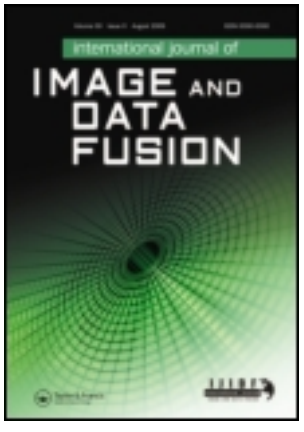

International Journal of Image and Data Fusion

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tidf20

3D building modeling using images and LiDAR: a review

Ruisheng Wang [a]

[a] Department of Geomatics Engineering , University of Calgary , Calgary , Canada

Published online: 05 Jul 2013.

To cite this article: Ruisheng Wang (2013) 3D building modeling using images and LiDAR: a review, International Journal of Image and Data Fusion, 4:4, 273-292, DOI: 10.1080/19479832.2013.811124

To link to this article: http://dx.doi.org/10.1080/19479832.2013.811124

International Journal of Image and Data Fusion, 2013
Vol. 4, No. 4, 273–292, http://dx.doi.org/10.1080/19479832.2013.811124

3D building modeling using images and LiDAR: a review

Ruisheng Wang*

Department of Geomatics Engineering, University of Calgary, Calgary, Canada

(*Received 21 February 2013; final version received 27 May 2013*)

3D modeling from images and LiDAR (Light Detection And Ranging) has been an active research area in the photogrammetry, computer vision, and computer graphics communities. In terms of literature review, a comprehensive survey on 3D building modeling that contains methods from all these fields will be beneficial. This article attempts to survey the state-of-the-art 3D building modeling methods in the areas of photogrammetry, computer vision, and computer graphics. The existing methods are grouped into three categories: 3D reconstruction from images, 3D modeling using range data, and 3D modeling using images and range data. The use of both data for 3D modeling is a sensor fusion approach, in which methods of image-to-LiDAR registration, upsampling, and image-guided segmentation are reviewed. For each category, the key problems are identified and solutions are addressed.

Keywords: 3D reconstruction; LiDAR; building modeling; registration; sensor fusion

1. Introduction

The interest in generating 3D models is motivated by a wide range of applications, such as video games, virtual, augmented, and mixed reality, 3D GPS navigation, solar potential analysis, and 3D Geographic Information Systems. The interest in 3D building models is also indicated by a survey from the European Organization for Experimental Photogrammetric Research, showing that 95% of participants were most interested in 3D building data within city models (Fuchs *et al.* 1998). In recent years, the demand for 3D photorealistic building models has dramatically increased (Wang *et al.* 2011). These needs arise particularly from 3D GPS navigation systems and online services such as Google Earth and Nokia maps. Currently, the creation of 3D photorealistic building models still lacks automation and is a labor-intensive process (Van Gool *et al*. 2007). Automatic solutions could significantly increase productivity, reduce costs, and be of enormous interest to location-based service providers such as Nokia and Google.

In the photogrammetry, computer vision, and computer graphics communities, much effort has been spent on the automatic creation of 3D models from images and LiDAR (Light Detection And Ranging). Photogrammetry is described in a computer vision textbook (Trucco and Verri 1998) as a noncontact imaging technique that obtains reliable and accurate object measurements. The emphasis on accuracy is the main characteristic of photogrammetry in comparison with computer vision. Computer vision concerns inferring properties (e.g., geometric and dynamic) of the 3D world from one or more digital images (Trucco and Verri 1998), and its emphasis is placed on automation. Computer graphics also concerns obtaining geometry from images, but the purpose is often for high-quality 3D rendering; this problem is also known as image-based modeling (Snavely 2008).

*Email: ruiswang@ucalgary.ca

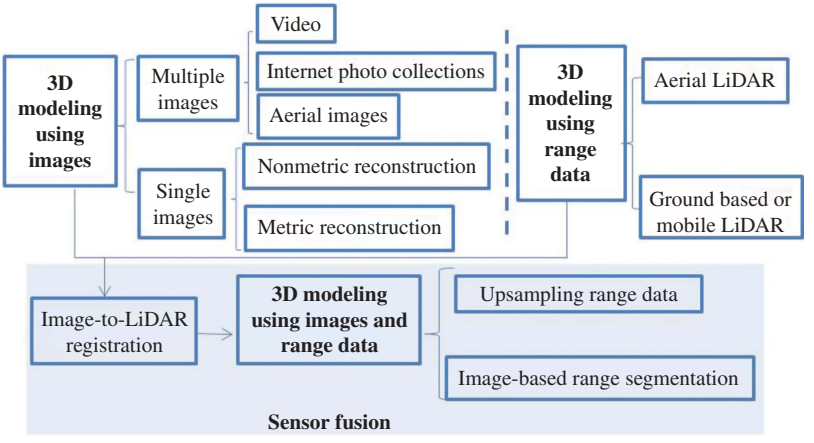

Figure 1. Criteria for classification of 3D building modeling methods.

A comprehensive review of the work in urban modeling and rendering from the computer graphics field has been presented by Vanegas *et al.* (2010). This paper covers the relevant methods from these three related subjects.

3D building models can be generated by various techniques such as airborne imaging using active sensors (e.g., LiDAR) and/or passive sensors (e.g., aerial camera), and ground-based or vehicle-borne sensing techniques (i.e., mobile mapping). Figure 1 shows the criteria for classification of 3D modeling methods. Image-based solutions have been widely researched. The types of input images for 3D reconstruction include single images (Criminisi *et al.* 2000, Hoiem *et al.* 2005, Wang and Ferrie 2008, Saxena *et al.* 2009), multiple images/videos (Fitzgibbon and Zisserman 1998, Pollefeys *et al.* 2008), including aerial images, and unorganized photo collections such as photos collected from the Internet (Snavely *et al.* 2006, Agarwa *et al.* 2009). 3D modeling using range data is becoming more and more popular nowadays. The research has moved from well-controlled laboratory environments (Curless and Levoy 1996, Hoover *et al.* 1996) to solving real-world problems using data collected from outdoor scenes. The existing methods in large-scale urban modeling include airborne LiDAR-based rooftop modeling (Zhou and Neumann 2008, Poullis and You 2009) and mobile LiDAR modeling of building facades (Zhao and Shibasaki 2001, Frueh and Zakhor 2003, Hohmann *et al.* 2009). In the photogrammetry and remote-sensing community, 3D building modeling using airborne or terrestrial LiDAR data has been an intensive and longstanding research problem (Vosselman and Suveg 2001, Rottensteiner 2003). Generally, the goal here is to generate polygonal mesh models to represent the real scene with much less data in comparison with original LiDAR. 3D modeling using the combination of image and range data is a sensor fusion approach that takes strength from each in order to overcome their limitations. Images normally have higher resolution and more visual information than range data, but require solution of the inverse optics problem (Palmer 1999, Zygmunt 2001) to derive the corresponding 3D information. Range data are noisy, sparse, and have less visual information, but already contain 3D information. The first step in sensor fusion is to solve the image-to-range registration problem, which is to align 2D images with the 3D range data, consisting of estimating the relative camera pose with respect to the range sensor. The registration result is critical not only for texture mapping 3D models but also for many vision applications such as image-based upsampling of range data and image-

guided range segmentation (Dias *et al*. 2003, Torres-Mendez and Dudek 2004, Diebel and Thrun 2005, Yang *et al*. 2007) that can eventually be used for 3D modeling tasks. In the following sections, the existing methods are categorized in terms of multiple or single images, aerial or mobile LiDAR, or combination of images and LiDAR.

2. Three-dimensional reconstruction from images

2.1 *Multiple images*

To reconstruct 3D geometry from images, camera poses have to be known. The earliest work in camera pose estimation is from photogrammetry. The so-called "space resection" or "spatial resection" in photogrammetry is to estimate a camera pose by measuring at least three feature points evenly distributed across an image (mainly aerial images), whose 3D coordinates are known, and normally georeferenced. The recovery of relative poses of two cameras is called "relative orientation" in photogrammetry. Finsterwalder (1937) stated that the relative orientation of two images can be established by using five pairs of homologous image points. Kruppa (1913) studied the possible solutions for the relative orientation problem. Based on these results, several five-point algorithms for estimating two-view geometry have been recently developed (Nister 2004a, Li and Hartley 2006). Since most stereo papers focus on pixel-wise correspondence from calibrated views in which the camera poses are known, the literature on stereo or multiple-view stereo is not included here. Excellent surveys can be found in Scharstein and Szeliski (2002) and Seitz *et al*. (2006).

For a large number of images, in certain scenarios such as a frame of video, the methods (Fitzgibbon and Zisserman 1998, Nister 2004b, Pollefeys *et al*. 2004) have automatically generated impressive 3D models. These methods start with a sequence of images taken by an uncalibrated camera under very short baselines (Remondino and El-Hakim 2006). The interest points are automatically extracted, tracked, or matched across views. The relations between multiple views are computed by using a bundle method (Triggs *et al*. 2000), and the 3D models are then automatically created by dense depth map generation (Scharstein and Szeliski 2002, Seitz *et al*. 2006). These approaches normally require good features in multiple images and very short baselines between consecutive images. In cases of occlusions, illumination changes, and lack of textures, the algorithms often fail. Bundle methods require an appropriate initialization, which can be difficult to obtain. In Debevec *et al.* (1996), the initial values for a nonlinear optimization are computed through scene constraints.

In contrast to a dense 3D reconstruction, sparse scene geometry can be recovered automatically from large, unorganized photo collections (Snavely *et al.* 2006, Agarwa *et al*. 2009). The core of the method is a robust "structure from motion" algorithm. In Snavely *et al.* (2006), the feature points in each image are first extracted by using a SIFT keypoint detector (Lowe 2004), and then the fundamental matrix for each pair image is robustly estimated using the eight-point algorithm (Hartley and Zisserman 2004) and RANSAC (Fischler and Bolles 1981). A set of camera parameters and 3D locations for matched points are incrementally estimated with algorithms such as Levenberg–Marquardt (Nocedal and Wright 2006). The recovered camera parameters and 3D points are then used for better rendering, transitioning, and navigating photographs. Based on this research, Microsoft has created a streaming multiresolution Web-based service called Photosynth (http://photosynth.net), and the open-source software "Bundler" (http://

phototour.cs.washington.edu/bundler/) is available for structure-from-motion (SfM) using unordered image collections.

Recent efforts have also tried to generate dense reconstructions from Internet photo collections (Goesele *et al*. 2007, Furukawa *et al*. 2010). Goesele *et al.* (2007) proposed the first multiview stereo method applied to Internet photo collections. To handle variations in the images, they developed an adaptive view selection procedure to automatically identify image subsets that are similar in appearance and scale. Furukawa *et al.* (2010) introduced an approach for enabling existing multiview stereo methods (Seitz *et al*. 2006) to operate on large unstructured photo collections. They formulated the overlapping clustering problem as a constrained optimization and developed a new merging method that robustly eliminates low-quality or erroneous points. The authors claimed that it is the first to demonstrate an unstructured multiview stereo approach at a city scale.

A representative approach in automatic generation of 3D photorealistic models from ground-level images captured along the streets has been presented in Xiao *et al*. (2009). This method produced visually compelling results with a strong assumption of building regularity, the Manhattan-world assumption. Limitations include limited camera field of view because the images were captured from a ground-based camera. The upper parts of large buildings may not be modeled. Instead of working on perspective images, the method (Micusik and Kosecka 2009) worked directly on the street view panoramic image sequences based on a piecewise planar structure assumption.

Interactive 3D reconstruction methods (Debevec *et al.* 1996, Cipolla *et al.* 1999, van den Hengel *et al*. 2007, Xiao *et al*. 2008) reconstruct 3D models with user intervention. Facade (Debevec *et al.* 1996) was one of the most successful image-based modeling systems (Cipolla *et al.* 1999). The core algorithm of the modeling part is based on Taylor and Kriegman (1995), which is a structure from motion algorithm exploiting constraints that are characteristic of architectural scenes. The user needs to manually match edges in the images with the edges in the model to recover parameters of both cameras and parametric polyhedral primitives for reconstruction of the architectural scene. The methods (Cipolla *et al.* 1999, Sinha *et al*. 2008) use vanishing point constraints for architectural scenes where parallel lines are abundant. A working system called PhotoBuilder was designed and implemented that allows a user to select a set of images either parallel or perpendicular in the world to build a model (Cipolla *et al.* 1999). In Sinha *et al*. (2008), the user draws outlines on the photographs to reconstruct piecewise planar 3D models of architectural structures and urban scenes from unordered photograph collections. Xiao *et al*. (2008) proposed to approximate orthographic images by fronto-parallel reference images for each facade. They decomposed a facade into rectilinear patches, and each patch was then augmented with a depth value optimized using the SfM depth data. The user intervention takes place in image space for controlling the decomposition and depth augmentation. The method (van den Hengel *et al*. 2007) interactively reconstructed 3D models by tracing the shape of the object over one or more frames of the video. Another representative interactive approach (Mueller *et al*. 2007) combines the procedural modeling pipeline with image analysis to produce photorealistic building models. In general, this type of method needs experts to specify rules to describe existing buildings.

There has also been a considerable amount of work involving 3D reconstruction from aerial images (Fischer *et al*. 1998, Gruen and Wang 1998, Zhu *et al.* 2004, Zebedin *et al*. 2006, Nyaruhuma *et al.* 2012, Xiao *et al.* 2012). The paper by Zebedin *et al*. (2006) is a representative work in city modeling from aerial images and was used for producing 3D models for Microsoft Bing maps. This work focuses on rooftop modeling and is complementary to ground-based methods (Xiao *et al*. 2008, 2009).

2.2 *Single images*

In contrast to the use of a number of images, 3D reconstruction from single images has drawn considerable attention from the computer vision and photogrammetry communities (Horry *et al.* 1997, van den Heuvel 1998, Liebowitz *et al.* 1999, Sturm and Maybank 1999, Criminisi *et al.* 2000, Oh *et al*. 2001, Zhang *et al.* 2001a, 2001b, Hoiem *et al.* 2005, Mueller *et al*. 2006, Wang and Ferrie 2008, Saxena *et al.* 2009). The existing single-view reconstruction methods may be roughly divided into two broad categories: nonmetric and metric reconstruction.

2.2.1 *Nonmetric reconstruction*

The nonmetric reconstruction methods do not reconstruct accurate geometry but focus on producing nice visual approximation (Horry *et al.* 1997, Oh *et al*. 2001, Zhang *et al.* 2001a, Hoiem *et al.* 2005, Saxena *et al.* 2009). "Automatic photo pop-up" (Hoiem *et al.* 2005) and "Make3D" (Saxena *et al.* 2009) are two recent representative works. The system in Hoiem *et al.* (2005) automatically constructs simple 3D models from a single outdoor image, based on the assumption that a scene is composed of a single ground plane, piecewise planar objects extruded vertically to the ground, and the sky. The main idea is to statistically model geometric classes defined by their orientation in the scene. They first classify each pixel as ground, vertical, or sky. These labels are then used to determine where to "cut" and "fold" in the image and produce a simple "pop-up"-type fly-through from an image. The method fails on scenes that do not satisfy this assumption.

"Make3D" (Saxena *et al.* 2009) is a supervised learning-based approach to address the problem of estimating depth maps from a single image. First, they create a training data set that consists of numerous images and their corresponding ground-truth depth maps collected from a laser scanner. Based on this training data set, they use a hierarchical multiscale Markov random field (MRF) to model the depths and the relation between depths at different image locations. Since they do not employ explicit assumptions about the structure of the scene, this enables their method to generalize well (Saxena *et al.* 2009).

Other methods like the paper by Zhang *et al.* (2001a) model free-form scenes by letting the user specify a sparse set of constraints, such as surface positions and normals, on a single painting or photograph and then optimizing for the best 3D model to satisfy these constraints. Instead of placing strong assumptions on either the shape or reflectance properties of the scene, they assume that the scene is represented by a piecewise continuous surface. Tour into the Picture (Horry *et al.* 1997) uses a spidery mesh to obtain a simple scene model from the image of the scene with a graphical user interface. This method produces impressive results for scenes that can be approximated by a one-point perspective (Hoiem *et al.* 2005).

2.2.2 *Metric reconstruction*

Metric reconstruction focuses on precise geometry recovery. Recovering 3D geometry from a single 2D image may have an infinite number of possible 3D interpretations without any assumptions. An illustration of this ill-posed problem is shown in Figure 2, which shows that the vertices of the cube and the polyhedron both project to the same 2D points.

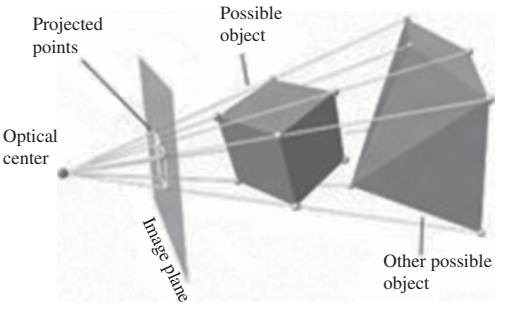

Figure 2. Illustration of the ill-posed problem: two distinct objects have the same 2D projections (Courtesy and copyright Grossmann 2002).

According to Grossmann (2002), the methods of 3D metric reconstruction from single images can be roughly classified as "model-based" and "constraint-based."

Urban structure is complicated with arbitrary shapes, but the majority of the buildings are rectilinear or piecewise planar. These assumptions are the basis for model-based methods (Debevec *et al.* 1996, Li *et al.* 2007). Most of the model-based approaches (Lowe 1991, Debevec *et al.* 1996) require a model-to-image fitting process, while constraint-based methods (van den Heuvel 1998, Liebowitz *et al.* 1999, Sturm and Maybank 1999, Criminisi *et al.* 2000, Zhang *et al.* 2001b) normally explore geometric properties, such as planarity, parallelism, and orthogonality, to reconstruct 3D geometry. The main idea of the model-based approaches is to obtain the reconstruction as a collection of "primitives" that best fits the image data (Grossmann 2002). Typically, the user needs to correspond to model edges (Debevec *et al.* 1996, Jelinek and Taylor 2001) with their image edges, and then the algorithms automatically compute model parameters such as width and height through finding a best fit to the image. The cost function that measures the disparity between the projected edges of the model and the edges marked in the image contains the unknown parameters such as camera interior and exterior parameters. Camera poses are computed by minimizing this cost function.

Model-based methods in the photogrammetry community are often called "CAD-based" (Grossmann 2002). The so-called CAD-based photogrammetry is due to the fact that photogrammetric tools are intended to be integrated with the existing CAD system (van den Heuvel 2000). The limitation of the model-based methods is obvious: they can only reconstruct objects for which models of the objects are available or can be decomposed into simpler shapes (Grossmann 2002).

The constraint-based methods exploit geometric properties inherent in the image such as parallelism and orthogonality for a reconstruction (Grossmann 2002). Although constraint-based approaches can treat more general scenes, they rely on a small collection of geometric properties that limits their applicability (Grossmann 2002). Vanishing point estimation and camera calibration are common steps in these types of methods. Caprile and Torre (1990) described methods to use vanishing points to recover intrinsic parameters from a single camera and extrinsic parameters from a pair of cameras. They showed that the orthocenter of the image triangle formed by the vanishing points is the principal point, and from the vanishing points corresponding to three mutually orthogonal directions at most three intrinsic parameters can be estimated. Although they were mainly

concerned with stereo cameras, their method is considered to be a means of access for the single-view calibration problem. In photogrammetry, the relation between vanishing points and camera calibration is well documented in Williamson and Brill (1990).

Criminisi *et al.* (2000) focus on scenes containing planes and parallel lines, and assume that at least a vanishing line of a reference plane and a vanishing point for a direction not parallel to the plane are available. Algorithms have been described to obtain measurements such as the distance between planes parallel to a reference plane, computing area, and length ratios on two parallel planes, and computing the camera's location. The limitation of this method is that the scene to be reconstructed must be composed of parallel planes linked by segments (Grossmann 2002). A Bayesian method is presented in Han and Zhu (2003) to recover 3D information from a single image. Multiple piecewise planar object reconstructions from single images are addressed through the constraints of connectivity and perspective symmetry (Liu and Stamos 2007). Huang and Cowan (2009) presented an automatic method of 3D reconstruction of a single indoor image.

3. Three-dimensional modeling using range data

There have been numerous projects for reconstructing statues, such as Stanford's "Digital Michelangelo Project" (http://graphics.stanford.edu/projects/mich) and IBM's "Piet`a Project" (http://www.research.ibm.com/pieta), historical sites (Guidi 2005, El-Hakim *et al*. 2007, 2008), building facades (Frueh and Zakhor 2003, Frueh *et al.* 2005, Nan *et al*. 2010, Zheng *et al*. 2010), and urban building models (Vosselman and Suveg 2001, Stamos and Allen 2002, Rottensteiner 2003, Verma *et al.* 2006, Chen and Chen 2008, Matei *et al*. 2008, Stamos *et al*. 2008, Zhou and Neumann 2008, 2009, 2010, 2011, Poullis and You 2009, Jozsa and Benedek 2013) using laser scanning data. Range segmentation plays an important role in generating 3D models. Hoover *et al.* (1996) provided a comprehensive study on comparison of existing segmentation algorithms.

3.1 *Range segmentation*

Previous range segmentation methods (Hoover *et al*. 1996) mainly focus on small objects in a well-controlled laboratory environment, which can be roughly classified into "edge based," "region based," and "hybrid" (Chen 2007). The edge-based segmentation extracts both depth and surface normal discontinuities and often fails if the image consists of weak edges. To avoid a fixed threshold, Bellon and Silva (2002) developed a method with an adaptive threshold based on the local neighborhood of each point for segmenting simple objects such as planar or polyhedral objects. Most edge-based segmentations work well on objects with simple shapes such as cubes and spheres and often fail to segment objects with complicated curved surfaces (Chen 2007). The region-based segmentation generates a group of pixels in terms of similar properties such as surface curvature. Some algorithms only work well on surfaces with simple shapes like polyhedral or spherical (e.g., Han *et al.* 1987, Jiang *et al.* 1996). Approaches to segmenting objects with arbitrary shapes often involve variable-order surface fitting and are computationally intensive (e.g., Besl and Jain 1988, Djebali *et al.* 2002). The region-based methods produce closed contours but can be limited in terms of providing accurate edge information (Chen 2007). Hybrid approaches combine edge detection with region growing (e.g., Yokoya and Levine 1989, Bhandarkar and Siebert 1992) and attempt to take advantage of strength from both sources (Chen 2007).

3.2 *3D building modeling using aerial LiDAR*

The research on range data processing has moved from well-controlled laboratory environments to real-world scenes. In recent years, creation of 3D building models from aerial LiDAR data has received considerable attention from the computer vision community (e.g., You *et al*. 2003, Verma *et al.* 2006, Matei *et al*. 2008, Poullis and You 2009, Zhou and Neumann 2010). In photogrammetry, this research has been under investigation since the late 1980s (e.g., Weidner and Foerstner 1995, Kraus and Pfeifer 1998, Ackermann 1999, Maas and Vosselman 1999, Vosselman 1999, Vosselman *et al*. 2001, Elaksher and Bethel 2002, Rottensteiner and Briese 2002, Oude Elberink and Vosselman 2011).

Two algorithms for creating building models using aerial LiDAR are presented in Maas and Vosselman (1999): The first algorithm analyzes invariant moments of the point clouds to extract building models, and the second algorithm uses the plane intersection principle to determine the roof structure. Vosselman *et al.* (2001) present two strategies to reconstruct building models from segmented planar surfaces and ground plans. They employ a 3D Hough transform to detect planar segments and merge them using least-squares estimation. Building modes are reconstructed by either detecting intersection lines and the so-called height jump edges or refining a coarse 3D model.

In computer vision, Frueh and Zakhor (2003) automatically created textured 3D building models using both ground- and aerial-based LiDAR data. The aerial LiDAR is employed to generate a Digital Surface Model (DSM) that contains terrain and building roof information that is normally difficult to obtain from the ground-based LiDAR. The DSM is then merged with the facade models using Monte Carlo localization. The redundant parts in the two meshes are eliminated and gaps are filled. You *et al.* (2003) proposed an interactive method for modeling buildings from aerial LiDAR data using predefined primitive libraries. Their method, resembling a model-based approach, is limited in terms of the availability of primitive libraries.

Recent methods (Verma *et al.* 2006, Matei *et al*. 2008, Zhou and Neumann 2008, 2009, Poullis and You 2009) introduced an automatic pipeline of creating 3D building modes from aerial LiDAR data with a focus on rooftop modeling. The common components include a classification algorithm to remove trees and noise, a segmentation algorithm to separate individual building patches and ground points, and a modeling algorithm to generate mesh models from building patches (Zhou and Neumann 2008). In the modeling stage, these methods first extract individual planar building roofs by using a plane-fitting algorithm, then use different heuristics to form a complete rooftop consisting of multiple planar pieces. For instance, Verma *et al.* (2006) employ a graph-based method to represent the relationships between various planar patches of a complex roof structure; Matei *et al.* (2008) regularize roof outlines by estimating the building orientation based on minimizing the number of vertices in a rectilinear approximation of the building; and Poullis and You (2009) create simple 3D models by simplifying boundaries of fitted planes. Instead of making assumptions on the angles between roof edges, Zhou and Neumann (2008, 2009) learn a set of principal directions of roof edges to align roof boundaries. In general, these methods all detect roofs using either predefined patterns such as planar shapes (Verma *et al.* 2006, Matei *et al*. 2008, Zhou and Neumann 2008, 2009, Poullis and You 2009) or user-defined primitive libraries (You *et al*. 2003), which are not generalized well to handle roofs with arbitrary shapes. To deal with roofs with arbitrary shapes, Zhou and Neumann (2010, 2011) extend the classic dual contouring (Ju *et al*. 2002) into a 2.5D method to produce crack-free models composed of arbitrarily shaped roofs and vertical walls connecting them.

3.3 *3D modeling using ground-based or mobile LiDAR*

This article focuses on building modeling using ground-based LiDAR in an outdoor environment. Unlike well-controlled laboratory environments, outdoor scenes contain buildings, trees, cars, pedestrians, and many other man-made or natural objects. Due to the positioning constraint for any ground-level data acquisition system, it is difficult to always obtain a complete and sufficient sampling of all the building surfaces. For instance, the rooftops and back of the buildings often cannot be scanned by a mobile LiDAR scanning system. The sampling rate of surface also varies in terms of its distance and orientation to the scanner (Chen and Chen 2008). For instance, the point clouds from the upper level of a high building are normally sparser than those from the lower level. Moreover, laser beams can penetrate transparent surfaces such as glass windows, doors, and walls, which result in insufficient LiDAR data collected from these surfaces. To deal with the data inconsistency, Wang *et al.* (2011) presented a combination of top-down with bottom-up approach to detect windows from mobile LiDAR data.

Ground-based laser scanning can be thought of as a type of stationary scanning method, in which multiple scans at different locations are needed to obtain "complete" data. Because building roofs are still often missing due to the viewpoint constraint, the meaning of "complete" is relative to the laser scanning data acquired from one location. To process the multiple scans, range registration is needed to combine the separated scans into a complete data set. A common 3D-to-3D registration method is the Iterative Closest Point (ICP) algorithm (Besl and McKay 1992). Stamos and Allen (2002) presented a typical 3D building modeling method using ground-based LiDAR. This is a bottom-up method that involves pointwise normal computation and classification using principal component analysis (PCA), cluster merging, surface fitting, and boundary extraction. Based on an assumption that the surfaces of a building can be represented as a bounded polyhedron, Chen and Chen (2008) detected planar regions and their intersections to generate piecewise planar building models from sparse laser scanning data. Schnabel *et al.* (2007a, 2007b) presented a top-down approach, which is based on a sequential RANSAC (Fischler and Bolles 1981) strategy, to detect planes, spheres, cylinders, cones, and tori from unorganized point clouds.

Mobile LiDAR data are continuously collected from the laser scanners mounted on a vehicle while driving at a posted speed. The earliest work by Zhao and Shibasaki (2001) employed two one-dimensional laser scanners mounted on the roof of a vehicle, scanning the scene horizontally and vertically, respectively. Based on a flat terrain assumption, a 3D model is reconstructed by registering and integrating horizontal and vertical scans. Frueh *et al.* (2003, 2005) developed a set of algorithms for generating textured facade meshes of cities from a series of vertical 2D surface scans and camera images. They classified points into foreground and background layers using a similar concept by Chang and Zakhor (2001) and detected major building structures in the depth images. Large holes in the background layer were filled.

A recent work (Nan *et al*. 2010) introduced an interactive tool to build detailed building models from mobile LiDAR data, in which the balconies of a building can be modeled well. The key idea is that the user-defined building blocks are automatically adjusted to fit well to the point data, considering the contextual relations with nearby similar blocks. Besides the research papers, some commercial 3D reconstruction and modeling software are also available in the market. A summary of existing commercial 3D building reconstruction software and university research prototypes is provided in the Appendix.

4. Three-dimensional modeling using images and range data

To use both images and range data for 3D modeling, these two different modalities have to be coregistered. The registration of multimodal data is the prerequisite for sensor fusions that have been widely studied (e.g., Li 2010, Zhang 2010, Liu *et al.* 2011, Brook *et al.* 2013). The problem of image-to-range registration involves alignment of the 2D image with the 2D projection of the range data, consisting of estimating the relative camera pose with respect to the range sensor. The registration result is critical not only for texture-mapping 3D models of large-scale scenes but also for applications such as image-based upsampling of range data (Torres-Mendez and Dudek 2004, Diebel and Thrun 2005, Yang *et al*. 2007, Dolson *et al*. 2010) and image-guided range segmentation (Matthieu *et al*. 2005, Barnea and Filin 2008). Upsampling range data is another means of using images with LiDAR data for 3D modeling. Sparse LiDAR point clouds, for instance those collected from single or cheap lasers, cannot represent the detailed geometry of the scene appropriately. Upsampled LiDAR data using camera images, which contain dense-enough points representing the scene, could be of significant value, e.g., enabling an inexpensive LiDAR/camera system to perform as well as an expensive LiDAR-only system. In the following, image-to-range registration, upsampling range data, and image-guided range segmentation are reviewed.

4.1 *Image-to-range registration*

Existing image-to-range registration methods range from keypoint-based matching (Becker and Haala 2007, Ding *et al.* 2008, Aguilera *et al.* 2009), structural features-based matching (Liu and Stamos 2005, 2007, Stamos *et al*. 2008, Wang and Neumann 2009), to mutual information-based registration (Mastin *et al.* 2009, Wang *et al.* 2012). The range data include terrestrial or aerial LiDAR. The images include vertical or oblique aerial images and ground-level images.

Keypoint-based matching (Becker and Haala 2007, Aguilera *et al.* 2009) is based on the similarity between laser reflectance images and the corresponding camera intensity images. First, each pixel of the laser reflectance image is encoded with its corresponding 3D coordinate. Then, the feature points are extracted by using either SIFT (Lowe 2004) or Foerstner operators (Foerstner and Guelch 1987) from both images. A robust matching strategy based on RANSAC (Fischler and Bolles 1981) and/or epipolar geometry constraint is employed to determine the correspondence pairs for computing fundamental matrix. Sensor registration is then achieved based on a robust camera spatial resection. Ding *et al.* (2008) registered oblique aerial images with a 3D model generated from aerial LiDAR data based on 2D and 3D corner features in the 2D images and 3D LiDAR model. The correspondence between extracted corners was based on the Hough transform and the generalized M-estimator sample consensus. The resultant corner matches are used in Lowe's algorithm (Lowe 2004) to refine camera parameters estimated from a combination of vanishing point computation and GPS/IMU readings. In general, the feature point extraction and robust matching are the key to a successful registration for these types of methods.

Instead of matching points, structural feature-based methods (Liu and Stamos 2005, Stamos *et al*. 2008, Wang and Neumann 2009) match structural features in both 2D and 3D spaces to estimate the relative camera poses. Direct matching of single-line features is error-prone due to the noise in both LiDAR and image data and the robustness of the detection algorithms. High-level structural features are helpful in increasing the robustness

of both detection and matching. Wang and Neumann (2009) registered aerial images with aerial LiDAR based on matching the so-called "3 Connected Segments," in which each linear feature contains three segments connected into a chain. They used a two-level RANSAC algorithm to refine the putative feature matches and estimated camera poses using methods (Hartley and Zisserman, 2004). Liu and Stamos (2005, 2007) extracted the so-called "rectangular parallelepiped" features, which are composed of vertical or horizontal 3D rectangular parallelepipeds in the LiDAR and 2D rectangles in the images, to estimate camera translation with a hypothesis-and-test scheme. Camera rotation was estimated based on at least two vanishing points. Since the vanishing points are required, their methods work well for ground-level data that are not efficient to handle aerial data with a weak perspective effect.

All the aforementioned methods are dependent on either the strong presence of parallel lines to infer vanishing points or on the availability of feature pair correspondences, which limits their applicability and robustness. Recent methods (Mastin *et al.* 2009, Wang *et al.* 2012) using statistical metrics, such as mutual information (Viola and Wells 1997), as a similarity measure for registering oblique aerial images and aerial LiDAR do not require any feature extraction process. This method searches for optimal camera poses through maximizing the mutual information between camera images and different attributes of LiDAR such as the LiDAR reflectance image, depth map, or a combination of both. Instead of using features, the mutual information method evaluates histograms and joint histograms using all the pixels in both images, which avoids the problems of feature extraction and correspondence.

Another method (Zhao *et al.* 2004) registered videos onto 3D point clouds through aligning a point cloud computed from the video with the one obtained from a range sensor. Although this method avoids the problems of feature extraction, it requires structure from motion techniques that are computationally expensive and limited in accuracy and robustness.

4.2 *Upsampling range data using images*

There appears to be relatively little work in using coregistered intensity images to upsample range data, at least in comparison to pure image-based super-resolution, e.g., Farsiu *et al.* (2004), Borman and Stevenson (1998), and Irani and Peleg (1991). One of the first attempts reported was based on MRFs (Torres-Mendez and Dudek 2004, Diebel and Thrun 2005). A common assumption here is that depth discontinuities correspond to intensity change in the images, which is not always true. The problem occurs when the image region with a similar color corresponds to objects with depth discontinuities. Yang *et al.* (2007) proposed an iterative bilateral filtering method for enhancing the resolution of range images. The authors also compared their approach with MRF, showing that this method allows for subpixel accuracy. Andreasson *et al.* (2006) compared five different interpolation schemes with the MRF method (Torres-Mendez and Dudek 2002, Diebel and Thrun 2005) and summarized four different metrics for confidence measures of interpolated range data. The assumption in their approach is similar to that of the MRF method that uses color similarity as an indication of depth similarity. Garro *et al.* (2009) projected points onto segmented color images and used bilinear interpolation to compute the depth value of the grid samples that belong to the same region. Dolson *et al.* (2010) addressed the problem of upsampling range data in dynamic environments based on a Gaussian framework.

There is some work that does super-resolution from depth data only. Basically the goal is to enhance the resolution by using multiple low-resolution depth maps that were obtained from close viewpoints (Kil *et al.* 2006, Schuon *et al*. 2009).

4.3 *Image-based range segmentation*

Camera images normally provide texture information for the generated 3D models (e.g., Stamos and Allen 2002, Frueh and Zakhor 2003, Brenner 2005, Chen and Stamos 2007, El-Hakim *et al*. 2007, Stamos *et al*. 2008, Pu and Vosselman 2009). Matthieu *et al*. (2005) and Barnea and Filin (2008) employed images for improving the accuracy of range segmentation, especially on object boundaries where accurate normal estimation is problematic. According to Matthieu (2005), most current work combines laser data with images independently and normally includes image-to-range registration and association of two steps (Matthieu *et al*. 2005). The association step involves integrating results from each of the data sources by performing parallel segmentations of both data (e.g., Matthieu *et al*. 2005, Barnea and Filin 2008), combining both segmentations to generate a grammar representation for a facade (Hohmann *et al*. 2009) and triangulating two camera images to generate 3D points for occluded parts filling in the point clouds (Dias *et al*. 2003).

5. Summary and conclusions

5.1 *Summary*

This article gives a general review in the areas of 3D modeling from images, LiDAR, and fusion of both. For image-based solutions, a fully automatic 3D reconstruction from video has generated impressive 3D models. Under arbitrary camera configurations such as Internet photos, sparse scene geometry can be recovered completely automatically. To generate dense reconstruction from such photo collections, view selection algorithms have to be developed to select images that are similar in appearance and scale so that multiview stereo methods can be applied. Interactive methods achieve best results but are difficult to scale up to large scale. 3D reconstruction from single images relies on different assumptions and fails when the assumptions are violated. Most methods in building modeling from aerial LiDAR focus on simple parametric building models such as polyhedral, prismatic, and flat roof models. The building models created from airborne solution are always coarse compared with those from the data collected from short distances such as ground-based or mobile LiDAR. Most methods using mobile LiDAR focus on generating photorealistic building models because of the sufficient geometric information contained in the point clouds. The registration between images and range data is a prerequisite for 3D modeling using both data sources. Most methods (e.g., Ding *et al.* 2008, Mastin *et al.* 2009, Wang and Neumann 2009) start with an initial registration and achieve optimal solutions through feature-based or mutual information registration. Methods of upsampling range data using images include deterministic (Andreasson *et al.* 2006, Yang *et al*. 2007, Garro *et al*. 2009) and probabilistic (Torres-Mendez and Dudek 2002, Diebel and Thrun 2005) approaches.

5.2 *Conclusions*

In the context of 3D building modeling, the mobile LiDAR technology has drawn intensive attention from the industrial and research communities due to its data collection

efficiency and capability of providing data with street-level details. It can be expected that 3D building modeling using mobile LiDAR data will be a trend. The challenging problems include data noise, data inconsistency such as windows, and incomplete data. A common method to address the incomplete data problem is to combine aerial with ground-based or mobile LiDAR to result in complete building models. In terms of sensor fusion, the registration of image and LiDAR is the prerequisite. The major challenges here are how to extract reliable features from 2D and 3D data and how to automatically match them. Mutual information registration does not require feature extraction and holds potential for building a robust registration system.

We hope that this study can promote interdisciplinary research collaborations among researchers with different backgrounds in the area of 3D building modeling. However, given the large amount of existing work, it is not possible to cover every work in this article. Readers may consider this article as a basis to further explore interesting work in the areas.

References

Ackermann, F., 1999. Airborne laser scanning present status and future expectations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54, 64–67.

Agarwa, S., *et al.*, 2009. Building Rome in a day. *In*: *International conference on computer vision*. 28 September – 2 October 2009, Kyoto, Japan.

Aguilera, D.G., Rodriguez Gonzalvez, P., and Gomez Lahoz, J., 2009. An automatic procedure for co-registration of terrestrial laser scanners and digital cameras. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 308–316.

Andreasson, H., Triebel, R., and Lilienthal, A., 2006. Vision-based interpolation of 3D laser scans. *In: Proceeding of the IEEE international conference on autonomous robots and agents*, *ICARA*. 12–14 December 2006, Palmerston North, New Zealand.

Barnea, S. and Filin, S., 2008. Segmentation of terrestrial laser scanning data by integrating range and image content. *In: The proceedings of XXIth ISPRS Congress*. 3–11 July 2008, Beijing, China.

Becker, S. and Haala, N., 2007. Combined feature extraction for facade reconstruction. *In*: P. Rönnholm, H. Hyyppä and J. Hyyppä, eds. *ISPRS workshop on laser scanning*. 12–14 September 2007, Espoo, Finland.

Bellon, O. and Silva, L., 2002. New improvements to range image segmentation by edge detection. *IEEE Signal Processing Letters*, 9, 43–45.

Besl, P.J. and Jain, R.C., 1988. Segmentation through variable-order surface fitting. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 10, 167–192.

Besl, P.J. and McKay, N.D., 1992. A method for registration of 3-D shapes. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14, 239–256.

Bhandarkar, S. and Siebert, A., 1992. Integrating edge and surface information for range image segmentation. *In*: *IEEE Proceeding of the Southeastcon '92*, vol. 25, 947–962. 12–15 April 1992, Birmingham, AL.

Borman, S. and Stevenson, R.L., 1998. Super-resolution from image sequences – a review. *In*: *Proceedings of the Midwest symposium on systems and circuits*, *MWSCAS '98*, 374–378. 9–12 August 1998, Notre Dame, IN.

Brenner, C., 2005. Building reconstruction from images and laser scanning. *International Journal of Applied Earth Observation and Geoinformation*, Theme Issue on "Data Quality in Earth Observation Techniques", 6 (3–4), 187–198.

Brook, A., Ben-Dor, E., and Richter, R., 2013. Modeling and monitoring urban built environment via multisource integrated and fused remote sensing data. *International Journal of Image and Data Fusion*, 4 (1), 2–32.

Caprile, B. and Torre, V., 1990. Using vanishing points for camera calibration. *International Journal of Computer Vision*, 4, 127–140.

Chang, N.L. and Zakhor, A., 2001. Constructing a multivalued representation for view synthesis. *International Journal of Computer Vision*, 45, 157–190.

Chen, C.C., 2007. *Range segmentation and registration for 3D modeling of large scale urban scenes*. Thesis (PhD). The City University of New York.

Chen, C. and Stamos, I., 2007. Range Image Segmentation for Modeling and Object Detection in Urban Scenes. *The 6th International Conference on 3-D Digital Imaging and Modeling*, 185–192. 21–23 August 2007, Montreal, Canada.

Chen, J. and Chen, B., 2008. Architectural modeling from sparsely scanned range data. *International Journal of Computer Vision*, 78, 223–236.

Cipolla, R., Robertson, D., and Boyer, E., 1999. Photobuilder – 3D models of architectural scenes from uncalibrated images. *In*: *Proceedings of the IEEE international conference on multimedia computing and systems* – volume 2, 25–31. 7–11 June 1999, Washington, DC.

Criminisi, A., Reid, I., and Zisserman, A., 2000. Single view metrology. *International Journal of Computer Vision*, 40 (2), 123–148.

Curless, B. and Levoy, M., 1996. A volumetric method for building complex models from range images. *In*: *Proceedings of SIGGRAPH*, 303–312. 4–9 August 1996, New Orleans, LA.

Debevec, P.E., Taylor, C.J., and Malik, J., 1996. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. *In*: *Proceedings of ACMSIGGRAPH*, 11–20. 4–9 August 1996, New Orleans, LA.

Dias, P., *et al.*, 2003. Registration and fusion of intensity and range data for 3D modelling of real world scenes. *In*: *International conference on 3D digital imaging and modeling*, 418. 6–10 October 2003, Banff, Canada.

Diebel, J. and Thrun, S., 2005. An application of Markov random fields to range sensing. *In*: Y. Bengio, D. Schuurmans, J. Lafferty, C.K.I. Williams and A. Culotta, eds. *Proceedings of conference on neural information processing systems* (NIPS). 7–10 December 2005, Vancouver, Canada.

Ding, M., Lyngbaek, K., and Zakhor, A., 2008. Automatic registration of aerial imagery with untextured 3D LiDAR models. *In*: *The proceeding of IEEE computer vision and pattern recognition*, 1–8. 24–26 June 2008, Anchorage, AK.

Djebali, M., Melkemi, M., and Sapidis, N.S., 2002. Range-image segmentation and model reconstruction based on a fit-and-merge strategy. *In*: *Symposium on solid modeling and applications*, 127–138. 17–21 June 2002, Saarbrücken, Germany.

Dolson, J., *et al.*, 2010. Upsampling range data in dynamic environments. *In*: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1141–1148.

Elaksher, A.F. and Bethel, J.S., 2002. Reconstructing 3D buildings from LiDAR data. *In*: *ISPRS Commission III*, *Symposium*. 9–13 September 2002, Graz, Austria.

El-Hakim, S., *et al.*, 2007. Detailed 3D modelling of castles. *International Journal of Architectural Computing (IJAC)*, 5, 199–220.

El-Hakim, S., *et al.*, 2008. Surface reconstruction of large complex structures from mixed range data – the erechtheion experience. *In*: *The proceedings of XXIth ISPRS congress*. 3–11 July 2008, Beijing, China.

Farsiu, S., *et al.*, 2004. Fast and robust multi-frame super-resolution. *In*: *IEEE Transactions on Image Processing*, 13, 1327–1344.

Finsterwalder, S., 1937. Die geometrischen grundlagen der photogrammetrie. Jahres- bericht Deutsche Mathem. Vereinigung, VI, 2, Teubner Verlag, Leipzig, 1–41, 1899. In: *Sebastian Finsterwalder zum 75*. *Geburtstag Gesellschaft fuer Photogram-metrie*, *Wichmann Verlag*, Berline.

Fischer, A., *et al.*, 1998. Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. *Computer Vision and Image Understanding*, 72 (2), 185–203.

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

Fitzgibbon, A. and Zisserman, A., 1998. Automatic 3D model acquisition and generation of new images from video sequence. *In*: *Proceedings of European signal processing conference*, 1261–1269. 8–11 September 1998, Rhodes, Greece.

Foerstner, W. and Guelch, E., 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features. *In*: *ISPRS intercommission workshop Interlaken*, 281–305. June 1987, Interlaken, Switzerland.

Frueh, C., Jain, S., and Zakhor, A., 2005. Data processing algorithms for generating textured 3D building facade meshes from laser scans and camera images. *International Journal of Computer Vision*, 61, 159–184.

Frueh, C. and Zakhor, A., 2003. Constructing 3D city models by merging ground-based and airborne views. *In*: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 16–22 June 2003, Madison, WI.

Fuchs, C., Guelch, E., and Foerstner, W., 1998. OEEPE survey on 3D-city models. *In*: *OEEPE Publication*, 9–124.

Furukawa, Y., *et al.*, 2010. Towards internet-scale multi-view stereo. *In*: *IEEE International Conference on Computer Vision and Pattern Recognition*. 13–18 June 2010, San Francisco, CA.

Garro, V., *et al.*, 2009. A novel interpolation scheme for range data with side information. *In*: *Proceedings of the conference for visual media production*, *CVMP '09*, 52–60. 12–13 November 2009, London, UK.

Goesele, M., *et al.*, 2007. Multi-view stereo for community photo collections. *In*: *IEEE International Conference on Computer Vision and Pattern Recognition*, 1–8. 18–23 June 2007, Minneapolis, MN.

Gool, L.V., *et al.*, 2007. Towards mass-produced building models. *Photogrammetric Image Analysis*, 209–220.

Grossmann, E., 2002. *Maximum likelihood 3D reconstruction from one or more uncalibrated views under geometric constraints*. Thesis (PhD). Universidade Tècnica de Lisboa – Instituto Superior Tècnico.

Gruen, A. and Wang, X., 1998. CC-modeler: a topology generator for 3-d city models. *Journal of Photogrammetry and Remote Sensing*, 53, 286–295.

Guidi, G., *et al.*, 2005. 3D digitization of a large model of imperial Rome. *In*: *Proceedings of the fifth international conference on 3-D digital imaging and modeling*, 565–572. 13–16 June 2005, Ottawa, Canada.

Han, J., Volz, R.A., and Mudge, T.N., 1987. Range image segmentation and surface parameter extraction for 3-D object recognition of industrial parts. *In*: *IEEE International Conference on Robotics and Automation*, 4, 380–386. March 1987, Raleigh, NC.

Han, F. and Zhu, S.-C., 2003. Bayesian reconstruction of 3d shapes and scenes from a single image. *In*: *Proceedings of the first IEEE international workshop on higher-level knowledge in 3D modeling and motion analysis*, 12–20. 17 October 2003, Nice, France.

Hartley, R.I. and Zisserman, A., 2004. *Multiple view geometry in computer vision.* 2nd ed. Cambridge: Cambridge University Press, ISBN: 0521540518.

Hohmann, B., *et al.*, 2009. Cityfi: high-quality urban reconstruction by fitting shape grammars to image and derived textured point clouds. *In*: F. Remondino, S. El-Hakim and L. Gonzo, eds. *Proceedings of the international workshop 3D-ARCH*. 25–28 February 2009, Trento, Italy.

Hoiem, D., Efros, A.A., and Hebert, M., 2005. Automatic photo pop-up. *ACM Transaction on Graphics*, 24, 577–584.

Hoover, A., *et al.*, 1996. An experimental comparison of range image segmentation algorithms. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 18, 673–689.

Horry, Y., Anjyo, K.-I., and Arai, K., 1997. Tour into the picture: using a spidery mesh interface to make animation from a single image. *In*: *SIGGRAPH '97 Proceedings of the 24th annual conference on computer graphics and interactive techniques*, 225–232. 3–8 August 1997, Los Angeles, CA.

Huang, J. and Cowan, B., 2009. Simple 3D reconstruction of single indoor image with perspective cues. *In: Proceedings of the Canadian conference on computer and robot vision*, 140–147. 25–27 May 2009, Kelowna, Canada.

Irani, M. and Peleg, S., 1991. Improving resolution by image registration. *CVGIP: Graph Models Image Processing*, 53, 231–239.

Jelinek, D. and Taylor, C.J., 2001. Reconstruction of linearly parameterized models from single images with a camera of unknown focal length. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 23, 767–773.

Jiang, X.Y., Bunke, H., and Meier, U., 1996. Fast range image segmentation using high level segmentation primitives. *In*: *Proceedings of the 3rd IEEE workshop on applications of computer vision* (WACV '96), 83–88. 2–4 December 1996, Sarasota, FL.

Jozsa, O. and Benedek, C., 2013. Analysis of 3D dynamic urban scenes based on LiDAR point cloud sequences. *KÉPAF 2013*.

Ju, T., *et al.*, 2002. Dual contouring of Hermite data. *ACM Transaction on Graphics*, 21, 339–346.

Kil, Y.J., Amenta, N., and Mederos, B., 2006. Laser scanner super-resolution. *In:* M. Botsch and B. Chen, eds. *Point Based Graphics*, 9–16.

Kraus, K. and Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 53, 193–203.

Kruppa, E., 1913. Zur ermittlung eines objectes aus zwei perspektiven mit innerer orientierung. *Sitzungsberichte der Mathematisch Naturwissenschaftlichen Kaiserlichen Akademie der Wissenschaften*, 122, 1939–1948.

Li, D., 2010. Remotely sensed images and GIS data fusion for automatic change detection. *International Journal of Image and Data Fusion*, 1 (1), 99–108.

Li, H. and Hartley, R., 2006. Five-point motion estimation made easy. *In*: *Proceedings of the 18th international conference on pattern recognition*, vol. 01. 20–24 August 2006, Hong Kong.

Li, Z., Liu, J., and Tang, X., 2007. A closed-form solution to 3D reconstruction of piecewise planar objects from single images. *In*: *Proceeding of IEEE computer vision and pattern recognition*, 1–6. 18–23 June 2007, Minneapolis, MN.

Liebowitz, D., Criminisi, A., and Zisserman, A., 1999. Creating architectural models from images. *In*: *Annual conference of the European association for computer graphics* (eurographics), vol. 18, 39–50. 7–11 September 1999, Milan, Italy.

Liu, L. and Stamos, I., 2005. Automatic 3D to 2D registration for the photorealistic rendering of urban scenes. *In*: *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, Volume 2, 137–143. 20–26 June 2005, San Diego, CA.

Liu, L. and Stamos, I., 2007. A systematic approach for 2d-image to 3d-range registration in urban environments. *In*: *The international conference on computer vision*, 1–8. 14–20 October 2007, Rio de Janeiro, Brazil.

Liu, C., Wu, H., and Zhang, Y., 2011. Extraction of urban 3D features from LiDAR data fused with aerial images using an improved mean shift algorithm. *Survey Review*, 43 (322), 402–414.

Lowe, D.G., 1991. Fitting parameterized three-dimensional models to images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 13, 441–450.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60, 91–110.

Maas, H.G. and Vosselman, G., 1999. Two algorithms for extracting building models from raw laser altimetry data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54, 153–163.

Mastin, A., Kepner, J., and Fisher, J., 2009. Automatic registration of LiDAR and optical images of urban scenes. *In*: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2639–2646. 25 June 2009, Miami Beach, FL.

Matei, B.C., *et al.*, 2008. Building segmentation for densely built urban regions using aerial LiDAR data. *In*: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*. 24–26 June 2008, Anchorage, AK.

Matthieu, D.A.B., *et al.*, 2005. Strategy for the extraction of 3D architectural objects from laser and image data acquired from the same viewpoint. *In*: S. El-Hakim, F. Remondino and L. Gonzo, eds. *International archives of photogrammetry*, *remote sensing and spatial information sciences*, vol. 36. 22–24 August 2005, Mestre-Venice, Italy.

Micusik, B. and Kosecka, J., 2009. Piecewise planar city 3D modeling from street view panoramic sequences. *In*: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2906–2912. 27 September – 4 October 2009, Kyoto, Japan.

Mueller, P., *et al.*, 2006. Procedural modeling of buildings. *ACM Transaction on Graphics*, 25, 614–623.

Mueller, P., *et al.*, 2007. Image-based procedural modeling of facades. *ACM Transaction on Graphics*, 26 (3), 85.

Nan, L., *et al.*, 2010. Smartboxes for interactive urban reconstruction. *ACM Transaction on Graphics*, 29 (4), 93.

Nister, D., 2004a. An efficient solution to the five-point relative pose problem. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 26, 756–777.

Nister, D., 2004b. Automatic passive recovery of 3D from images and video. *In*: *Proceedings of the 3D data processing*, *visualization*, *and transmission*, *2nd international symposium*, *3DPVT '04*, 438–445. 6–9 September 2004, Thessaloniki, Greece.

Nocedal, J. and Wright, S.J., 2006. *numerical optimization*. 2nd ed. Berlin: Springer.

Nyaruhuma, A.P., Gerke, M., and Vosselman, G., 2012. Verification of 2D building outlines using oblique airborne images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 71, 62–75.

Oh, B.M., *et al.*, 2001. Image-based modeling and photo editing. *In*: *Proceedings of the 28th annual conference on computer graphics and interactive techniques*, *SIGGRAPH '01*, 433–442. 12–17 August 2001, Los Angeles, CA.

Oude Elberink, S. and Vosselman, G., 2011. Quality analysis of 3D building models reconstructed from airborne laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66 (1), 56–68.

Palmer, S., 1999. *Vision science: from photons to phenomenology.* Cambridge, MA: MIT Press.

Pollefeys, M., *et al.*, 2004. Visual modeling with a hand-held camera. *International Journal of Computer Vision*, 59, 207–232.

Pollefeys, M., *et al.*, 2008. Detailed real-time urban 3D reconstruction from video. *International Journal of Computer Vision*, 78 (2), 143–167.

Poullis, C. and You, S., 2009. Automatic reconstruction of cities from remote sensor data. *In*: *Proceeding of IEEE conference on computer vision and pattern recognition* (CVPR). 20–25 June 2009, Miami Beach, FL.

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

Remondino, F. and Ei-Hakim, S., 2006. Image-based 3D modeling: a review. *The Photogrammetric Record*, 21 (115), 269–291.

Rottensteiner, F., 2003. Automatic generation of high-quality building models from LiDAR data. *IEEE Computer Graphics and Applications*, 23, 42–50.

Rottensteiner, F. and Briese, C., 2002. A new method for building extraction in urban areas from high-resolution LiDAR data. *In*: *ISPRS commission III*, *symposium*, 295–301. 9–13 September 2002, Graz, Austria.

Saxena, A., Sun, M., and Ng, A.Y., 2009. Make 3D: learning 3D scene structure from a single still image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31, 824–840.

Scharstein, D. and Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47, 7–42.

Schnabel, R., Wahl, R., and Klein, R., 2007a. RANSAC based out-of-core point-cloud shape detection for city-modeling. *Schriftenreihe des DVW, Terrestrisches Laser-Scanning* (TLS 2007)).

Schnabel, R., Wahl, R., and Klein, R., 2007b. Efficient RANSAC for point-cloud shape detection. *Computer Graphics Forum*, 26, 214–226.

Schuon, S., *et al.*, 2009. Lidarboost: depth superresolution for TOF 3D shape scanning. *In*: *Proceeding of the IEEE conference on computer vision and pattern recognition*. 20–25 June 2009, Miami Beach, FL.

Seitz, S.M., *et al.*, 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. *In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition*, volume 1, 519–528. 17–22 June 2006, New York, NY.

Sinha, S.N., *et al.*, 2008. Interactive 3D architectural modeling from unordered photo collections. *In*: *ACM SIGGRAPH Asia*, No. 159, 1–10. 10–13 December 2008, Singapore.

Snavely, K.N., 2008. *Scene reconstruction and visualization from internet photo collections*. Thesis (PhD). University of Washington.

Snavely, N., Seitz, S.M., and Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. *In*: *ACM SIGGRAPH*, 835–846.

Stamos, I., *et al.*, 2008. Integrating automated range registration with multiview geometry for the photorealistic modelling of large-scale scenes. *International Journal of Computer Vision*, 78, 237–260.

Stamos, I. and Allen, P.K., 2002. Geometry and texture recovery of scenes of large scale. *Computer Vision and Image Understanding*, 88, 94–118.

Sturm, P.F. and Maybank, S.J., 1999. A method for interactive 3D reconstruction of piecewise planar objects from single images. *In*: *Proceeding of BMVC*, 265–274. 13–16 September 1999, Nottingham, UK.

Taylor, C.J. and Kriegman, D.J., 1995. Structure and motion from line segments in multiple images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 17, 1021–1032.

Torres-Mendez, L.A. and Dudek, G., 2004. Reconstruction of 3d models from intensity images and partial depth. *In*: *Proceedings of the 19th national conference on artificial intelligence*, AAAI'04, 476–481. 25–29 July 2004, San Jose, CA.

Triggs, B., *et al.*, 2000. Bundle adjustment – a modern synthesis. *In*: B. Triggs, A. Zisser-man, and R. Szeliski, eds. *Vision algorithms: theory and practice, vol. 1883 of Lecture Notes in Computer Science.* Berlin: Springer-Verlag, 298–372.

Trucco, E. and Verri, A., 1998. *Introductory techniques for 3-D computer vision*. Upper Saddle River, NJ: Prentice Hall.

van den Heuvel, A., 1998. Vanishing point detection for architectural photogrammetry. *International Archives of Photogrammetry and Remote Sensing*, 32, 652–659.

van den Heuvel, F.A., 2000. Trends in cad-based photogrammetric measurement. *International Archives of Photogrammetry and Remote Sensing*, 33 (5/2), 852–863.

van den Hengel, A., *et al.*, 2007. Video-trace: rapid interactive scene modelling from video. *ACM Transaction on Graphics*, 26 (3), 86.

Vanegas, C.A., *et al.*, 2010. Modelling the appearance and behaviour of urban spaces. *Computer Graphics Forum*, 29 (1), 25–42.

Verma, V., Kumar, R., and Hsu, S., 2006. 3D building detection and modeling from aerial LiDAR data. *In*: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition* – volume 2, 2213–2220. 17–22 June 2006, New York, NY.

Viola, P. and Wells, W.M.III, 1997. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24, 137–154.

Vosselman, G., 1999. Building reconstruction using planar faces in very high density height data. *In: ISPRSGIS99*, 87–92.

Vosselman, G., *et al.*, 2001. 3D building model reconstruction from point clouds and ground plans. *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXIV-3/W4, 37–43.

Vosselman, G. and Suveg, I., 2001. Map based building reconstruction from laser data and images. *In*: *Proceedings of automatic extraction of man-made objects from aerial and space images*, (III), Ascona, 11–15 June, Balkema Publishers, 231–239.

Wang, R., Bach, J., and Ferrie, F., 2011. Window detection from mobile LiDAR data. *In*: *IEEE workshop on applications of computer vision* (WACV), 5–7 January, Kona, Hawaii.

Wang, R. and Ferrie, F., 2008. Camera localization and building reconstruction from single monocular images. *In: CVPR workshop on visual localization for mobile platforms*, 11–20. 28 June 2008, Anchorage, AK.

Wang, R., Ferrie, F., and Macfarlane, J., 2012. Automatic registration of mobile LiDAR and spherical panoramas. *In*: *The IEEE computer vision and pattern recognition workshop on point cloud processing in computer vision*, 33–40. 16 June 2012, Providence, RI.

Wang, L. and Neumann, U., 2009. A robust approach for automatic registration of aerial images with untextured aerial LiDAR data. *In: The proceeding of the IEEE conference on computer vision and pattern recognition*, 2623–2630. 20–25 June 2009, Miami, FL.

Weidner, U. and Foerstner, W., 1995. Towards automatic building reconstruction from high resolution digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50, 38–49.

Williamson, J.R. and Brill, M.H., 1990. *Dimensional analysis through perspective – a reference manual*. Dubuque, IA: Kendal/Hunt Publishing Co.

Xiao, J., *et al.*, 2008. Image-based façade modelling. *ACM Transaction on Graphics*, 27 (161), 1–10.

Xiao, J., *et al.*, 2009. Image-based street-side city modelling. *ACM Transaction on Graphics*, 28 (114), 1–12.

Xiao, J., Gerke, M., and Vosselman, G., 2012. Building extraction from oblique airborne imagery based on robust façade detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 68, 157–165.

Yang, Q., *et al.*, 2007. Spatial-depth super resolution for range images. *In*: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1–8. 18–23 June 2007, Minneapolis, MN.

Yokoya, N. and Levine, M.D., 1989. Range image segmentation based on differential geometry: a hybrid approach. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 11, 643–649.

You, S., *et al.*, 2003. Urban site modeling from LiDAR. *In: Proceeding of the international conference on computational science and its applications: Part III*, *ICCSA'03*, 579–588. 18–21 May 2003, Montreal, Canada.

Zebedin, L., *et al.*, 2006. Towards 3d map generation from digital aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60 (no. 6), 413–427.

Zhang, J., 2010. Multi-source remote sensing data fusion: status and trends. *International Journal of Image and Data Fusion*, 1 (1), 5–24.

Zhang, L., *et al.*, 2001a. Single view modeling of free-form scenes. *In: Proceedings of computer vision and pattern recognition*, 990–997.

Zhang, Z., Zhang, J., and Zhang, S., 2001b. 3D building reconstruction from single images. *In*: *Workshop on automatic engineering surveying of channel*.

Zhao, W., Nister, D., and Hsu, S., 2004. Alignment of continuous video onto 3d point clouds. *In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition*, 964–971. 27 June – 2 July 2004, Washington, DC.

Zhao, H. and Shibasaki, R., 2001. Reconstructing urban 3d model using vehicle-borne laser range scanners. *In*: *Proceeding of the third international conference on 3-D digital imaging and modeling*, 349–356. 28 May – 1 June 2001, Quebec City, Canada.

Zheng, Q., *et al.*, 2010. Non-local scan consolidation for 3D urban scenes. *ACM Transaction on Graphics*, 29 (4), 94.

Zhou, Q. and Neumann, U., 2008. Fast and extensible building modeling from airborne LiDAR data. *In: ACM SIGSPATIAL GIS*. 5–7 November 2008, Irvine, CA.

Zhou, Q.-Y. and Neumann, U., 2009. A streaming framework for seamless building reconstruction from large-scale aerial LiDAR data. *In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition*, 2759–2766. 20–25 June 2009, Miami, FL.

Zhou, Q.-Y. and Neumann, U., 2010. 2.5D dual contouring: a robust approach to creating building models from aerial LiDAR point clouds. *In: The proceeding of European conference on computer vision*, 115–128. 5–11 September 2010, Heraklion, Greece.

Zhou, Q.-Y. and Neumann, U., 2011. 2.5D building modeling with topology control. *In: The proceedings of the IEEE computer society conference on computer vision and pattern recognition*. 20–25 June 2011, Colorado Springs, CO.

Zhu, Z., Hanson, A.R., and Riseman, E.M., 2004. Generalized parallel-perspective stereo mosaics from airborne video. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 26, 226–237.

Zygmunt, P., 2001. Perception viewed as an inverse problem. *Vision Research*, 24, 3145–3161.

Appendix

Table 1. Existing commercial 3D building reconstruction systems and research prototypes.

System	Developer/ researcher	Input data	Description
CC-Modeler	CyberCity AG & ETH, Zurich	Calibrated stereo pan of aerial images	Semi-automated photogrammetric 3D reconstruction system
inJECT	Inpho GmbH & Bonn university, Germany	Calibrated single, stereo, or multiple overlapping aerial images	Semi-automated Constructive Solid Geometry based approach
Ascender	University of Massachusetts	Calibrated multiple aerial (nadir and oblique) images	Automated 3D building model reconstruction
SiteCity	Digital Mapping Laboratory, CMU,	Calibrated multiple aerial (nadir and oblique) images	Semi-automated photogrammetric 3D reconstruction system
ImageModeler	RealViz & INRIA, France	At least two photos taken from different positions	Accurate 3D measurement and modelling from photos
PhotoBuilder	Oxford University, UK	Uncalibrated two or more photos	Vanishing points based method to 3D reconstruction
Nverse Photo	Precision Lightworks, USA	Two or more aerial images	A series of plug-in components
Shape Capture	ShapeQuest Inc. & NRC, Canada,	Single or more photos	Accurate 3D measurement and modelling from single or more photos
PhotoModeler	Eos Systems, Canada	Single or more photos	Accurate 3D measurement and modelling from single or more photos
PhotoGenesis	Plenoptics Ltd, UK	Uncalibrated single or more photos	Semi-automated model-based 3D reconstruction system
Photosynth	Microsoft	Internet photos	Sparse 3D model generation for navigating images in 3D space
Pix4UAV	Pix4D, Switzerland	Aerial images	Automatic 3D model generation from aerial images
C3	Apple, USA	Aerial images	Automatic 3D model generation from aerial images
Edgewise	ClearEdge3D, USA	Range data	3D modeling using range data