Bischof H.: Efficient Sparse 3D Reconstruction by Space Sweeping. In *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)* (June 2006), IEEE, pp. 527–534. 7

[Can86] Canny J.: A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8*, 6 (Nov. 1986), 679–698. 15

[CLCvG08] Cornelis N., Leibe B., Cornelis K., van Gool L.: 3D Urban Scene Modeling Integrating Recognition and Reconstruction. *International Journal of Computer Vision 78*, 2-3 (Oct. 2008), 121–141. 18, 19

[CLP10] Chauve A.-L., Labatut P., Pons J.-P.: Robust piecewise-planar 3D reconstruction and completion from large-scale unstructured point data. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (June 2010), IEEE, pp. 1261–1268. 7

[CML*12] Ceylan D., Mitra N. J., Li H., Weise T., Pauly M.: Factored Facade Acquisition using Symmetric Line Arrangements. *Computer Graphics Forum (Proc. EG'12) 31*, 1 (May 2012). 16

[Col96] Collins R. T.: A space-sweep approach to true multi-image matching. In *Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1996), IEEE Comput. Soc. Press, pp. 358–363. 6, 10

[COSH11] Crandall D., Owens A., Snavely N., Huttenlocher D.: Discrete-continuous optimization for large-scale structure from motion. In *CVPR 2011* (June 2011), IEEE, pp. 3001–3008. 7

[CR99] Cipolla R., Robertson D.: 3D models of architectural scenes from uncalibrated images and vanishing points. In *Proceedings 10th International Conference on Image Analysis and Processing* (1999), vol. 0, IEEE Comput. Soc, pp. 824–829. 9

[CRB99] Cipolla R., Robertson D., Boyer E.: Photobuilder - 3d models of architectural scenes from uncalibrated images. *IEEE INT. CONF. ON MULTIMEDIA COMPUTING AND SYSTEMS* (1999), 25 – 31. 9

[CS08] Cech J., Sara R.: Windowpane Detection based on Maximum Aposteriori Probability Labeling. In *Image Analysis - From Theory to Applications* (2008). 16

[CT99] Coorg S., Teller S.: Extracting textured vertical facades from controlled close-range imagery. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)* (1999), IEEE Comput. Soc, pp. 625–632. 10, 14

[CW11] Campbell J. B., Wynne R. H.: *Introduction to Remote Sensing, Fifth Edition*. Guilford Publications, 2011. 2, 18

[DF08] Drauschke M., Förstner W.: Selecting appropriate features for detecting buildings and building parts. In *21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS), Beijing, China* (2008). 16

[DTC00] Dick A., Torr P., Cipolla R.: Automatic 3D Modelling of Architecture. In *British Machine Vision Conference, BMVC* (2000), pp. 273–289. 10

[DTC04] Dick A., Torr P. H. S., Cipolla R.: Modelling and Interpretation of Architecture from Several Images. *International Journal of Computer Vision 60*, 2 (Nov. 2004), 111–134. 10

[DTM96] Debevec P. E., Taylor C. J., Malik J.: Modeling and rendering architecture from photographs. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH '96* (New York, New York, USA, 1996), ACM Press, pp. 11–20. 8, 9, 10, 14

[EhWG05] El-hakim S., Whiting E., Gonzo L.: 3D modelling with reusable and integrated building blocks. In *7 th Conference on Optical 3-D Measurement Techniques* (Vienna, 2005), pp. 3–5. 9

[EhWGG05] El-hakim S., Whiting E., Gonzo L., Girardi S.: 3D Reconstruction of Complex Architectures from Multiple Data. In *Proceedings of the ISPRS Working Group V/4 Workshop 3D-ARCH 2005: "Virtual Reconstruction and Visualization of Complex Architectures"* (Mestre-Venice, Italy, 2005). 9

[ELS08] Eisenacher C., Lefebvre S., Stamminger M.: Texture Synthesis From Photographs. *Computer Graphics Forum 27*, 2 (Apr. 2008), 419–428. 14

[FB81] Fischler M. A., Bolles R. C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM 24*, 6 (June 1981), 381–395. 6

[FCSS09a] Furukawa Y., Curless B., Seitz S. M., Szeliski R.: Manhattan-world stereo. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (June 2009), IEEE, pp. 1422–1429. 7, 12

[FCSS09b] Furukawa Y., Curless B., Seitz S. M., Szeliski R.: Reconstructing Building Interiors from Images. In *2009 IEEE 12th International Conference on Computer Vision* (2009), IEEE. 7

[FCSS10] Furukawa Y., Curless B., Seitz S. M., Szeliski R.: Towards Internet-scale Multi-view Stereo. In *2010 IEEE Conference on Computer Vision and Pattern Recognition* (2010), IEEE, p. to appear. 7

[FFGG*10] Frahm J.-M., Fite-Georgel P., Gallup D., Johnson T., Raguram R., Wu C., Jen Y.-H., Dunn E., Clipp B., Lazebnik S., Pollefeys M.: Building Rome on a Cloudless Day. In *Computer Vision - ECCV 2010* (Berlin, Heidelberg, 2010), Daniilidis K., Maragos P., Paragios N., (Eds.), vol. 6314 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 368–381. 2, 8, 20

[FJZ05] Frueh C., Jain S., Zakhor A.: Data Processing Algorithms for Generating Textured 3D Building Facade Meshes from Laser Scans and Camera Images. *International Journal of Computer Vision 61*, 2 (Feb. 2005), 159–184. 18

[FP07] Furukawa Y., Ponce J.: Accurate, Dense, and Robust Multi-View Stereopsis. In *2007 IEEE Conference on Computer Vision and Pattern Recognition* (June 2007), IEEE, pp. 1–8. 7

[FP08] Furukawa Y., Ponce J.: Carved Visual Hulls for Image-Based Modeling. *International Journal of Computer Vision 81*, 1 (Mar. 2008), 53–67. 7

[FP09] Furukawa Y., Ponce J.: Accurate, Dense, and Robust Multi-View Stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99 (2009), 1–1. 7

[FQ10] Fang T., Quan L.: Resampling structure from motion. In *Computer Vision - ECCV 2010* (Sept. 2010), pp. 1–14. 7

[FZ03] Frueh C., Zakhor A.: Constructing 3D city models by merging aerial and ground views. *IEEE Computer Graphics and Applications 23*, 6 (Nov. 2003), 52–61. 18

[FZ04] Frueh C., Zakhor A.: An Automated Method for Large-Scale, Ground-Based City Model Acquisition. *International Journal of Computer Vision 60*, 1 (Oct. 2004), 5–24. 18

[GCS06] Goesele M., Curless B., Seitz S.: Multi-View Stereo Revisited. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)* (2006), IEEE, pp. 2402–2409. 7

[GFM*07] Gallup D., Frahm J.-M., Mordohai P., Yang Q., Pollefeys M.: Real-Time Plane-Sweeping Stereo with Multiple Sweeping Directions. In *2007 IEEE Conference on Computer Vision and Pattern Recognition* (June 2007), IEEE, pp. 1–8. 7

[GFP10] Gallup D., Frahm J.-M., Pollefeys M.: Piecewise planar and non-planar stereo for urban scene reconstruction.

In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (June 2010), IEEE, pp. 1418–1425. 7

[GG84] GEMAN S., GEMAN D.: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6*, 6 (Nov. 1984), 721–741. 14, 17

[GH97] GUPTA R., HARTLEY R.: Linear pushbroom cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence 19*, 9 (1997), 963–975. 14

[GKCN08] GRÖGER G., KOLBE T. H., CZERWINSKI A., NAGEL C.: OpenGIS city geography markup language (CityGML) encoding standard. *Open Geospatial Consortium Inc. Reference number of this OGC{\textregistered} project document: OGC* (2008). 2

[GKKP07] GRAMMATIKOPOULOS L., KALISPERAKIS I., KARRAS G., PETSA E.: Automatic multi-view texture mapping of 3D surface projections. In *Proceedings of the 2nd ISPRS International Workshop 3D-ARCH 2007: "3D Virtual Reconstruction and Visualization of Complex Architectures"* (ETH Zurich, Switzerland, 2007). 14

[GKVH09] GRZESZCZUK R., KOŠECKÁ J., VEDANTHAM R., HILE H.: Creating compact architectural models by georegistering image collections. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops* (Sept. 2009), IEEE, pp. 1718–1725. 14, 19

[GPF10] GALLUP D., POLLEFEYS M., FRAHM J.-M.: 3D reconstruction using an n-layer heightmap. In *Proceedings of the 32nd DAGM conference on Pattern recognition* (Sept. 2010), pp. 1–10. 7

[GSC*07] GOESELE M., SNAVELY N., CURLESS B., HOPPE H., SEITZ S. M.: Multi-View Stereo for Community Photo Collections. In *2007 IEEE 11th International Conference on Computer Vision* (Oct. 2007), IEEE, pp. 1–8. 7

[GSGA05] GEORGIADIS C., STEFANIDIS A., GYFTAKIS S., AGOURIS P.: Image Orientation for Interactive Tours of Virtually-Modeled Sites. In *Proceedings of the ISPRS Working Group V/4 Workshop 3D-ARCH 2005: "Virtual Reconstruction and Visualization of Complex Architectures"* (Mestre-Venice, Italy, 2005). 14

[GW08] GONZÁLEZ R. C., WOODS R. E.: *Digital image processing*. Prentice Hall, 2008. 15

[HAA97] HORRY Y., ANJYO K.-I., ARAI K.: Tour into the picture. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques - SIGGRAPH '97* (New York, New York, USA, Aug. 1997), ACM Press, pp. 225–232. 10

[Hav05] HAVEMANN S.: *Generative Mesh Modeling*. Phd, Technische Universität Braunschweig, 2005. 13, 17

[HEH05] HOIEM D., EFROS A. A., HEBERT M.: Automatic photo pop-up. *ACM Transactions on Graphics 24*, 3 (July 2005), 577. 10

[HHKF10] HOHMANN B., HAVEMANN S., KRISPEL U., FELLNER D.: A GML shape grammar for semantically enriched 3D building models. *Computers & Graphics 34*, 4 (Aug. 2010), 322–334. 13

[HKHF09] HOHMANN B., KRISPEL U., HAVEMANN S., FELLNER D.: CityFit - High-quality urban reconstructions by fitting shape grammars to images and derived textured point clouds. In *Proceedings of the 3rd ISPRS International Workshop 3D-ARCH 2009: "3D Virtual Reconstruction and Visualization of Complex Architectures"* (Trento, Italy, 25-28 February 2009, 2009). 13, 17

[HLEL06] HAYS J., LEORDEANU M., EFROS A. A., LIU Y.: Discovering Texture Regularity as a Higher-Order Correspondence Problem. In *Computer Vision - ECCV 2006* (2006), Leonardis A., Bischof H., Pinz A., (Eds.), vol. 3952, Springer Berlin Heidelberg, pp. 522–535. 15

[HPFP09] HAUGEARD J.-E., PHILIPP-FOLIGUET S., PRECIOSO F.: Windows and facades retrieval using similarity on graph of contours. In *2009 16th IEEE International Conference on Image Processing (ICIP)* (Nov. 2009), IEEE, pp. 269–272. 16

[HS88] HARRIS C., STEPHENS M.: A Combined Corner and Edge Detector. In *Alvey vision conference* (Manchester, 1988), pp. 147–151. 7, 15

[HTF09] HASTIE T., TIBSHIRANI R., FRIEDMAN J. H.: *The elements of statistical learning: data mining, inference, and prediction*, 2 ed. Springer, 2009. 15

[HYN03] HU J., YOU S., NEUMANN U.: Approaches to large-scale urban modeling. *IEEE Computer Graphics and Applications 23*, 6 (Nov. 2003), 62–69. 18

[HZ04] HARTLEY R., ZISSERMAN A.: *Multiple View Geometry in Computer Vision*. {Cambridge University Press}, Mar. 2004. 2, 5, 6

[HZ05] HAN F., ZHU S.-C.: Bottom-up/Top-Down Image Parsing by Attribute Graph Grammar. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* (2005), vol. 2, IEEE, pp. 1778–1785. 13

[HZ09] HAN F., ZHU S.-C.: Bottom-up/top-down image parsing with attribute grammar. *IEEE Transactions on Pattern Analysis and Machine Intelligence 31*, 1 (Jan. 2009), 59–73. 13

[IZB07] IRSCHARA A., ZACH C., BISCHOF H.: Towards Wiki-based Dense City Modeling. In *2007 IEEE 11th International Conference on Computer Vision* (Oct. 2007), IEEE, pp. 1–8. 2, 7, 18

[IZKB12] IRSCHARA A., ZACH C., KLOPSCHITZ M., BISCHOF H.: Large-scale, dense city reconstruction from user-contributed photos. *Computer Vision and Image Understanding 116*, 1 (Jan. 2012), 2–15. 18

[JRH03] JAYNES C., RISEMAN E., HANSON A. R.: Recognition and reconstruction of buildings from multiple aerial images. *Computer Vision and Image Understanding 90*, 1 (Apr. 2003), 68–98. 19

[JTC09] JIANG N., TAN P., CHEONG L.-F.: Symmetric architecture modeling with a single image. *ACM Transactions on Graphics 28*, 5 (Dec. 2009), 1. 5, 10

[KBH06] KAZHDAN M., BOLITHO M., HOPPE H.: Poisson surface reconstruction. 61–70. 7

[KMO11] KORAH T., MEDASANI S., OWECHKO Y.: Strip Histogram Grid for efficient LIDAR segmentation from urban environments. In *CVPR 2011 WORKSHOPS* (June 2011), IEEE, pp. 74–81. 12

[KN04] KIM Z., NEVATIA R.: Automatic description of complex buildings from multiple images. *Computer Vision and Image Understanding 96*, 1 (Oct. 2004), 60–95. 19

[KP10] KARANTZALOS K., PARAGIOS N.: Large-Scale Building Reconstruction Through Information Fusion and 3-D Priors. *IEEE Transactions on Geoscience and Remote Sensing 48*, 5 (May 2010), 2283–2296. 19

[KR07a] KORAH T., RASMUSSEN C.: 2D Lattice Extraction from Structured Environments. In *2007 IEEE International Conference on Image Processing* (2007), vol. 2, IEEE, pp. II – 61–II – 64. 17

[KR07b] KORAH T., RASMUSSEN C.: Spatiotemporal Inpainting for Recovering Texture Maps of Occluded Building Facades. *IEEE Transactions on Image Processing 16*, 9 (Sept. 2007), 2262–2271. 14

[KST*09] KOUTSOURAKIS P., SIMON L., TEBOUL O., TZIRITAS G., PARAGIOS N.: Single view reconstruction using shape grammars for urban environments. In *2009 IEEE 12th International Conference on Computer Vision* (Sept. 2009), IEEE, pp. 1795–1802. 17

[KZ02] KOŠECKÁ J., ZHANG W.: Video Compass. In *Computer Vision - ECCV 2002* (Berlin, Heidelberg, Apr. 2002), Heyden A., Sparr G., Nielsen M., Johansen P., (Eds.), vol. 2353 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 29–32. 10, 11

[KZ04] KOLMOGOROV V., ZABIH R.: What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence 26*, 2 (Feb. 2004), 147–59. 14, 19

[KZ05] KOŠECKÁ J., ZHANG W.: Extraction, matching, and pose recovery based on dominant rectangular structures. *Computer Vision and Image Understanding 100*, 3 (Dec. 2005), 274–293. 11

[KZN08] KUMAR N., ZHANG L., NAYAR S.: What Is a Good Nearest Neighbors Algorithm for Finding Similar Patches in Images? In *Computer Vision - ECCV 2008* (Berlin, Heidelberg, 2008), Forsyth D., Torr P., Zisserman A., (Eds.), vol. 5303 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 364–378. 6

[KZZL10] KANG Z., ZHANG L., ZLATANOVA S., LI J.: An automatic mosaicking method for building facade texture mapping using a monocular close-range image sequence. *ISPRS Journal of Photogrammetry and Remote Sensing 65*, 3 (May 2010), 282–293. 14

[LA09] LOURAKIS M. I. A., ARGYROS A. A.: SBA: A software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software 36*, 1 (Mar. 2009), 1–30. 6

[LBHL08] LIU Y., BELKINA T., HAYS J., LUBLINERMAN R.: Image de-fencing. In *2008 IEEE Conference on Computer Vision and Pattern Recognition* (June 2008), IEEE, pp. 1–8. 15

[LCOZ*11] LIN J., COHEN-OR D., ZHANG H., LIANG C., SHARF A., DEUSSEN O., CHEN B.: Structure-preserving retargeting of irregular 3D architecture. *ACM Transactions on Graphics 30*, 6 (Dec. 2011), 1. 2

[LCT04] LIU Y., COLLINS R. T., TSIN Y.: A computational model for periodic pattern perception based on frieze and wallpaper groups. *IEEE Transactions on Pattern Analysis and Machine Intelligence 26*, 3 (Mar. 2004), 354–71. 15

[LCZ99] LIEBOWITZ D., CRIMINISI A., ZISSERMAN A.: Creating Architectural Models from Images. *Computer Graphics Forum 18*, 3 (Sept. 1999), 39–50. 9, 10

[LDZPD10] LAFARGE F., DESCOMBES X., ZERUBIA J., PIERROT-DESEILLIGNY M.: Structural approach for building reconstruction from a single DSM. *IEEE Transactions on Pattern Analysis and Machine Intelligence 32*, 1 (Jan. 2010), 135–47. 19

[LE06] LOY G., EKLUNDH J.-O.: Detecting Symmetry and Symmetric Constellations of Features. In *Computer Vision - ECCV 2006* (Berlin, Heidelberg, 2006), Leonardis A., Bischof H., Pinz A., (Eds.), vol. 3952 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 508–521. 15

[LHN00] LEE S. C., HUERTAS A., NEVATIA R.: Modeling 3-D complex buildings with user assistance. In *Proceedings Fifth IEEE Workshop on Applications of Computer Vision* (2000), IEEE Comput. Soc, pp. 170–177. 9

[LHXS05] LIU Y., HAYS J., XU Y.-Q., SHUM H.-Y.: Digital papercutting. In *ACM SIGGRAPH 2005 Sketches on - SIGGRAPH '05* (New York, New York, USA, 2005), ACM Press, p. 99. 15

[LJN02a] LEE S. C., JUNG S. K., NEVATIA R.: Automatic Integration of Facade Textures into 3D Building Models with a Projective Geometry Based Line Clustering. *Computer Graphics Forum 21*, 3 (Sept. 2002), 511–519. 9

[LJN02b] LEE S. C., JUNG S. K., NEVATIA R.: Automatic pose estimation of complex 3D building models. In *Sixth IEEE Workshop on Applications of Computer Vision, 2002. (WACV 2002). Proceedings.* (2002), IEEE Comput. Soc, pp. 148–152. 9

[LJN02c] LEE S. C., JUNG S. K., NEVATIA R.: Integrating ground and aerial views for urban site modeling. In *Object recognition supported by user interaction for service robots* (2002), vol. 4, IEEE Comput. Soc, pp. 107–112. 9

[LL02] LHUILLIER M., LONG Q.: Match propagation for image-based modeling and rendering. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 8 (Aug. 2002), 1140–1146. 7

[LLH04] LIU Y., LIN W.-C., HAYS J.: Near-regular texture analysis and manipulation. *ACM Transactions on Graphics 23*, 3 (Aug. 2004), 368. 15

[LLS04] LEIBE B., LEONARDIS A., SCHIELE B.: Combined Object Categorization and Segmentation With An Implicit Shape Model. *IN ECCV WORKSHOP ON STATISTICAL LEARNING IN COMPUTER VISION* (2004), 17 – 32. 16

[LM11a] LAFARGE F., MALLET C.: Building large urban environments from unstructured point data. In *2011 International Conference on Computer Vision* (Nov. 2011), IEEE, pp. 1068–1075. 20

[LM11b] LARSEN C., MOESLUND T.: 3D Reconstruction of Buildings with Automatic Facade Refinement. In *Advances in Visual Computing* (Berlin, Heidelberg, 2011), Bebis G., Boyle R., Parvin B., Koracin D., Wang S., Kyungnam K., Benes B., Moreland K., Borst C., DiVerdi S., Yi-Jen C., Ming J., (Eds.), vol. 6938 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 451–460. 10

[LMWY09] LIU J., MUSIALSKI P., WONKA P., YE J.: Tensor completion for estimating missing values in visual data. In *2009 IEEE 12th International Conference on Computer Vision* (Sept. 2009), IEEE, pp. 2114–2121. 14, 15

[LMWY12] LIU J., MUSIALSKI P., WONKA P., YE J.: Tensor Completion for Estimating Missing Values in Visual Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence 99*, PP (Jan. 2012). 14

[LN03] LEE S. C., NEVATIA R.: Interactive 3D building modeling using a hierarchical representation. In *First IEEE International Workshop on Higher-Level Knowledge in 3D Modeling and Motion Analysis, 2003. HLK 2003.* (2003), IEEE Comput. Soc, pp. 58–65. 9

[LN04] LEE S. C., NEVATIA R.: Extraction and integration of window in a 3D building model from ground view images. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* (2004), vol. 2, IEEE, pp. 113–120. 9, 15, 17

[Low04] LOWE D. G.: Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision 60*, 2 (Nov. 2004), 91–110. 5, 7, 15

[LPK09] LABATUT P., PONS J.-P., KERIVEN R.: Hierarchical shape-based surface reconstruction for dense multi-view stereo. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops* (Sept. 2009), IEEE, pp. 1598–1605. 7

[LQ05] LHUILLIER M., QUAN L.: A quasi-dense approach to surface reconstruction from uncalibrated images. *IEEE transactions on pattern analysis and machine intelligence 27*, 3 (Mar. 2005), 418–33. 7

[LS07] LIU L., STAMOS I.: A systematic approach for 2D-image to 3D-range registration in urban environments. In *2007 IEEE 11th International Conference on Computer Vision* (Oct. 2007), IEEE, pp. 1–8. 11

[LWZ*08] LI X., WU C., ZACH C., LAZEBNIK S., FRAHM J.-M.: Modeling and Recognition of Landmark Image Collections Using Iconic Scene Graphs. In *Computer Vision - ECCV 2008* (Berlin, Heidelberg, 2008), Forsyth D., Torr P., Zisserman A., (Eds.), vol. 5302 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 427–440. 7

[LZ98] LIEBOWITZ D., ZISSERMAN A.: Metric rectification for perspective images of planes. In *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1998), IEEE Comput. Soc, pp. 482–488. 5, 9, 10

[LZS*11] LI Y., ZHENG Q., SHARF A., COHEN-OR D., CHEN B., MITRA N. J.: 2D-3D fusion for layer decomposition of urban facades. In *2011 International Conference on Computer Vision* (Nov. 2011), IEEE, pp. 882–889. 11

[MAW*07] MERRELL P., AKBARZADEH A., WANG L., MORDOHAI P., FRAHM J.-M., YANG R., NISTER D., POLLEFEYS M.: Real-Time Visibility-Based Fusion of Depth Maps. In *2007 IEEE 11th International Conference on Computer Vision* (Oct. 2007), IEEE, pp. 1–8. 7

[MBB10] MITRA N. J., BRONSTEIN A., BRONSTEIN M.: Intrinsic regularity detection in 3D geometry. In *ECCV 2010* (Sept. 2010), pp. 398–410. 15

[MGP06] MITRA N. J., GUIBAS L. J., PAULY M.: Partial and approximate symmetry detection for 3D geometry. *ACM Transactions on Graphics 25*, 3 (July 2006), 560. 15

[MGP07] MITRA N. J., GUIBAS L. J., PAULY M.: Symmetrization. *ACM Transactions on Graphics 26*, 3 (July 2007), 63. 15

[MK09] MIČUŠÍK B., KOŠECKÁ J.: Piecewise planar city 3D modeling from street view panoramic sequences. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (Miami, FL, June 2009), IEEE, pp. 2906–2912. 7

[MK10] MIČUŠÍK B., KOŠECKÁ J.: Multi-view Superpixel Stereo in Urban Environments. *International Journal of Computer Vision 89*, 1 (Mar. 2010), 106–119. 7

[MKF09] MASTIN A., KEPNER J., FISHER J.: Automatic registration of LIDAR and optical images of urban scenes. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (Miami, FL, June 2009), IEEE, pp. 2639–2646. 19

[MMWvG11] MATHIAS M., MARTINOVIC A., WEISSENBERG J., VAN GOOL L.: Procedural 3D Building Reconstruction Using Shape Grammars and Detectors. In *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission* (May 2011), IEEE, pp. 304–311. 13

[MP08] MCCANN J., POLLARD N. S.: Real-time gradient-domain painting. *ACM Transactions on Graphics 27*, 3 (Aug. 2008), 1. 14

[MR05] MAYER H., REZNIK S.: Building Façade Interpretation from Image Sequences. In *Proceedings of the ISPRS Workshop CMRT 2005* (Vienna, 2005), vol. XXXVI, pp. 55–60. 16

[MR06] MAYER H., REZNIK S.: MCMC Linked with Implicit Shape Models and Plane Sweeping for 3D Building Facade Interpretation in Image Sequences. In *PCV '06, Photogrammetric Computer Vision* (2006), ISPRS Comm. III Symposium, IAPRS, pp. 130–135. 16

[MR07] MAYER H., REZNIK S.: Building facade interpretation from uncalibrated wide-baseline image sequences. *ISPRS Journal of Photogrammetry and Remote Sensing 61*, 6 (Feb. 2007), 371–380. 16

[MRM*10] MUSIALSKI P., RECHEIS M., MAIERHOFER S., WONKA P., PURGATHOFER W.: Tiling of Ortho-Rectified Façade Images. In *Proceedings of the 26th Spring Conference on Computer Graphics - SCCG '10* (New York, New York, USA, May 2010), ACM Press, p. 117. 16

[MvGV09] MOONS T., VAN GOOL L., VERGAUWEN M.: 3D Reconstruction from Multiple Images Part 1: Principles. *Foundations and Trends in Computer Graphics and Vision 4*, 4 (2009), 287–404. 2, 5

[MWR*09] MUSIALSKI P., WONKA P., RECHEIS M., MAIERHOFER S., PURGATHOFER W.: Symmetry-Based Façade Repair. In *Vision, Modeling, and Visualization (VMV)* (2009), Magnor M. A., Rosenhahn B., Theisel H., (Eds.), DNB, pp. 3–10. 14, 15

[MWW12] MUSIALSKI P., WIMMER M., WONKA P.: Interactive Coherence-Based Façade Modeling. *Computer Graphics Forum (Proceedings of EUROGRAPHICS 2012) 31*, 2 (May 2012), to appear. 16, 17

[MZWvG07] MÜLLER P., ZENG G., WONKA P., VAN GOOL L.: Image-based procedural modeling of facades. *ACM Transactions on Graphics 26*, 3 (July 2007), 85. 16

[Nis04] NISTÉR D.: An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence 26*, 6 (June 2004), 756–77. 5

[NN01] NORONHA S., NEVATIA R.: Detection and modeling of buildings from multiple aerial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*, 5 (May 2001), 501–518. 19

[NP02] NEVATIA R., PRICE K.: Automatic and interactive modeling of buildings in urban environments from aerial images. In *Proceedings. International Conference on Image Processing* (2002), vol. 1, IEEE, pp. 525–528. 19

[NS06] NISTER D., STEWENIUS H.: Scalable Recognition with a Vocabulary Tree. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)* (2006), IEEE, pp. 2161–2168. 6, 18

[NSZ*10] NAN L., SHARF A., ZHANG H., COHEN-OR D., CHEN B.: SmartBoxes for interactive urban reconstruction. *ACM Transactions on Graphics 29*, 4 (July 2010), 1. 11, 17

[NTC11] NIANJUAN JIANG, TAN P., CHEONG L.-F.: Multi-view repetitive structure detection. In *2011 International Conference on Computer Vision* (Nov. 2011), IEEE, pp. 535–542. 16

[OCDD01] OH B. M., CHEN M., DORSEY J., DURAND F.: Image-based modeling and photo editing. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques - SIGGRAPH '01* (New York, New York, USA, 2001), ACM Press, pp. 433–442. 10

[OR05] ORTIN D., REMONDINO F.: Generation of Occlusion-free Images for Realistic Texture Mapping. In *Proceedings of the ISPRS Working Group V/4 Workshop 3D-ARCH 2005: "Virtual Reconstruction and Visualization of Complex Architectures"* (Mestre-Venice, Italy, 2005). 14

[PBCL10] PARK M., BROCKLEHURST K., COLLINS R. T., LIU Y.: Translation-symmetry-based perceptual grouping with applications to urban scenes. In *Computer Vision - ACCV 2010* (Berlin, Heidelberg, Nov. 2010), Springer, pp. 329–342. 16

[PGB03] PÉREZ P., GANGNET M., BLAKE A.: Poisson image editing. *ACM Transactions on Graphics 22*, 3 (July 2003), 313. 14

[PMW*08] Pauly M., Mitra N. J., Wallner J., Pottmann H., Guibas L. J.: Discovering structural regularity in 3D geometry. *ACM Transactions on Graphics 27*, 3 (Aug. 2008), 1. 15, 16

[PNF*08] Pollefeys M., Nistér D., Frahm J.-M., Akbarzadeh A., Mordohai P., Clipp B., Engels C., Gallup D., Kim S.-J., Merrell P., Salmi C., Sinha S., Talton B., Wang L., Yang Q., Stewénius H., Yang R., Welch G., Towles H.: Detailed Real-Time Urban 3D Reconstruction from Video. *International Journal of Computer Vision 78*, 2-3 (Oct. 2008), 143–167. 7, 18

[PSG*06] Podolak J., Shilane P., Golovinskiy A., Rusinkiewicz S., Funkhouser T.: A planar-reflective symmetry transform for 3D shapes. *ACM Transactions on Graphics 25*, 3 (July 2006), 549. 15

[PSK06] Pavić D., Schönefeld V., Kobbelt L.: Interactive image completion with perspective correction. *The Visual Computer 22*, 9-11 (Aug. 2006), 671–681. 14

[PV09a] Pu S., Vosselman G.: Building Facade Reconstruction by Fusing Terrestrial Laser Points and Images. *Sensors 9*, 6 (June 2009), 4525–4542. 17

[PV09b] Pu S., Vosselman G.: Knowledge based reconstruction of building models from terrestrial laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing 64*, 6 (Nov. 2009), 575–584. 12, 17

[PV09c] Pu S., Vosselman G.: Refining building facade models with images. In *ISPRS Workshop, CMRT09 - City Models, Roads and Traffic* (2009), vol. XXXVIII, pp. 3–4. 17

[PvGV*04] Pollefeys M., van Gool L., Vergauwen M., Verbiest F., Cornelis K., Tops J., Koch R.: Visual Modeling with a Hand-Held Camera. *International Journal of Computer Vision 59*, 3 (Sept. 2004), 207–232. 7, 18

[PY09a] Poullis C., You S.: Automatic Creation of Massive Virtual Cities. In *2009 IEEE Virtual Reality Conference* (Mar. 2009), IEEE, pp. 199–202. 20

[PY09b] Poullis C., You S.: Automatic reconstruction of cities from remote sensor data. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (June 2009), IEEE, pp. 2775–2782. 20

[PY09c] Poullis C., You S.: Photorealistic large-scale urban city model reconstruction. *IEEE Transactions on Visualization and Computer Graphics 15*, 4 (2009), 654–69. 20

[PY11] Poullis C., You S.: 3D Reconstruction of Urban Areas. In *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission* (May 2011), IEEE, pp. 33–40. 20

[RB07] Ripperda N., Brenner C.: Data driven rule proposal for grammar based façade reconstruction. In *PIA07. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* (2007). 13, 17

[RB09] Ripperda N., Brenner C.: Application of a Formal Grammar to Facade Reconstruction in Semiautomatic and Automatic Environments. In *Proceedings of 12th AGILE Conference on GIScience* (Hannover, Germany, 2009). 13, 17

[RC02] Rother C., Carlsson S.: Linear Multi View Reconstruction and Camera Recovery Using a Reference Plane. *International Journal of Computer Vision 49*, 2 (2002), 117–141–141. 9, 10, 11

[RFP08] Raguram R., Frahm J.-M., Pollefeys M.: A Comparative Analysis of RANSAC Techniques Leading to Adaptive Real-Time Random Sample Consensus. In *Computer Vision - ECCV 2008* (Berlin, Heidelberg, 2008), Forsyth D., Torr P., Zisserman A., (Eds.), vol. 5303 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 500–513. 6

[RGL04] Roman A., Garg G., Levoy M.: Interactive design of multi-perspective images for visualizing urban landscapes. *IEEE Visualization 2004* (2004), 537–544. 14

[Rip08] Ripperda N.: Determination of Facade Attributes for Facade Reconstruction. In *ISPRS Congress Beijing 2008, Proceedings of Commission III* (2008), pp. 285–290. 13, 17

[RM07] Reznik S., Mayer H.: Implicit shape models, model selection, and plane sweeping for 3d facade interpretation. In *PIA07. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* (2007), vol. 36, pp. 173–178. 16

[Rot00] Rother C.: A new approach for vanishing point detection in architectural environments. *IN PROC. 11TH BRITISH MACHINE VISION CONFERENCE* (2000), 382 – 391. 10

[RWY95] Reisfeld D., Wolfson H., Yeshurun Y.: Context-free attentional operators: The generalized symmetry transform. *International Journal of Computer Vision 14*, 2 (Mar. 1995), 119–130. 15

[SA00a] Stamos I., Allen P.: Integration of range and image sensing for photo-realistic 3D modeling. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)* (2000), vol. 2, IEEE, pp. 1435–1440. 11

[SA00b] Stamos I., Allen P. K.: 3-D model construction using range and image data. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000* (2000), IEEE Comput. Soc, pp. 531–536. 11

[SA01] Stamos I., Allen P. K.: Automatic registration of 2-D with 3-D imagery in urban environments. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001* (2001), IEEE Comput. Soc, pp. 731–736. 11

[SA02] Stamos I., Allen P. K.: Geometry and Texture Recovery of Scenes of Large Scale. *Computer Vision and Image Understanding 88*, 2 (Nov. 2002), 94–118. 11

[SB03] Schindler K., Bauer J.: A model-based method for building reconstruction. In *First IEEE International Workshop on Higher-Level Knowledge in 3D Modeling and Motion Analysis, 2003. HLK 2003.* (2003), IEEE Comput. Soc, pp. 74–82. 16

[SBM*10] St'ava O., Beneš B., Mech R., Aliaga D. G., Krištof P.: Inverse Procedural Modeling by Automatic Generation of L-systems. *Computer Graphics Forum 29*, 2 (2010). 12

[SCD*06] Seitz S., Curless B., Diebel J., Scharstein D., Szeliski R.: A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR'06)* (2006), vol. 1, IEEE, pp. 519–528. 6

[SGSS08] Snavely N., Garg R., Seitz S. M., Szeliski R.: Finding paths through the world's photos. *ACM Transactions on Graphics 27*, 3 (Aug. 2008), 1. 7

[SHB08] Sonka M., Hlavac V., Boyle R.: *Image processing, analysis, and machine vision*. Thompson Learning, 2008. 15

[SHFH11] Shen C.-H., Huang S.-S., Fu H., Hu S.-M.: Adaptive partitioning of urban facades. *ACM Transactions on Graphics 30*, 6 (Dec. 2011), 1. 12, 16

[SHS98] Shum H.-Y., Han M., Szeliski R.: Interactive construction of 3D models from panoramic mosaics. In *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1998), IEEE, pp. 427–433. 10

[SJW*11] Sunkel M., Jansen S., Wand M., Eisemann E., Seidel H.-P.: Learning Line Features in 3D Geometry. *Computer Graphics Forum 30*, 2 (Apr. 2011), 267–276. 17

[SK03] Seitz S. M., Kim J.: Multiperspective imaging. *IEEE Computer Graphics and Applications 23*, 6 (Nov. 2003), 16–19. 14

[SKD06] Schindler G., Krishnamurthy P., Dellaert F.: Line-Based Structure from Motion for Urban Environments. In *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)* (June 2006), IEEE, pp. 846–853. 6

[SS99] Schapire R. E., Singer Y.: Improved Boosting Algorithms Using Confidence-rated Predictions. *Machine Learning 37*, 3 (1999), 297. 16

[SS02a] Scharstein D., Szeliski R.: A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision 47*, 1 (2002), 7–42–42. 6, 7

[SS02b] Shum H.-Y., Szeliski R.: Construction of Panoramic Image Mosaics with Global and Local Alignment. *International Journal of Computer Vision 48*, 2 (July 2002), 151–152. 14

[SSG*10] Snavely N., Simon I., Goesele M., Szeliski R., Seitz S. M.: Scene Reconstruction and Visualization From Community Photo Collections. *Proceedings of the IEEE 98*, 8 (Aug. 2010), 1370–1390. 6, 7

[SSS06] Snavely N., Seitz S. M., Szeliski R.: Photo tourism: Exploring Photo Collections in 3D. *ACM Transactions on Graphics 25*, 3 (July 2006), 835. 2, 7

[SSS07] Snavely N., Seitz S. M., Szeliski R.: Modeling the World from Internet Photo Collections. *International Journal of Computer Vision 80*, 2 (Dec. 2007), 189–210. 7

[SSS*08] Sinha S. N., Steedly D., Szeliski R., Agrawala M., Pollefeys M.: Interactive 3D architectural modeling from unordered photo collections. *ACM Transactions on Graphics 27*, 5 (Dec. 2008), 1. 9, 14

[SSS09] Sinha S. N., Steedly D., Szeliski R.: Piecewise planar stereo for image-based rendering. In *2009 IEEE 12th International Conference on Computer Vision* (Sept. 2009), IEEE, pp. 1881–1888. 7

[Str05] Strang G.: *Linear Algebra and Its Applications*. Brooks Cole, 2005. 14

[SWK07] Schnabel R., Wahl R., Klein R.: Efficient RANSAC for Point-Cloud Shape Detection. *Computer Graphics Forum 26*, 2 (June 2007), 214–226. 11

[SZ99] Schaffalitzky F., Zisserman A.: Geometric grouping of repeated elements within images. pp. 13–22. 15

[Sze06] Szeliski R.: Image Alignment and Stitching: A Tutorial. *Foundations and Trends in Computer Graphics and Vision 2*, 1 (2006), 1–104. 14

[Sze11] Szeliski R.: *Computer Vision*. Texts in Computer Science. Springer London, London, Nov. 2011. 2

[SZG*09] Schweiger F., Zeisl B., Georgel P. F., Schroth G., Steinbach E., Navab N.: Maximum Detector Response Markers for SIFT and SURF. In *Int. Workshop on Vision, Modeling and Visualization (VMV)* (2009). 6

[TCLH06] Tsai F., Chen C.-H., Liu J.-K., Hsiao K.-H.: Texture Generation and Mapping Using Video Sequences for 3D Building Models. In *Innovations in 3D Geo Information Systems* (Berlin, Heidelberg, 2006), Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg, pp. 429–438. 15

[Tel98] Teller S.: Automated Urban Model Acquisition: Project Rationale and Status. In *Image Understanding Workshop* (1998), pp. 445–462. 10

[TK95] Taylor C. J., Kriegman D. J.: Structure and motion from line segments in multiple images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 17*, 11 (1995), 1021–1032. 6, 9

[TKO08] Tan Y. K. A., Kwoh L. K., Ong S. H.: Large scale texture mapping of building facades. In *ISPRS Congress Beijing 2008, Proceedings of Commission V* (2008). 14

[TKS*11] Teboul O., Kokkinos I., Simon L., Koutsourakis P., Paragios N.: Shape grammar parsing via Reinforcement Learning. In *CVPR 2011* (June 2011), IEEE, pp. 2273–2280. 17

[TL07] Tsai F., Lin H.: Polygon-based texture mapping for cyber city 3D building models. *International Journal of Geographical Information Science 21*, 9 (Oct. 2007), 965–981. 14

[TLH06] Tsai F., Liu J.-K., Hsiao K.-H.: Morphological Processing of Video for 3D Building Model Visualization. In *Proceedings of 27 th Asian Conference on Remote Sensing (ACRS2006), Ulaanbaatar, Mongolia* (2006), pp. 1–6. 15

[TLL*11] Talton J. O., Lou Y., Lesser S., Duke J., Měch R., Koltun V.: Metropolis procedural modeling. *ACM Transactions on Graphics 30*, 2 (Apr. 2011), 1–14. 12

[TLLH05] Tsai F., Lin H.-C., Liu J.-K., Hsiao K.-H.: Semiautomatic texture generation and transformation for cyber city building models. In *Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International* (2005), pp. 4980–4983. 15

[TMHF99] Triggs B., McLauchlan P. F., Hartley R. I., Fitzgibbon A. W.: Bundle Adjustment - A Modern Synthesis. *Lecture Notes In Computer Science; Vol. 1883* (1999), 298. 6

[TMT10] Toshev A., Mordohai P., Taskar B.: Detecting and parsing architecture at city scale from range data. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (June 2010), IEEE, pp. 398–405. 12, 13

[TS10] Tyleček R., Šára R.: A weak structure model for regular pattern recognition applied to facade images. In *Computer Vision - ACCV 2010* (Nov. 2010), Springer-Verlag, pp. 450–463. 17

[TSKP10] Teboul O., Simon L., Koutsourakis P., Paragios N.: Segmentation of building facades using procedural shape priors. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (June 2010), IEEE, pp. 3105–3112. 17

[TTMvG01] Turina A., Tuytelaars T., Moons T., van Gool L.: Grouping via the Matching of Repeated Patterns. Singh S., Murshed N., Kropatsch W., (Eds.), Lecture Notes in Computer Science, Springer, pp. 250–259. 15

[TTvG01] Turina A., Tuytelaars T., van Gool L.: Efficient grouping under perspective skew. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001* (2001), vol. 1, IEEE Comput. Soc, pp. I–247–I–254. 15

[TvG11] Tingdahl D., van Gool L.: A Public System for Image Based 3D Model Generation. *Computer Vision/Computer Graphics Collaboration Techniques 6930* (2011), 262–273. 7

[VAB10] Vanegas C. A., Aliaga D. G., Beneš B.: Building Reconstruction using Manhattan-World Grammars. *2010 IEEE Conference on Computer Vision and Pattern Recognition* (2010), to appear. 13

[VAB12] VANEGAS C. A., ALIAGA D. G., BENES B.: Automatic Extraction of Manhattan-World Building Masses from 3D Laser Range Scans. *IEEE transactions on visualization and computer graphics*, 99 (Jan. 2012), 1–1. 12

[VAW*10] VANEGAS C. A., ALIAGA D. G., WONKA P., MÜLLER P., WADDELL P. A., WATSON B.: Modelling the Appearance and Behaviour of Urban Spaces. *Computer Graphics Forum 29*, 1 (Mar. 2010), 25–42. 2, 12

[vdH01] VAN DEN HEUVEL F. A.: Object Reconstruction from a Single Architectural Image Taken with an Uncalibrated Camera. *Photogrammetrie Fernerkundung Geoinformation 4* (2001), 247 – 260. 10

[vdHDT*06] VAN DEN HENGEL A., DICK A., THORMÄHLEN T., WARD B., TORR P. H. S.: Building models of regular scenes from structure and motion. *British Machine Vision Conference 2006* (2006). 9

[vdHDT*07a] VAN DEN HENGEL A., DICK A., THORMÄHLEN T., WARD B., TORR P. H. S.: A shape hierarchy for 3D modelling from video. *Computer graphics and interactive techniques in Australasia and South East Asia* (2007), 63. 9

[vdHDT*07b] VAN DEN HENGEL A., DICK A., THORMÄHLEN T., WARD B., TORR P. H. S.: Interactive 3D Model Completion. In *9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007)* (Dec. 2007), IEEE, pp. 175–181. 9

[vdHDT*07c] VAN DEN HENGEL A., DICK A., THORMÄHLEN T., WARD B., TORR P. H. S.: VideoTrace: rapid interactive scene modelling from video. *ACM Transactions on Graphics 26*, 3 (July 2007), 86. 14

[vGZ97] VAN GOOL L., ZISSERMAN A.: Automatic 3D model building from video sequences. *European Transactions on Telecommunications 8*, 4 (July 1997), 369–378. 6

[vGZBM07] VAN GOOL L., ZENG G., BORRE F. V. D., MÜLLER P.: Towards Mass-produced Building Models. In *PIA07. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* (2007), Institute of Photogrammetry and Cartography, Technische Universitaet Muenchen, pp. 209–220. 17

[VvG06] VERGAUWEN M., VAN GOOL L.: Web-based 3D Reconstruction Service. *Machine Vision and Applications 17*, 6 (2006), 411. 7

[WACS11] WU C., AGARWAL S., CURLESS B., SEITZ S. M.: Multicore bundle adjustment. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (2011), IEEE, pp. 3057–3064. 6, 16

[WDF08] WENZEL S., DRAUSCHKE M., FÖRSTNER W.: Detection of repeated structures in facade images. *Pattern Recognition and Image Analysis 18*, 3 (Sept. 2008), 406–411. 16

[WFP10] WU C., FRAHM J.-M., POLLEFEYS M.: Detecting Large Repetitive Structures with Salient Boundaries. In *Computer Vision - ECCV 2010 - Lecture Notes in Computer Science* (2010), Daniilidis K., Maragos P., Paragios N., (Eds.), vol. 6312, Springer Berlin / Heidelberg, pp. 142–155–155. 16

[WH01] WANG X., HANSON A. R.: Surface Texture and Microstructure Extraction from Multiple Aerial Images. *Computer Vision and Image Understanding 83*, 1 (July 2001), 1–37. 14, 15

[WSB05] WILCZKOWIAK M., STURM P., BOYER E.: Using geometric constraints through parallelepipeds for calibration and 3D modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27*, 2 (Feb. 2005), 194–207. 5

[WTT*02] WANG X., TOTARO S., TAILL F., HANSON A. R., TELLER S.: Recovering facade texture and microstructure from real-world images. In *Photogrammetric Computer Vision, ISPRS Commission III, Symposium 2002* (Graz, Austria, 2002), no. September, pp. A381–368. 14, 15

[WYN07] WANG L., YOU S., NEUMANN U.: Semiautomatic registration between ground-level panoramas and an orthorectified aerial image for building modeling. In *2007 IEEE 11th International Conference on Computer Vision* (2007), IEEE, pp. 1–8. 19

[WZ02] WERNER T., ZISSERMAN A.: New Techniques for Automated Architectural Reconstruction from Photographs. *Lecture Notes In Computer Science; Vol. 2351* (2002). 10

[XFT*08] XIAO J., FANG T., TAN P., ZHAO P., OFEK E., QUAN L.: Image-based façade modeling. *ACM Transactions on Graphics 27*, 5 (Dec. 2008), 1. 14, 17, 19

[XFZ*09] XIAO J., FANG T., ZHAO P., LHUILLIER M., QUAN L.: Image-based street-side city modeling. *ACM Transactions on Graphics 28*, 5 (Dec. 2009), 1. 10, 19

[YP03] YANG R., POLLEFEYS M.: Multi-resolution real-time stereo on commodity graphics hardware. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.* (2003), IEEE Comput. Soc, pp. I–211–I–217. 7

[YWR09] YIN X., WONKA P., RAZDAN A.: Generating 3D Building Models from Architectural Drawings: A Survey. *IEEE Computer Graphics and Applications 29*, 1 (Jan. 2009), 20–30. 2

[ZBKB08] ZEBEDIN L., BAUER J., KARNER K., BISCHOF H.: Fusion of Feature- and Area-Based Information for Urban Buildings Modeling from Aerial Imagery. In *Computer Vision - ECCV 2008* (2008), Springer Berlin Heidelberg, pp. 873–886. 19

[ZFPW03] ZOMET A., FELDMAN D., PELEG S., WEINSHALL D.: Mosaicing new views: the crossed-slits projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence 25*, 6 (June 2003), 741–754. 14

[Zhe03] ZHENG J. Y.: Digital route panoramas. *IEEE Multimedia 10*, 3 (July 2003), 57–67. 14

[ZLPW06] ZOMET A., LEVIN A., PELEG S., WEISS Y.: Seamless image stitching by minimizing false edges. *IEEE Transactions on Image Processing 15*, 4 (Apr. 2006), 969–977. 14

[ZN08] ZHOU Q.-Y., NEUMANN U.: Fast and extensible building modeling from airborne LiDAR data. *Geographic Information Systems* (2008). 12

[ZN09] ZHOU Q.-Y., NEUMANN U.: A streaming framework for seamless building reconstruction from large-scale aerial LiDAR data. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (June 2009), IEEE, pp. 2759–2766. 20

[ZN10] ZHOU Q.-Y., NEUMANN U.: 2.5D dual contouring: a robust approach to creating building models from Aerial LiDAR point clouds. In *ECCV 2010* (Sept. 2010), pp. 115–128. 11, 12

[ZN11] ZHOU Q.-Y., NEUMANN U.: 2.5D building modeling with topology control. In *CVPR 2011* (June 2011), IEEE, pp. 2489–2496. 12, 20

[ZPB07] ZACH C., POCK T., BISCHOF H.: A Globally Optimal Algorithm for Robust TV-L1 Range Image Integration. In *2007 IEEE 11th International Conference on Computer Vision* (Oct. 2007), IEEE, pp. 1–8. 7

[ZQ11] ZHAO P., QUAN L.: Translation symmetry detection in a fronto-parallel view. In *CVPR 2011* (June 2011), IEEE, pp. 1009–1016. 16

[ZSW*10] ZHENG Q., SHARF A., WAN G., LI Y., MITRA N. J., COHEN-OR D., CHEN B.: Non-local scan consolidation for 3D urban scenes. *ACM Transactions on Graphics 29*, 4 (July 2010), 1. 11