3D Measurements of Buildings and Environment for Harbor Simulators

Report UMINF 09.19

Niclas Börlin* Christina Igasto†

Department of Computing Science
Umeå University

October 15, 2009

Abstract

Oryx Simulations develops and manufactures real-time physics simulators for training of harbor crane operator in several of the world's major harbors. Currently, the modelling process is labor-intensive and a faster solution that can produce accurate, textured models of harbor scenes is desired. The accuracy requirements vary across the scene, and in some areas accuracy can be traded for speed. Due to the heavy equipment involved, reliable error estimates are important throughout the scene.

This report surveys the scientific literature of 3D reconstruction algorithms from aerial and terrestrial imagery and laser scanner data. Furthermore, available software solutions are evaluated.

The conclusion is that the most useful data source is terrestrial images, optionally complemented by terrestrial laser scanning. Although robust, automatic algorithms exist for several low-level subproblems, no automatic high-level 3D modelling algorithm exists that satisfy all the requirements. Instead, the most successful high-level methods are semi-automatic, and their respective success depend on how well user input is incorporated into an efficient workflow.

Furthermore, the conclusion is that existing software cannot handle the full suite of varying requirements within the harbor reconstruction problem. Instead we suggest that a 3D reconstruction toolbox is implemented in a high-level language, Matlab. The toolbox should contain state-of-the-art low-level algorithms that can be used as "building blocks" in automatic or semi-automatic higher-level algorithms. All critical algorithms must produce reliable error estimates.

The toolbox approach in Matlab will be able to simultaneously support basic research of core algorithms, evaluation of problem-specific high-level algorithms, and production of industry-grade solutions that can be ported to other programming languages and environments.

*niclas.borlin@cs.umu.se
†Maiden name: Christina Olsén

Contents

1 Introduction

1.1 Background

Oryx simulations[1] develops and manufactures real-time physics simulator for e.g. harbor environments. Among its customers are the harbors in Gothenburg, Rotterdam, Kuala Lumpur, and Shanghai. The simulators are used for education of harbor crane operators. Currently, the items within the simulator environment are hand-modelled and therefore a large amount of objects present in a harbor scene are not modelled. Furthermore, the surrounding is only introduced in a limited fashion into the simulation, resulting in a synthetic "look-and-feel". Recently, customers have presented the desire to have more realistically looking environment simulators. This would not only be more aesthetically pleasing but only be beneficial to training and smooth the transition between the training and real-world environment.

1.2 Aim

The aim of this pilot study is twofold: 1) Survey existing algorithms and software for creating textured 3D models of objects and the surrounding environment from images and other information sources. 2) Unless a software solution is available for the harbor reconstruction problem, formulate an implementation project with the necessary capabilities. Of the general requirements we mention speed, flexibility, and error estimates: Since the crane operators are to operate *real* heavy equipment after training, it is of paramount importance to have reliable error estimates of the measured values that comprise their virtual training environment.

2 Harbor modelling requirements

A harbor scene has different objects with different capture requirements. Furthermore, the requirements on the captured environment differ. In this context, objects are generally considered man-made whereas the environment is not.

Objects may be classified into *active objects*, *obstacles*, and *landmarks*. Potential attributes to reconstruct are shape (geometry), position, and texture. The environment consists of *work areas*, the *general area* and *the horizon*. Attributes to reconstruct are the topography (shape and position), and texture.

2.1 Active objects

The objects with the highest requirements for geometry and texture are the *active objects*. Active objects are objects that can be manipulated in the simulation environment, e.g. cargo containers or pallets. However, their exact position do not need to be recovered.

[1]http://www.oryx.se

2.2 Work areas

The areas with the highest requirement for topography and texture are the *work areas* where the active objects are to be manipulated. Examples are container storage areas or loading-unloading areas for pallets.

2.3 The general area

The general area consist of everything except the work areas. Parts of the general area may be used for transporting objects, but no manipulation of active objects generally takes place in the general area.

The exact topography of the general area do not need to be known with a high precision, and a high-quality texture is generally not need. However, in some areas, e.g. road junctions, the road markings may have to be of high quality.

Within the general area, *obstacles* and *landmarks* are placed.

2.4 Obstacles

Obstacles are objects that are not intended to be manipulated. However, they should not be bumped into during e.g. a transportation. As such, they have medium requirements on geometry and texture. Furthermore, their position should be known with medium precision. Examples of obstacles include "concrete pigs" and light towers.

2.5 Landmarks

Landmarks are buildings that an operator can use for navigation. Most buildings outside the work area are considered landmarks. The requirement for the exact position, size and texture are comparably low. However, they must still look "good enough" from the important viewpoints within the scene.

2.6 The horizon

The horizon consist of the part of the environment that is considered "far enough" away not to have to be individually modeled. However, if the real scene has an interesting horizon, e.g. a city skyline, the horizon may still be important for navigation and realism. The horizon is considered to have medium requirement for the angular position and texture.

3 Other requirements

The harbor is a busy workplace, and site access for data capture may thus be limited. Furthermore, the cost for data acquisition should not be too high. Finally, the visualization quality is especially important from select viewpoint, e.g. at the top of the work cranes and loading/unloading areas.

4 Literature study

4.1 Background

The studied literature falls mainly within the research fields of *Photogrammetry*, *Computer Vision*, and, to a lesser extent, *Computer Graphics* and *Surveying*. See Appendix A for a list of sources and Appendix B for a list of grouped references.

Photogrammetry[2] has developed since the mid-1850:s, originally as a technique for creating accurate topographical maps (McGlone et al. 2004, Ch. 1). Only recently, digital images have become standard input, and some 3D measurements is still performed manual on analog aerial images. Photogrammetry carries a strong statistical tradition, with error analysis and blunder detection being an integral part of most algorithms.

Surveying (or *Land surveying*) has historically been used longer than photogrammetry to construct maps. Surveying techniques include angle measurements between distinct point by a *theodolite*. Modern surveying is typically by *tacheometry*, where a laser theodolite can measure both angles and distances. For optimal accuracy and identification, highly reflective synthetic targets can be used. Often the theodolite is combined with a Global Positioning System (GPS) receiver for geo-referencing (Grussenmeyer et al. 2008).

Computer Vision has developed from the desire to make computers "see"(Hartley and Zisserman 2003, Foreword), i.e. to detect, measure, analyze, and understand the 3D environment. Computer Vision has a solid foundation in mathematics, especially in projective geometry and linear algebra. Many algorithms are oriented towards full automation. The interest in 3D reconstruction from the *Computer Graphics* area is based on the desire to capture and visualize *real* scenes rather than synthetic ones. The main strength of the research field lies in rendering and visualization.

4.2 3D reconstruction methods — overview

The 3D reconstruction methods presented in the literature differ in four major aspects; sensor type, sensor platform, algorithmic approach, and error treatment. The sensor type can be range-based (laser scanning, LIDAR[3]) or image-based. Either acquisition mode can be terrestrial (ground-based) or aerial (airborne). The algorithmic approaches differ widely based on the input and output requirements. Finally, the methods differ in their approach to errors, from a rigorous error analysis with presented precision values in object space coordinates, e.g. m, to error analysis in image coordinates or no error analysis at all, i.e. "it looks fine".

[2] from *photos*—light, *gramma*—something drawn or written, and *metron*—to measure
[3] LIght Detection and Ranging, "laser radar"

4.3 Sensor type and platform

4.3.1 Laser scanner

Most laser scanners measure the time-of-flight between an emitted laser pulse and its reflection. One ("line scanners") or two ("image scanners") rotating mirrors enable the laser to "scan" its surrounding. In principle, the recorded time is used to calculate the coordinates of one 3D point. However, more advanced scanners exist that record multiple echos per pulse, the reflected intensity, and even color (Akca and Gruen 2007; Remondino et al. 2005; Rottensteiner et al. 2007). Laser scanners can either be terrestrial (*TLS — Terrestrial Laser Scanners*) or aerial (*LIDAR*).

The basic algorithm for 3D reconstruction with a laser scanner is (see e.g. Remondino (2006b, Ch. 1)):

1. *Acquisition* of a in a scanner-local coordinate system.
2. *Co-registration* of multiple point clouds into a common, global, coordinate system.
3. *Segmentation and structuring* of the point cloud, surface generation.
4. Extraction of *texture data*.

4.3.2 Image-based techniques

Image-based techniques are today almost entirely based on digital still and video cameras. Both types of cameras can either be single or mounted in stereo or in multi-nocular[4] configurations. Airborne or spaceborne cameras are custom-built whereas many consumer digital cameras today have a high enough quality to be used for 3D measurements (Fraser and Cronk 2009). Classical aerial imagery is taken in regular patterns at high altitude (2000-5000 m) with nadir-mounted[5] cameras. Some modern cameras are so called *pushbroom cameras*, consisting of three to four lines angled forward, nadir, and backward (McGlone et al. 2004, Ch. 8). *Low-level* aerial imagery can either be obtained by nadir-mounted or oblique-looking cameras mounted on an Unmanned Aerial Vehicle (UAV) or out the window of a low-flying aircraft.

In principle, all image-based techniques use the following algorithm to calculate 3D points from the input images (see e.g. Remondino (2006b, Ch. 1)):

1. Image acquisition.
2. Detection and measurement of *feature points*, e.g. corners, in each image.
3. *Matching* of feature points between images, i.e. which 2D points correspond to the same 3D point?

[4] camera configurations with more than two cameras
[5] looking straight down

4. Calculation of the *relative orientation* between (pairs of) images, i.e. the relative position and orientation of the camera stations at the instants when the images were taken.

5. *Triangulation*, i.e. calculation of object point coordinates. This will generate a “cloud” of 3D points expressed in a local coordinate system.

6. *Co-registration* of multiple point clouds into a common, global, coordinate system (optional).

7. *Fine-tuning* of calculated object points and camera coordinates (optional).

8. *Point cloud densification*, i.e. measurements of more points (optional).

9. *Segmentation and structuring* of the point cloud, surface generation.

10. Extraction of *texture data*.

In addition to the above steps, *calibration* of each camera is required to obtain high-quality results. This can be performed separately or in conjunction with the point cloud processing.

If two cameras are fixed to a stereo rig, the rig itself can be calibrated. This corresponds to determining the relative orientation between the rig-mounted cameras. If this process is performed prior to step 4 of the algorithm above, the relative orientation problem reduces to calculating the relative orientation between successive image pairs.

4.4 Algorithms for subproblems

4.4.1 Camera calibration

The purpose of *camera calibration* is to calculate parameters internal to the camera. We distinguish between two different types of parameters; *linear* and *non-linear*. The most important linear parameter is the *(effective) focal length*, which is generally not the same value as the focal length written on the camera or stored in the image. The effect of the non-linear parameters is commonly called *lens distortion*, and has the effect that projections of straight lines are not straight (Figure 1). Most mathematics of photogrammetry and computer vision relies on that *no lens distortion is present*, or equivalently that the images or the measured coordinates are corrected for lens distortion. Such a corrected “camera” is said to be *straight-line-preserving* (see Figure 2). Lens distortion can only be ignored in low precision application or with cameras with very long focal lengths (>500mm).

Camera calibration is typically performed by taking multiple images of a calibration object, see Figure 3. For optimal results, camera calibration should be performed separate to the 3D reconstruction (Remondino and Fraser 2006). If that is not possible, the internal camera parameters may be estimated together with the object coordinates (“self-calibration” or “auto-calibration”) (Hartley

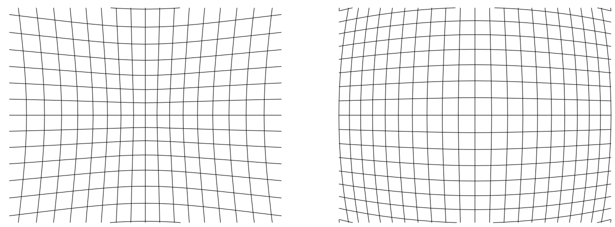

Figure 1: Lines straight in object space, bent by lens distortion. Left: pincushion distortion. Right: barrel distortion.

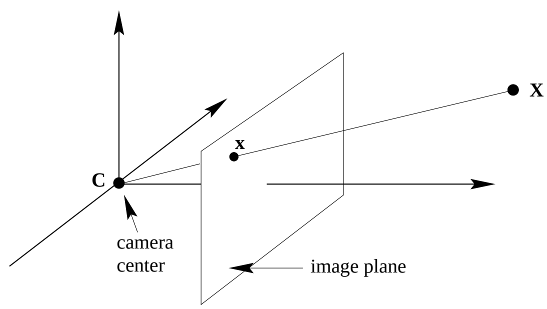

Figure 2: In a straight-line-preserving camera, the object point **X**, the camera center **C**, and the projected point **x** are *collinear*, i.e. on a straight line. The distance between the image plane and the camera center is known as the (effective) focal length. In this figure, the image plane is presented in front of the camera center instead.

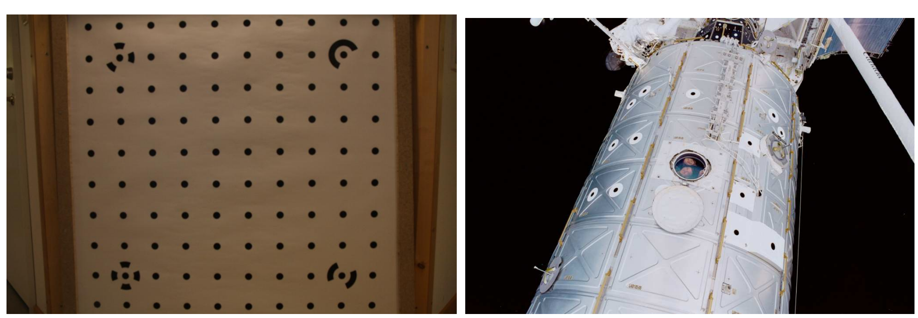

Figure 3: Left: A image of a calibration object with artificial targets (black circles). The targets have known three-dimensional coordinates. The code rings around four of the targets are used for identification. Right: Artificial targets attached to the outside of the Destiny lab attached to the International Space Station. Image credit: NASA.

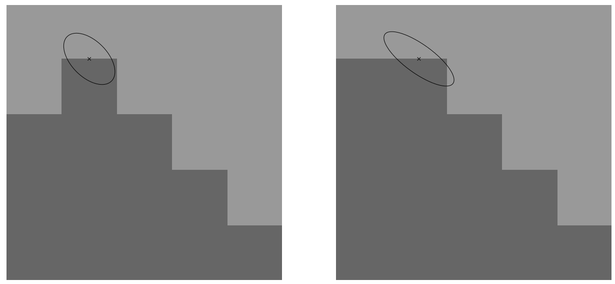

Figure 4: Two corners detected by the Förstner operator (Förstner and Gülch 1987) in synthetic images. The ellipses describe the uncertainty of each corner.

et al. 1992; Duan et al. 2008) or during the fine-tuning stage (Fraser 1997), at the cost of a reduced quality of the result.

In order to obtain useful 3D information, the camera calibration information has to be added at some stage of the reconstruction. Some algorithms only require the non-linear parameters to be known, i.e. that the cameras are straight-line-preserving (Devernay and Faugeras 2001).

4.4.2 Feature point detection

A *feature point* is a point or an area[6] of an image that is likely to be found and recognized in other images. Typical feature points are corners and circular features, although many textured areas will also be good feature points. In industrial applications, artificial *targets* are often added to a scene. These targets provide good feature points and are sometimes *coded* to aid automatic identification (Fraser and Cronk 2009), see Figure 3.

Most feature point detectors are automatic — they take an image as input and generates a list of 2D coordinates where feature points have been detected. Some detectors furthermore estimate the uncertainty of each 2D coordinate, see Figure 4. In addition, each feature point may be accompanied by a *descriptor* that describe the surrounding of the detected point, such as the size of the feature and the dominant direction within the region containing the feature, see Figure 5. The purpose of the descriptors is to enable matching of feature points detected in different images, i.e. to enable identification of the *same* 3D point viewed e.g. from different distances and/or directions.

In a comparison by Remondino (2006a), the methods by Förstner and Gülch (1987) and Heitger et al. (1992) had the highest precision of the detected 2D coordinates. Other common feature point detectors include the Harris detector (Harris and Stephens 1988), SUSAN (Smith and Brady 1997), the KLT tracker (Tomasi and Kanade 1991), and SIFT (Lowe 2004). The KLT tracker is especially common in videogrammetry.

[6] For simplicity, this report does not distinguish between *point detectors* and *region detectors*, found in some of the literature.

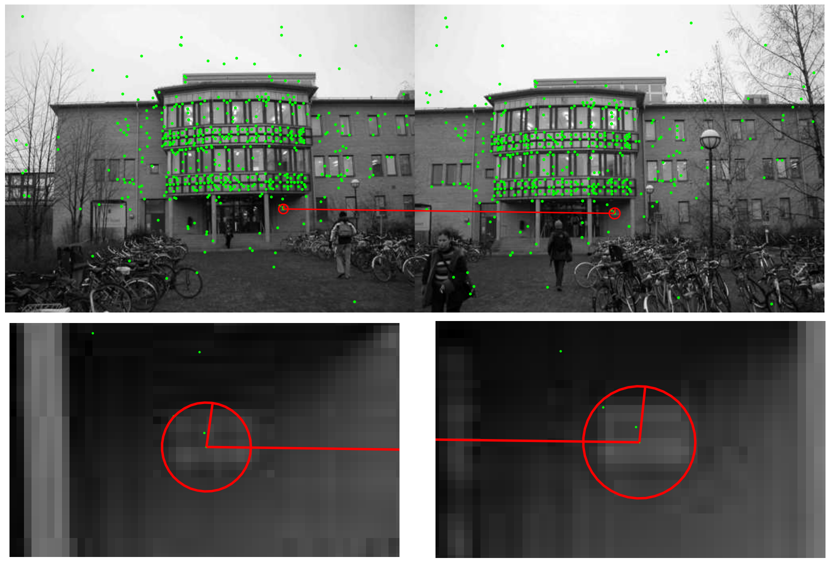

Figure 5: Top row: Feature points found with the SIFT detector (Lowe 2004) in two images of the same building. One match is highlighted. Bottom row: Zoom of the matched points in the images, indicating the size and dominant orientation of the feature, a sign on the wall.

4.4.3 Feature point matching

In order to extract 3D information from 2D images, a *correspondence* between points in different images must be established. This process is called *matching*. Feature points can be matched based on the image content around them or from the descriptors calculated by the feature point detector. Furthermore, if the relative orientation between two images is known, the matching can be restricted to *epipolar lines* (see Figure 6 (left)) rather than the whole image. Furthermore, if a third image is used, the matching ambiguities can be further reduced (Shashua 1997; Schaffalitzky and Zisserman 2002), see Figure 6 (right). Among the feature point detectors compared by Mikolajczyk and Schmid (2003), the SIFT descriptor (Lowe 2004) had the highest tolerance to changes in viewing geometry.

4.4.4 Combined feature point detection and matching

The Least Squares Template Matching (LSTM) technique performs the matching and precise location of the matches simultaneously. The basic algorithm compares patches between images while allowing a controlled geometric and radiometric deformation (Gruen 1985, 1996), see Figure 7. The LSTM algorithm is an iterative procedure that uses initial estimates of the match positions and other geometrical parameters. If the initial estimates are good and the image

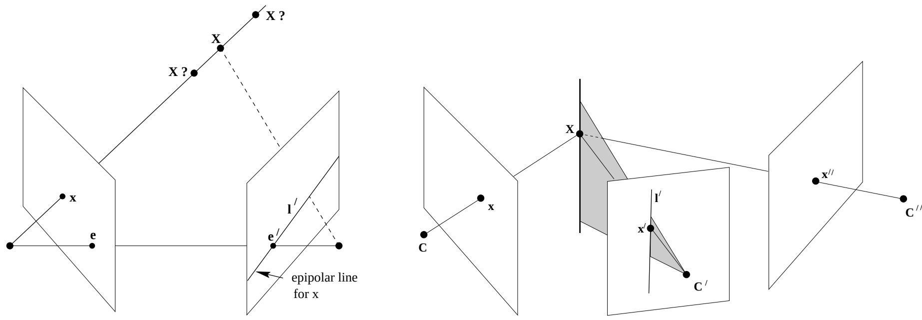

Figure 6: Left: An epipolar line restricts the search for corresponding points. Potential matches outside the epipolar line do not correspond to the same 3D point. Right: Adding a third image further reduces the ambiguity. A match found in along the epipolar line of image 2 must have a corresponding match in the third image.

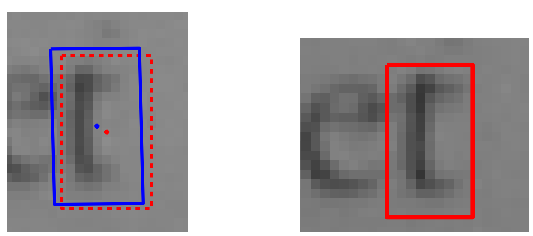

Figure 7: The patch (image region marked in the right image) is matched to the image to the left by Least Squares Template Matching (LSTM). Given the initial position (dashed red) in the left image, the algorithm updates the shape and position of the region (final value, solid blue). In this example, the region was allowed to deform by an affine transformation (shift, rotate, scale, and shear).

contains enough information (texture), the algorithm will converge to more precise matching coordinates, including error estimates. Otherwise, the algorithm may fail to converge, indicating a failed matching. LSTM can also use epipolar information (Baltsavias 1991), in which case the algorithm is known as Multi-Photo Constrained Template Matching. Furthermore, LSTM can be used to fine-tune the position of feature point matched by other detectors (Remondino and Ressl 2006), see Figure 8.

4.4.5 Relative orientation

The calculation of the *relative orientation* between different images is central to image-based techniques. A poor estimation of the relative orientation will affect the calculation of 3D object coordinates, both in terms of the object point precision and potential blunders due to incorrect matches.

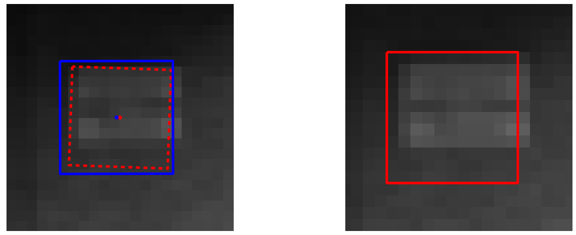

Figure 8: LSTM update of the SIFT match in Figure 5. The highlighted region in the right image is matched to the left image. The initial values from SIFT (dashed, red) are updated by LSTM (solid, blue).

Assuming that the correspondence problem has been solved, the relative orientation can be calculated from image point correspondences. The *eight-point algorithm* (Hartley 1997; Hartley and Zisserman 2003) is the most important algorithm for calculating the relative orientation for straight-line-preserving cameras. A recent important development is the *five-point algorithm* (Nistér 2004; Stewénius et al. 2006) that uses camera calibration information and is thus more stable.

If the correspondence between image points and *control points* (object points with known 3D coordinates) is known, the position and orientation of the image can be calculated for single images. This process is called *resection* (Grussenmeyer and Khalil 2002) . Two of the most important resection algorithms are the *Direct Linear Transformation (DLT)* (Abdel-Aziz and Karara 1971; Mikhail et al. 2001, Ch. 9.3) (straight-line-preserving) and Grunert (Haralick et al. 1994; McGlone et al. 2004, Ch. 11.1.3.4) (calibrated). Recent work on calibrated resection include Schweighofer and Pinz (2006); Olsson et al. (2009).

Automatic relative orientation is still subject to active research Läbe and Förstner (2006); Remondino and Ressl (2006); Läbe et al. (2008); Frahm et al. (2009), and challenges remain for especially low-textured images or wide-baseline images with little overlap and large differences in viewpoint.

4.4.6 Triangulation

The calculation of 3D coordinates for an object point given corresponding points with a known relative orientation is known as *triangulation* (Hartley and Zisserman 2003, Ch. 12.2; Förstner et al. 2004, Ch. 11.2). Beside the calculated 3D coordinate, an estimate of the *error* can also be produced (Förstner et al. 2004, Ch. 11.2.7; Heuel 2004; Förstner and Wrobel 2004, Ch. 2.3.5).

4.4.7 Fine-tuning (bundle adjustment)

Bundle adjustment is the process of simultaneously estimating object points and camera position and orientations. The process has been used in photogrammetry since the late 1950:ies (Brown 1976). After a seminal paper by Triggs et al.

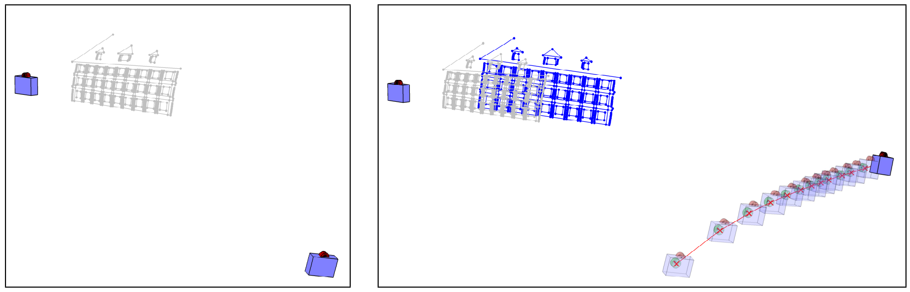

Figure 9: Bundle adjustment example with only two cameras, the leftmost of which is kept stationary. Left: The initial approximations of the object points (gray) and rightmost camera. Right: The final values of the object points (blue) and rightmost camera. The camera positions during the iteration process is also indicated.

(2000), it is now widely accepted within the Computer Vision community as well.

The process is iterative and require initial approximation from e.g. relative orientation and triangulation (Mikhail et al. 2001, Ch. 5.8). The initial approximation are iteratively modified until the projection of the estimated object points into the cameras at the estimated positions matches the measured image coordinates as closely as possible, see Figure 9. Besides the updated estimates of the object points and camera positions and orientations, error estimates of all estimated parameters are also produced (Förstner and Wrobel 2004, Ch. 2.2.5). Bundle adjustment is necessary in order to get high-quality results. At the same time, the method is general, and can handle any number of cameras and object points as well as other 3D geometric objects and scene constraints, e.g. that two estimated planes should be orthogonal.

Recent development in bundle adjustment include Börlin et al. (2004); Lourakis and Argyros (2005); Dickscheid et al. (2008); Lourakis and Argyros (2009).

4.4.8 Densification of the point cloud

The point cloud generated by general feature point matching is generally sparse and unevenly distributed over the images. However, once the relative orientation between two or more images is known, it is possible to use guided matching to densify the point cloud. Several "dense stereo" algorithms have been developed. The basic algorithm matches intensity variations along scan-lines of image pairs. Continuity constraints enable calculation of almost one depth value per pixel in well-textured image areas (Scharstein and Szeliski 2002; Yoon and Kweon 2008). Extension algorithms work on three or more images (Gallup et al. 2007; Seitz et al. 2006). Most algorithms are focused toward short-baseline input, i.e. situations where the camera motion between images is small, such as e.g. with video data. However, the algorithms work comparably well on wide-baseline images if the imaged regions are almost planar. Many algorithms focus on speed and report only qualitative accuracy results. Some recent exceptions are Seitz et al. (2006); Mordohai et al. (2007); Zhang et al. (2009).

The Least Squares Template Matching (LSTM) technique (Gruen 1985; Baltsavias 1991) can also be used for densification. Compared to "dense stereo" methods, LSTM generates less dense point clouds. However, an advantage is that bad matches can be detected and excluded and precision values can be calculated for each generated match (Grün et al. 2004; Remondino and Zhang 2006; Remondino et al. 2008).

4.4.9 Co-registration of point clouds

Image-based methods generally generate a point cloud in the same coordinate system, since new images are added sequentially and the relative orientation is determined from feature point matches with the existing images. However, if two point clouds are generated without any common images, or from other sensors such as laser scanners, the point clouds need to be *co-registered*, i.e. transformed into the same coordinate system.

If three or more point correspondences between the clouds are known, the transformation is a simple rigid-body transformation (Arun et al. 1987; Förstner et al. 2004, Ch. 11.1.6). For optimal robustness, point correspondences between different clouds can be determined from artificial targets in the scene (Akca 2003) or by manual assignment.

Without point correspondences, the co-registration problem is hard. Several algorithms have been presented in the literature, either based on point-point (or line-line) correspondences (Besl and McKay 1992; Früh and Zakhor 2004; Barnea and Filin 2008; Boström et al. 2008; Brenner et al. 2008; Stamos et al. 2008; González-Aguilera et al. 2009) or surface matching (Gruen and Akca 2005; Pottmann et al. 2006; Akca 2007; Bae and Lichti 2008). Other techniques are based on global features of the point cloud (Johnson and Hebert 1999; Huber and Hebert 2003; Bucksch and Lindenbergh 2008) or by matching with CAD models (Rabbani and van den Heuvel 2004; Rabbani et al. 2007). Methods for matching point clouds detected by hybrid camera-laser scanner hardware have also been developed (Wendt 2007; Smith et al. 2008). In order to work well, the algorithms for unorganized point clouds require good initial values (unless known from the hardware) or require good global coverage of the scene.

4.4.10 Object extraction and model generation

If the points have been labelled in previous stages, automatically or manually, object extraction usually consists of fitting primitives, e.g. planar surfaces or edges, to the labelled points (Debevec et al. 1996; Gruen and Wang 1998; Koch 2005; Yang et al. 2009). Many automatic algorithms have been developed to detect simple planar surfaces in unlabelled terrestrial and aerial laser scanner data (Rottensteiner 2003; Yu et al. 2008; Tarsha-Kurdi et al. 2008). Other algorithms search for known complex geometric objects, e.g. CAD models, in point clouds generated from images or laser scanner data (Rabbani and van den

Heuvel 2004; Ferrari et al. 2006; Rabbani et al. 2007; Leibe et al. 2008). An approach that does not try to group points into geometric objects is to generate a 3-D mesh directly from the point cloud (Früh and Zakhor 2002; Akbarzadeh et al. 2006; Gallup et al. 2007; Mordohai et al. 2007; Pollefeys et al. 2008).

Beside the actual measurements, it is also possible to add *scene constraints*, e.g. that two surfaces should be orthogonal, to aid the model generation (Debevec 1996; El-Hakim 2002).

4.4.11 Texture extraction

Once the camera positions and the 3D geometry of an object has been calculated, determining what part of an image that can be used as a texture map on each surface is trivial. However, in order to generate a convincingly looking textured 3D model, some practical aspects need to be considered, e.g. occlusion (Debevec et al. 1998; Ortin and Remondino 2005) and differences in lighting (Kim and Pollefeys 2008; Troccoli and Allen 2008). Furthermore, several papers have been presented that try to analyze repeated patterns in textures (Zalesny et al. 2005; Mayer and Reznik 2007; Müller et al. 2007; Wenzel and Förstner 2008). However, the problem in the general case is still unsolved.

4.4.12 Panoramic image stitching

If images are acquired by a rotating camera, no 3-D information may be inferred (Remondino and Börlin 2004). However, it is still possible to “stitch” the images together to form a panorama as described by e.g. Brown and Lowe (2007).

4.5 Reconstruction approaches and the type of input data

This section reports some of the major high-level approaches presented in the literature. The choice of approach is strongly correlated to the type of input data, and this section is thus structured by the type of input data.

4.5.1 Video-based reconstruction

The video-based approaches, exemplified by Akbarzadeh et al. (2006); Mordohai et al. (2007); Pollefeys et al. (2008); Frahm et al. (2009) are characterized by a high level of automation based on video input. The automation is successful mainly due to the small image distortion between consecutive video frames. No specific assumptions are made about the reconstructed environment and the result is usually in the form of textured 3D meshes. Several methods are able to generate automatic results in real-time, albeit without error estimates.

4.5.2 Reconstruction from aerial/satellite imagery

Methods based on aerial and/or satellite imagery are characterized by a high level of automation (Zhang and Gruen 2006). The automation is mainly achieved due to the regularity of the images capturing process. The output is usually in the form of a 2.5D Digital Surface Map (DSM), including error estimates. A 2.5D DSM is a 2D grid of positions that describes the elevation above a vertical datum, such as mean sea level. A DSM typically cover a large area and is useful for generating topographic maps. However, due to its low resolution and that each position only has one height value associated with it, a DSM does not contain any information about e.g. building facades.

4.5.3 Reconstruction from laser scanner data

Methods based on laser scanner data, either aerial (LIDAR) or terrestrial (TLS), focus on analyzing huge point clouds (Rottensteiner 2003; Tarscha-Kurdi et al. 2007; Barnea and Filin 2008; Tarsha-Kurdi et al. 2008), mainly to detect common points for point cloud co-registration and planar surfaces for building detection. The result is usually with error estimates but without textures.

4.5.4 Image-based reconstruction

There is a multitude of papers about image-based 3D reconstruction in the literature. Some example applications include: city model generation from aerial images (Gruen and Wang 1998), modeling of *Arc de Triomphe* in Paris from tourist images (El-Hakim 2002), reconstruction of buildings using scene constraints (Debevec 2003), high-resolution reconstruction of cultural heritage objects (The Standing Buddhas of Bamyan) from multi-resolution images (Grün et al. 2004; Remondino and Niederoest 2004), and a fully automatic measurements for industrial applications (Fraser and Cronk 2009). As stated by Remondino and El-Hakim (2006, abstract), "... image-based modelling ... remains the most complete, economical, portable, flexible and widely used approach". However, no unified method exist that cover all reconstruction problem. Indeed, from the same paper

> ... there is no single modelling technique able to satisfy all requirements of high geometric accuracy, portability, full automation, photo-realism and low cost as well as flexibility and efficiency (Remondino and El-Hakim 2006, p. 272).

4.5.5 Combination of image and laser scanner data

Methods that use a combination of images and laser scanner data are also common. Fruh and Zakhor (2003), used aerial images and laser scans of part of a city to generate a textured DSM of the ground and rooftops. A laser-scanner and camera-equipped car was used to acquire terrestrial data. The terrestrial data was used to determine the path within the DSM driven by the car and to generate texture maps of the vertical building facades.

Rabbani and van den Heuvel (2004) used laser scans to reconstruct an industrial site. Multiple scans were used to model the bulk of the site. Images were taken to complement the laser scan data in regions that were hard to reach for the laser scanner due to the crowded scene. The image data was used to aid interpretation of the point cloud and to provide complementary measurements of edges within the scene.

In an opposite approach, Gonzo et al. (2004) used aerial images and tacheometric data to generate an overall model of an Italian castle. Detail was added to the model from terrestrial images and laser scans. The tacheometric and laser scans were considered optional in their setup, i.e. they suggest that image-only methods would be viable.

Several papers have compared tacheometry, laser scanning, and photogrammetry to reconstruct complicated buildings, e.g. castles (Remondino et al. 2005; Landes et al. 2007; Grussenmeyer et al. 2008). The consensus is that either technique has its strength and weaknesses, but that they complement each other well. However, a comparative paper by Strecha et al. (2008) challenges the consensus by posing the question on whether image-based methods can completely replace close-range laser scanning.

In a recent paper, Remondino et al. (2009) suggests the following workflow for modeling of complicated architectures:

1. Use surveying to obtain a high-accuracy reference grid.
2. Acquire low-level oblique aerial images to model the majority of the model.
3. Take terrestrial images to for complementary modelling and detailed modelling of parts of the scene.
4. Use medium-range (1–50m) laser scanner of interiors less suited for image-based modeling.

4.6 Automatic vs. semi-automatic reconstruction

A 3-D reconstruction application should ideally satisfy the following requirements (El-Hakim 2002):

1. High geometric accuracy.
2. Capturing all details.
3. Photo-realism.
4. Full automation.
5. Low cost.
6. Portability.
7. Flexibility in applications.
8. Efficiency in model size.

Thus, automation is a highly wanted requirement. However, the only available fully automated techniques either generate incomplete textured meshes without any error information (Akbarzadeh et al. 2006; Mordohai et al. 2007; Pollefeys et al. 2008; Frahm et al. 2009) or require a substantial modification of the scene (Fraser and Cronk 2009). Indeed, many authors suggest that semi-automated methods are the most efficient, and e.g. Mayer (2008, p. 217) concludes that: "Key factors determining the practical usefulness of a system are *thorough testing* as well as an *optimized user interaction*" (our emphasis).

5 "Reconstruction" software

In this section, some software related to 3-D reconstruction from images are described. Evaluated software is listed in sections 5.1–5.3.

5.1 Google Sketchup

Google Sketchup[7] is a free tool for constructing 3D models that can be uploaded to the Internet and viewed by Google Earth. It can generate textured models but does not perform any measurements from images.

5.2 Microsoft Photosynth

Microsoft Photosynth[8] is a free[9] tool for organizing images of a scene. The application uses SIFT (Lowe 2004) features to determine the relative orientation, followed by bundle adjustment with restricted camera self-calibration to improve the estimated 3D coordinates and camera positions. The camera self-calibration is restricted to the focal length and two radial lens distortion parameters. For details, see Snavely et al. (2006, 2008).

Presently (June 2009) it is not possible to view the result on your local machine. Instead, it is necessary to upload the result to the Photosynth web site, where it is automatically made publicly available. The result is presented as a virtual 3D world, where it is possible to change the viewpoint between different calculated camera coordinates. Snavely et al. (2008, p. 191) writes

> ...our objective is *not* to synthesize a photo-realistic view of the world from all viewpoints *per se*, but to browse a specific collection of photographs in a 3D spatial context that gives a *sense* of the geometry of the underlying scene

(their emphasis) and as of this writing it is not possible to extract 3D information from the generated result.

[7] http://sketchup.google.com
[8] http://photosynth.net
[9] However, a Microsoft Live ID is required.

5.3 Photomodeler

Photomodeler[10] is a photogrammetric software from EOS Systems, Inc. The basic software is based on manual measurements, although an automatic camera calibration component is included. Add-ons include automatic measurement of coded targets and dense matching. The software can export 3-D textured models as well as the raw 3-D data, including positions of object points and cameras. Furthermore, uncertainty estimates for the calculated positions are also available. The list price (June 2009) for the software is between USD 1000 (basic version) and USD 2700 (complete version).

5.4 ShapeCapture/ShapeScan

ShapeCapture and ShapeScan are two software from ShapeQuest Inc[11]. According to the ShapeQuest homepage, ShapeCapture provides the capability for 3D Modeling from images, Camera Calibration, Accurate 3D Coordinate Measurement, Photogrammetry, Texture Mapping, Automatic Target Extraction and Stereo Matching. Furthermore, the ShapeScan software is be able to work with both images and point clouds acquired by a laser scanner. However, ShapeCapture/ShapeScan are unable to measure stored parametric models of e.g. buildings. The list prices (June 2009) for ShapeCapture and ShapeScan are USD 1600 and USD 8000, respectively.

5.5 ImageModeler

Autodesk ImageModeler is a reconstruction software sold by Autodesk Inc[12]. According to the Autodesk homepage, ImageModeler is able to reconstruct photo-realistic objects, scenes and cities from images as well as take measurements of buildings. The software can export the results in FBX, Maya, OBJ, and DWG formats. However, the FAQ[13] does not mention anything about extracting the precision of the calculated values. The list price (June 2009) of the software is USD 1500.

5.6 Other photogrammetric software

Other photogrammetric software include Australis[14] and iWitness PRO[14], V-STARS[15], and DPA-Pro[16] oriented mainly at industrial applications.

---

[10] http://www.photomodeler.com
[11] http://www.shapecapture.com
[12] http://www.autodesk.com/imagemodeler
[13] http://images.autodesk.com/adsk/files/imagemodeler2009_faq0.pdf
[14] http://www.photometrix.com.au
[15] http://www.geodetic.com
[16] http://www.aicon.de

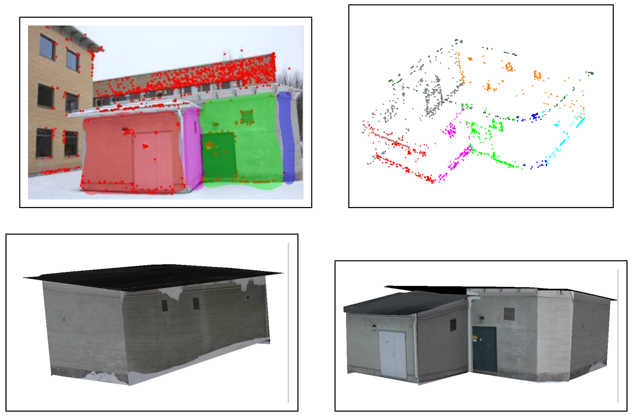

Figure 10: Reconstruction results from Fors Nilsson and Grundberg (2009). Top left: Manual "painting" of image regions corresponding to different object surfaces. Top right: Resulting point cloud labelling. Bottom row: Two textured views of the reconstructed result.

6 Proof of concept

A proof of concept for 3D measurements of buildings from image data is presented in the work by Fors Nilsson and Grundberg (2009). The reconstruction is based on planar surfaces and uses calibrated cameras. Initially, the acquired images are ordered sequentially by the operator. Feature detection and matching is performed automatically between consecutive images and the relative orientation and triangulation is calculated using robust methods. The triangulation result is presented to the operator for quality control. Adequate results are added to the sequential registration process. Otherwise, the process is re-run with different parameters. At any stage, bundle adjustment may be run to fine-tune the reconstructed data.

Object measurements is also semi-automatic. The operator marks image parts that correspond to the same planar object surface with a "paint" tool. Labelling of the reconstructed point cloud is automatically inferred from the images, and the planes are calculated from the labelled points. The result is presented to the user that has the possibility to repeat the calculation after removing mis-labelled or wrongly calculated points. Finally, several planes are combined into a polyhedral object. See Figure 10 for an example.

In agreement with Mayer (2008), the report stresses the importance of optimizing the workflow for the user and that in some cases a too high degree of automation may be of more harm than good. As future work the authors suggest fitting points to other surface primitives than planes. Furthermore, reconstructed composite primitives could be stored in a "library" and later retrieved when a similarly composed object is to be reconstructed. Other suggestions include detail enhancement on facades with dense stereo methods and guided matching to increase the number of matched surface feature points.

7 Summary and discussion

7.1 Input data for harbor modelling

Some harbor scenes cover a substantial area, and conventional aerial methods carry the advantage of being able to cover large areas quickly. However, aerial data also carry a substantial cost. Furthermore, it is believed that on some sites, obtaining the proper permissions may be difficult or even impossible. Obtaining low-level aerial data from e.g. a UAV (unmanned aerial vehicle) could be easier and cheaper, but the maturity of the technology is unclear. However, an interesting possibility is that the harbor cranes provide elevated vantage points that may be utilized instead of low-level aerial images.

Instead, we argue that we should focus on terrestrial images and possibly laser scanner data. A possible solution is to model the general area from terrestrial image data. The work areas could either be modelled from image data or from short-range laser scanner data. An interesting prospect would be to use high-resolution images acquired by a calibrated stereo rig, something that increases the robustness of e.g. relative orientation and triangulation.

7.2 Software

There is no software that satisfy all our requirements. Of the investigated software, ShapeCapture/ShapeScan (Section 5.4) seem to be the best candidate for the harbor reconstruction problem. However, the software lacks the possibility to construct parameterized 3D models of e.g. buildings that can later be retrieved and used to speed up measurements of repeated structures. Furthermore, there is a question mark on the quality of the dense point clouds generated by ShapeScan from image data.

7.3 Potential research areas

Among the identified potential research areas within this field are

- Camera calibration, especially stereo rig calibration.
- Bundle adjustment, especially bundle adjustment with scene constraints.
- Optimization of camera networks.
- 3D reconstruction for physical simulation.

7.4 The 3D reconstruction toolbox

We suggest that a *toolbox* is implemented with algorithms for 3D reconstruction. The main data source should be terrestrial images and laser scan data and optionally tacheometry data. State-of-the-art automatic low-level algorithms for well-understood subproblems should be implemented “as is”. However, algorithm without error estimates should be avoided unless they can be augmented to include error estimates. Higher-level algorithms, automatic or semi-automatic, can be combined from the low-level algorithms and evaluated to find efficient measurement workflows.

The toolbox is suggested to be implemented in a high-level language, Matlab[17]. Matlab supports efficient implementation of many automatic 3D reconstruction algorithms. Furthermore, it has enough support for GUI implementations to enable the necessary efficiency evaluations. Additionally, non-GUI-algorithm can be ported with relative ease to efficient implementations in low-level languages such as C/C++. More toolbox details are presented in Appendix C.

It is our belief that a 3D reconstruction toolbox in Matlab will be a flexible tool since it can simultaneously support basic research of core algorithms, evaluation of problem-specific high-level algorithms, and production of industry-grade results that can be spawned and ported to other environments. The development will be incremental, and by using the toolbox in different 3D reconstruction projects it is possible to learn from different cases.

A potential application of the toolbox on the harbor reconstruction problem would be to model the ground as a 3-D topographic mesh and reconstruct it from image data and possibly laser scanner data. Ground texture data would automatically be extracted from the images. Distant landmarks with restricted image coverage could be modelled using scene constraints or stored composite primitives as suggested in Section 6. Composite primitives could also be used for obstacle modelling and modelling of active objects. Whether the geometric quality of the reconstructed models is sufficient for modelling of active objects remains an open question and will be interesting to investigate. The horizon could finally be reconstructed from image data acquired from e.g. the harbor cranes using panoramic image stitching as described in Section 4.4.12.

[17]http://www.mathworks.com

References

Abdel-Aziz, Y. I. and Karara, H. M. (1971). Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. In *Proceedings of ASP Symposium on Close-range Photogrammetry*, pages 1–18, University Illinois at Urbana-Champaign, Urbana, IL.

Akbarzadeh, A., Frahm, J.-M., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Merrell, P., Phelps, M., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewénius, H., Yang, R., Welch, G., Towles, H., Nistér, D., and Pollefeys, M. (2006). Towards urban 3d reconstruction from video. In *Proc. 3DPVT'06*, pages 1–8, Chapel Hill, North Carolina, USA. IEEE.

Akca, D. (2003). Full automatic registration of laser scanner point clouds. In *Proc. of Optical 3-D Measurement Techniques VI*, volume I, pages 330–337, Zurich, Switzerland. ISPRS.

Akca, D. (2007). Matching of 3d surfaces and their intensities. *ISPRS J Photogramm*, 62(2):112 – 121.

Akca, D. and Gruen, A. (2007). Generalized least squares multiple 3d surface matching. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVI(3/W52):1–7.

Arun, K. S., Huang, T. S., and Blodstein, S. D. (1987). Least-squares fitting of two 3-d point sets. *IEEE T Pattern Anal*, 9(5):698–700.

Bae, K.-H. and Lichti, D. D. (2008). A method for automated registration of unorganised point clouds. *ISPRS J Photogramm*, 63(1):36 – 54.

Baltsavias, E. P. (1991). *Multiphoto Geometrically Constrained Matching*. PhD thesis, Institute of Geodesy and Photogrammetry, ETH, Zürich, Switzerland.

Barnea, S. and Filin, S. (2008). Keypoint based autonomous registration of terrestrial laser point-clouds. *ISPRS J Photogramm*, 63(1):19 – 35.

Besl, P. J. and McKay, N. D. (1992). A method for registration of 3-d shapes. *IEEE T Pattern Anal*, 14(2):239–256.

Börlin, N., Grussenmeyer, P., Eriksson, J., and Lindström, P. (2004). Pros and cons of constrained and unconstrained formulation of the bundle adjustment problem. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXV(B3):589–594.

Boström, G., Gonçalves, J. G., and Sequeira, V. (2008). Controlled 3d data fusion using error-bounds. *ISPRS J Photogramm*, 63(1):55 – 67.

Brenner, C., Dold, C., and Ripperda, N. (2008). Coarse orientation of terrestrial laser scans in urban environments. *ISPRS J Photogramm*, 63(1):4 – 18.

Brown, D. C. (1976). The bundle adjustment — progress and prospects. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 21(3):33 pp.

Brown, M. and Lowe, D. G. (2007). Automatic panoramic image stitching using invariant features. *Int J Comp Vis*, 74(1):59–73.

Bucksch, A. and Lindenbergh, R. (2008). CAMPINO — a skeletonization method for point cloud processing. *ISPRS J Photogramm*, 63(1):115 – 127.

Debevec, P. E. (1996). *Modeling and Rendering Architecture from Photographs.* PhD thesis, University of California at Berkeley.

Debevec, P. E. (2003). Image-based techniques for digitizing environments and artifacts. In *Proceedings of the 4th International Conference on 3D Digital Imaging and Modeling (3DIM 2003)*, pages 234–242, Banff, Canada. IEEE. Invited paper.

Debevec, P. E., Taylor, C. J., and Malik, J. (1996). Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. *Proceedings of SIGGRAPH 96*, pages 11–20.

Debevec, P. E., Yu, Y., and Borshukov, G. D. (1998). Efficient view-dependent image-based rendering with projective texture-mapping. In Drettakis, G. and Max, N., editors, *Proc. Eurographics Rendering Workshop*, pages 105–116, Viennea, Austria.

Devernay, F. and Faugeras, O. (2001). Straight lines have to be straight. *Mach Vision Appl*, 13(1):14–24.

Dickscheid, T., Läbe, T., and Förstner, W. (2008). Benchmarking automatic bundle adjustment results. In *21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, volume B3a, pages 7–12, Beijing, China.

Duan, C., Meng, X., and Wang, L. (2008). 3d reconstruction from uncalibrated images taken from widely separated views. In *Cybernetics and Intelligent Systems, 2008 IEEE Conference on*, pages 58–62.

El-Hakim, S. (2002). Semi-automatic 3d reconstruction of occluded and unmarked surfaces from widely separated views. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 34(5):143–148.

Ferrari, V., Tuytelaars, T., and Gool, L. V. (2006). Simultaneous object recognition and segmentation from single or multiple model views. *Int J Comp Vis*, 67(2):159–188.

Fors Nilsson, H. and Grundberg, D. (2009). Plane-based close range photogrammetric reconstruction of buildings. Master's thesis, Department of Computing Science, Umeå University.

Förstner, W. and Gülch, E. (1987). A fast operator for detection and precise location of distinct points, corners and circular features. In *Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken.

Förstner, W. and Wrobel, B. (2004). *Mathematical Concepts in Photogrammetry*, chapter 2, pages 15–180. IAPRS, 5 edition.

Förstner, W., Wrobel, B., Paderes, F., Craig, R., Fraser, C., and Dolloff, J. (2004). *Analytical Photogrammetric Operations*, chapter 11, pages 763–948. IAPRS, 5 edition.

Frahm, J.-M., Pollefeys, M., Clipp, B., Gallup, D., Raguram, R., Wu, C., and Zach, C. (2009). 3d reconstruction of architectural scenes from uncalibrated video sequences. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVIII(5/W1):7 pp.

Fraser, C. S. (1997). Digital camera self-calibration. *ISPRS J Photogramm*, 52(4):149–159.

Fraser, C. S. and Cronk, S. (2009). A hybrid measurement approach for close-range photogrammetry. *ISPRS J Photogramm*, 64(3):328 – 333.

Früh, C. and Zakhor, A. (2002). Data processing algorithms for generating textured 3d building façade meshes from laser scans and camera images. In *1st International Symposium on 3D Data Processing Visualization and Transmission (3DPVT 2002)*, pages 834–849, Padova, Italy. IEEE Computer Society.

Fruh, C. and Zakhor, A. (2003). Constructing 3d city models by merging aerial and ground views. *IEEE Comput Graphics Appl*, 23(6):52–61.

Früh, C. and Zakhor, A. (2004). An automated method for large-scale, ground-based city model acquisition. *Int J Comp Vis*, 60(1):5–24.

Gallup, D., Frahm, J.-M., Mordohai, P., Yang, Q., and Pollefeys, M. (2007). Real-time plane-sweeping stereo with multiple sweeping directions. In *Proc. CVPR*, pages 1–8, Minneapolis, Minnesota, USA. IEEE.

González-Aguilera, D., Rodríguez-Gonzálvez, P., and Gómez-Lahoz, J. (2009). An automatic procedure for co-registration of terrestrial laser scanners and digital cameras. *ISPRS J Photogramm*, 64(3):308 – 316.

Gonzo, L., El-Hakim, S., Picard, M., Girardi, S., and Whiting, E. (2004). Photo-realistic 3-d reconstruction of castles with multiple- sources image-based techniques. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 35(B5):120–125.

Gruen, A. (1996). Least squares matching: a fundamental measurement algorithm. In Atkinson, K. B., editor, *Close Range Photogrammetry and Machine Vision*, chapter 8, pages 217–255. Whittles, Caithness, Scotland.

Gruen, A. and Akca, D. (2005). Least squares 3d surface and curve matching. *ISPRS J Photogramm*, 59(3):151 – 174.

Gruen, A. and Wang, X. (1998). Cc-modeler: a topology generator for 3-d city models. *ISPRS J Photogramm*, 53(5):286 – 295.

Gruen, A. W. (1985). Adaptive least squares correlation: A powerful image matching technique. *S Afr J of Photogrammetry*, 14(3):175–187.

Grün, A., Remondino, F., and Zhang, L. (2004). Photogrammetric reconstruction of the great buddha of bamiyan, afghanistan. *Photogramm Rec*, 19(107):177–199.

Grussenmeyer, P. and Khalil, O. A. (2002). Solutions for exterior orientation in photogrammetry: A review. *Photogramm Rec*, 17:615–634.

Grussenmeyer, P., Landes, T., Voegtle, T., and Ringle, K. (2008). Comparison methods of terrestrial laser scanning, photogrammetry and tacheometry data for recording of cultural heritage buildings. In *21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, volume B5, pages 213–218, Beijing, China.

Haralick, R. M., Lee, C.-N., Ottenberg, K., and Nölle, M. (1994). Review and analysis of solutions of the three point perspective pose estimation problem. *Int J Comp Vis*, 13(3):331–356.

Harris, C. J. and Stephens, M. (1988). A combined corner and edge detector. In *4th Alvey Vision Conference*, pages 147–151, Manchester.

Hartley, R. I. (1997). In defense of the eight-point algorithm. *IEEE T Pattern Anal*, 19(6):580–593.

Hartley, R. I., Gupta, R., and Chang, T. (1992). Stereo from uncalibrated cameras. In *CVPR'1992*, pages 761–764.

Hartley, R. I. and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, 2nd edition.

Heitger, F., Rosenthalter, L., von der Heydt, R., Peterhans, E., and Kuebler, O. (1992). Simulation of neuronal contour mechanisms: from simple to end-stopped cells. *Vision Res*, 32(5):963–981.

Heuel, S. (2004). *Uncertain Projective Geometry: Statistical Reasoning for Polyhedral Object Reconstruction*. Number 3008 in Lecture Notes in Computer Science. Springer, Berlin.

Huber, D. F. and Hebert, M. (2003). Fully automatic registration of multiple 3d data sets. *Image Vis Comput*, 21(7):637 – 650.

Johnson, A. and Hebert, M. (1999). Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE T Pattern Anal*, 21(5):433–449.

Kim, S. J. and Pollefeys, M. (2008). Robust radiometric calibration and vignetting correction. *IEEE T Pattern Anal*, 30(4):562–576.

Koch, R. (2005). 3-d surface reconstruction from stereoscopic image sequences. In *Proceedings of IEEE International Conference on Computer Vision ICCV'05*, pages 109–114, Beijing, China. IEEE.

Läbe, T., Dickscheid, T., and Förstner, W. (2008). On the quality of automatic relative orientation procedures. In *21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, volume B3b-1, pages 37–42, Beijing, China.

Läbe, T. and Förstner, W. (2006). Automatic relative orientation of images. In *Proc. of the 5th Turkish-German Joint Geodetic Days*, page 6 pp, Berlin, Germany.

Landes, T., Grussenmeyer, P., Voegtle, T., and Ringle, K. (2007). Combination of terrestrial recording techniques for 3d object modelling regarding topographic constraints. example of the castle of haut-andlau, alsace, france. In *Proceedings of XXI Intl CIPA Symposium*, page 6 pp, Athens, Greece. CIPA.

Leibe, B., Leonardis, A., and Schiele, B. (2008). Robust object detection with interleaved categorization and segmentation. *Int J Comp Vis*, 77(1-3):259–289.

Lourakis, M. I. A. and Argyros, A. A. (2005). Is Levenberg-Marquardt the most efficient optimization algorithm for implementing bundle adjustment? In *Proceedings of IEEE International Conference on Computer Vision ICCV'05*, volume 2, pages 1526–1531, Beijing, China. IEEE.

Lourakis, M. I. A. and Argyros, A. A. (2009). Sba: A software package for generic sparse bundle adjustment. *ACM TOMS*, 36(1):30 pp.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int J Comp Vis*, 60(2):91–110.

Mayer, H. (2008). Object extraction in photogrammetric computer vision. *ISPRS J Photogramm*, 63(2):213 – 222.

Mayer, H. and Reznik, S. (2007). Building facade interpretation from uncalibrated wide-baseline image sequences. *ISPRS J Photogramm*, 61(6):371 – 380.

McGlone, C., Mikhail, E., and Bethel, J., editors (2004). *Manual of Photogrammetry*. ASPRS, 5th edition.

Mikhail, E. M., Bethel, J. S., and McGlone, J. C. (2001). *Introduction to Modern Photogrammetry*. Wiley.

Mikolajczyk, K. and Schmid, C. (2003). A performance evaluation of local descriptors. In *Computer Vision and Pattern Recognition*, volume 2, pages 257–263, Madison, WI, USA. IEEE Computer Society.

Mordohai, P., Frahm, J.-M., Akbarzadeh, A., Clipp, B., Engels, C., Gallup, D., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewénius, H., Towles, H., Welch, G., Yang, R., Pollefeys, M., and Nistér, D. (2007). Real-time video-based reconstruction of urban environments. In *Proc. 3D-ARCH'2007*, Zürich, Switzerland. ISPRS.

Müller, P., Zeng, G., Wonka, P., and Gool, L. V. (2007). Image-based procedural modeling of facades. *ACM TOG*, 26(3):85.

Nistér, D. (2004). An efficient solution to the five-point relative pose problem. *IEEE T Pattern Anal*, 26(6):756–770.

Olsson, C., Kahl, F., and Oskarsson, M. (2009). Branch-and-bound methods for euclidean registration problems. *IEEE T Pattern Anal*, 31(5):783–794.

Ortin, D. and Remondino, F. (2005). Occlusion-free image generation for realistic texture mapping. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVI(5/W17):7 pp.

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., and Towles, H. (2008). Detailed real-time urban 3d reconstruction from video. *Int J Comp Vis*, 78(2-3):143–167.

Pottmann, H., Huang, Q.-X., Yang, Y.-L., and Hu, S.-M. (2006). Geometry and convergence analysis of algorithms for registration of 3d shapes. *Int J Comp Vis*, 67(3):277–296.

Rabbani, T., Dijkman, S., van den Heuvel, F., and Vosselman, G. (2007). An integrated approach for modelling and global registration of point clouds. *ISPRS J Photogramm*, 61(6):355 – 370.

Rabbani, T. and van den Heuvel, F. (2004). 3d industrial reconstruction by fitting csg models to a combination of images and point clouds. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 35(B5):7–12.

Remondino, F. (2006a). Detectors and descriptors for photogrammetric applications. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVI(3):49–54.

Remondino, F. (2006b). *Image-based modeling for object and human reconstruction.* PhD thesis, Institute of Geodesy and Photogrammetry, ETH Zürich, ETH Hoenggerberg, Zürich, Swizerland.

Remondino, F. and Börlin, N. (2004). Photogrammetric calibration of image sequences acquired with a rotating camera. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXIV(5/W16). Panoramic Photogrammetry Workshop, Dresden, Germany.

Remondino, F. and El-Hakim, S. (2006). Image-based 3D modelling: A review. *Photogramm Rec*, 21(115):269–291.

Remondino, F., El-Hakim, S., Girardi, S., Rizzi, A., Benedetti, S., and Gonzo, L. (2009). 3d virtual reconstruction and visualization of complex architectures - the 3d-arch project. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVIII(5/W1):9 pp.

Remondino, F., El-Hakim, S. F., Gruen, A., and Zhang, L. (2008). Development and performance analysis of image matching for detailed surface reconstruction of heritage objects. *IEEE Signal Proc Mag*, 25(4):55–64.

Remondino, F. and Fraser, C. (2006). Digital camera calibration methods: Considerations and comparisons. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVI(5):266–272.

Remondino, F., Guarnieri, A., and Vettore, A. (2005). 3d modeling of close-range objects: Photogrammetry or laser scanning? In *Proc. SPIE-IS&T Electronic Imaging: Videometrics VIII*, volume 5665, pages 216–225, San Jose, California.

Remondino, F. and Niederoest, J. (2004). Generation of high-resolution mosaic for photo-realistic texture-mapping of cultural heritage 3d models. In Cain, K., Chrysanthou, Y., Niccolucci, F., and Silberman, N., editors, *Proc. of 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST)*, pages 85–92, Brussels, Belgium.

Remondino, F. and Ressl, C. (2006). Overview and experiences in automated markerless image orientation. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVI(3):248–254.

Remondino, F. and Zhang, L. (2006). Surface reconstruction algorithms for detailed close-range object modeling. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVI(3):117–123.

Rottensteiner, F. (2003). Automatic generation of high-quality building models from lidar data. *IEEE Comput Graphics Appl*, 23(6):42–50.

Rottensteiner, F., Trinder, J., Clode, S., and Kubik, K. (2007). Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis. *ISPRS J Photogramm*, 62(2):135 – 149.

Schaffalitzky, F. and Zisserman, A. (2002). Multi-view matching for unordered image sets, or "how do i organise my holiday snaps?". In *ECCV'02*, volume 1, pages 414–431, Copenhagen, Denmark.

Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int J Comp Vis*, 47(1-3):7–42.

Schweighofer, G. and Pinz, A. (2006). Robust pose estimation from a planar target. *IEEE T Pattern Anal*, 28(12):2024–2030.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR'06*, volume 1, pages 519–528.

Shashua, A. (1997). Trilinear tensor: The fundamental construct of multiple-view geometry and its applications. In *Proc Int. Workshop on Algebraic Frames for the Perception Action Cycle (AFPAC)*, volume 1315 of *LNCS*, pages 190–206, Kiel, Germany. Springer.

Smith, E. R., King, B. J., Stewart, C. V., and Radke, R. J. (2008). Registration of combined range-intensity scans: Initialization through verification. *Comp Vis Imag Under*, 110(2):226 – 244.

Smith, S. M. and Brady, J. M. (1997). SUSAN — A new approach to low level image processing. *Int J Comp Vis*, 23(1):45–78.

Snavely, N., Seitz, S. M., and Szeliski, R. (2006). Photo tourism: Exploring photo collections in 3d. In *SIGGRAPH Conference Proceedings*, pages 835–846, New York, NY, USA. ACM Press.

Snavely, N., Seitz, S. M., and Szeliski, R. (2008). Modeling the world from internet photo collections. *Int J Comp Vis*, 80(2):189–210.

Stamos, I., Liu, L., Chen, C., Wolberg, G., Yu, G., and Zokai, S. (2008). Integrating automated range registration with multiview geometry for the photorealistic modeling of large-scale scenes. *Int J Comp Vis*, 78(2-3):237–260.

Stewénius, H., Engels, C., and Nistér, D. (2006). Recent developments on direct relative orientation. *ISPRS J Photogramm*, 60(4):284–294.

Strecha, C., von Hansen, W., Gool, L. V., Fua, P., and Thoennessen, U. (2008). On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *CVPR'08*, pages 1–8.

Tarscha-Kurdi, F., Landes, T., Grussenmeyer, P., and Koehl, M. (2007). Model-driven and data-driven approaches using lidar data: Analysis and comparison. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 36(3/W49A):6p.

Tarsha-Kurdi, F., Landes, T., and Grussenmeyer, P. (2008). Extended RANSAC algorithm for automatic detection of building roof planes from lidar data. *The Photogrammetric Journal of Finland*, 21(1):97–109.

Tomasi, C. and Kanade, T. (1991). Detection and tracking of point features. Technical Report CMU-CS-91-132, Carnegie Mellon University.

Triggs, B., McLauchlan, P., Hartley, R., and Fitzgibbon, A. (2000). Bundle adjustment — A modern synthesis. In *Vision Algorithms: Theory and Practice, Proceedings of the International Workshop on Vision Algorithms*, volume 1883 of *Lecture Notes in Computer Science*, pages 298–372. Spring Verlag.

Troccoli, A. and Allen, P. (2008). Building illumination coherent 3d models of large-scale outdoor scenes. *Int J Comp Vis*, 78(2-3):261–280.

Wendt, A. (2007). A concept for feature based data registration by simultaneous consideration of laser scanner data and photogrammetric images. *ISPRS J Photogramm*, 62(2):122 – 134.

Wenzel, S. and Förstner, W. (2008). Semi-supervised incremental learning of hierarchical appearance models. In *21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, volume B3b-2, pages 399–404, Beijing, China.

Yang, Q., Wang, L., Yang, R., Stewenius, H., and Nister, D. (2009). Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. *IEEE T Pattern Anal*, 31(3):492–504.

Yoon, K.-J. and Kweon, I. S. (2008). Distinctive similarity measure for stereo matching under point ambiguity. *Comp Vis Imag Under*, 112(2):173 – 183.

Yu, G., Grossberg, M., Wolberg, G., and Stamos, I. (2008). Think globally, cluster locally: A unified framework for range segmentation. In *Proceedings of 3DPVT'08 — the Fourth Internation Symposium on 3D Data Processing, Visualization and Transmissions*, page 8 pp, Atlanta, GA, USA.

Zalesny, A., Ferrari, V., Caenen, G., and Gool, L. V. (2005). Composite texture synthesis. *Int J Comp Vis*, 62(1/2):161–176.

Zhang, G., Jia, J., Wong, T.-T., and Bao, H. (2009). Consistent depth maps recovery from a video sequence. *IEEE T Pattern Anal*, 31(6):974–988.

Zhang, L. and Gruen, A. (2006). Multi-image matching for dsm generation from ikonos imagery. *ISPRS J Photogramm*, 60(3):195 – 211.

A Sources

A.1 Journals covered

- Image and Vision Computing, 2008–2009.
- International Journal of Computer Vision, 2004–2009.
- IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007–2009.
- ISPRS Journal of Photogrammetry and Remote Sensing, 2004–2009.
- Computer Vision and Image Understanding, 2008–2009.
- IEEE Computer Graphics and Applications, 2004–2009.

A.2 Conferences

- All IEEE conferences 2007–2009.
- XXI International CIPA Symposium, 01-06 October 2007, Athens, Greece.
- 3D-ARCH 2009: "3D Virtual Reconstruction and Visualization of Complex Architectures", Trento, Italy, 25-28 February 2009.
- BenCOS 2007: "Towards Benchmarking Automated Calibration, Orientation, and Surface Reconstruction from Images", Minneapolis, MN, June 23rd, 2007.

A.3 Research groups

- Prof. Luc Van Gool, Computer Vision Laboratory, ETH Zürich, `http://www.vision.ee.ethz.ch/research/index.en.html`.
- Ass. prof. Marc Pollefeys, Dept. Computer Science, University of North Carolina at Chapel Hill, `http://www.cs.unc.edu/~marc/research.html`.
- Christian Früh, Video and Image Processing Lab, University of California at Berkeley, `http://www-video.eecs.berkeley.edu/~frueh/3d/`.
- Prof. Wolgang Förstner, Institute of Geodesy and Geoinformation, Universität Bonn, `http://www.ipb.uni-bonn.de/foerstner/`.
- Prof. Pierre Grussenmeyer, Photogrammetry & Geomatics Group, National Institute of Applied Sciences of Strasbourg, `http://photogeo.insa-strasbourg.fr/index.htm`.

B Classified Reference List

B.1 Feature point detection and matching

Akgül, C. B., Sankur, B., Yemez, Y., and Schmitt, F. (2009). 3d model retrieval using probability density-based shape descriptors. *IEEE T Pattern Anal*, 31(6):1117–1133.

Avidan, S., Moses, Y., and Moses, Y. (2007). Centralized and distributed multi-view correspondence. *Int J Comp Vis*, 71(1):46–69.

Baker, S. and Matthews, I. (2004). Lucas-kanade 20 years on: A unifying framework. *Int J Comp Vis*, 56(3):221–255.

Baltsavias, E. P. (1991). *Multiphoto Geometrically Constrained Matching*. PhD thesis, Institute of Geodesy and Photogrammetry, ETH, Zürich, Switzerland.

Bartoli, A. (2008). Groupwise geometric and photometric direct image registration. *IEEE T Pattern Anal*, 30(12):2098–2108.

Bay, H., Ess, A., Tuytelaars, T., and van Gool, L. (2008). Speeded-up robust features (SURF). *Comp Vis Imag Under*, 110(3):346–359.

Bay, H., Tiytelaars, T., and Gool, L. V. (2006). Surf: Speeded up robust features. In *Proc Europ Conf Computer Vision*, Graz, Austria.

Bhowmick, P., Pradhan, R. K., and Bhattacharya, B. B. (2009). Approximate matching of digital point sets using a novel angular tree. *IEEE T Pattern Anal*, 31(5):769–782.

Bleyer, M. and Gelautz, M. (2005). A layered stereo matching algorithm using image segmentation and global visibility constraints. *ISPRS J Photogramm*, 59(3):128 – 150.

Bray, M., Koller-Meier, E., and Gool, L. V. (2007). Smart particle filtering for high-dimensional tracking. *Comp Vis Imag Under*, 106(1):116–129.

Bretzner, L. and Lindeberg, T. (1996). Feature tracking with automatic selection of spatial scales. Technical Report ISRN KTH NA/P-96/21-SE, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, Sweden.

Bretzner, L. and Lindeberg, T. (1998). Feature tracking with automatic selection of spatial scales. *Comp Vis Imag Under*, 71(3):385–392.

Brügelmann, R. and Förster, W. (1992). Noise estimation for color edge extraction. In Förstner, W. and Winter, S., editors, *Robust Computer Vision*, pages 90–107. Wichmann, Karlsruhe.

Carneiro, G. and Jepson, A. (2007). Flexible spatial configuration of local image features. *IEEE T Pattern Anal*, 29(12):2089–2104.

Chien, H.-J., Chen, C.-Y., Chen, C.-F., Su, Y.-M., and Chang, Y.-Y. (2008). Adaptive 3d reconstruction system with improved recovery of miscoded region to automatically adjust projected light patterns. In *Image and Vision Computing New Zealand, 2008. IVCNZ 2008. 23rd International Conference*, pages 1–6.

Choi, O. and Kweon, I. S. (2009). Robust feature point matching by preserving local geometric consistency. *Comp Vis Imag Under*, 113(6):726 – 742.

Chum, O. and Matas, J. (2008). Optimal randomized RANSAC. *IEEE T Pattern Anal*, 30(8):1472–1482.

Colombo, C., Del Bimbo, A., and Pernici, F. (2005). Metric 3d reconstruction and texture acquisition of surfaces of revolution from a single uncalibrated view. *IEEE T Pattern Anal*, 27(1):99–114.

Corso, J. and Hager, G. (2009). Image description with features that summarize. *Comp Vis Imag Under*, 113(4):446 – 458.

Crowley, J. L., Riff, O., and Piater, J. H. (2002). Fast computation of characteristic scale using a half-octave pyramid. In *CogVis 2002, International Workshop on Cognitive Computing*, Zurich.

Damon, J., Giblin, P., and Haslinger, G. (2009). Local image features resulting from 3-dimensional geometric features, illumination, and movement: I. *Int J Comp Vis*, 82(1):25–47.

Douxchamps, D. and Chihara, K. (2009). High-accuracy and robust localization of large control markers for geometric camera calibration. *IEEE T Pattern Anal*, 31(2):376–383.

Dowson, N. and Bowden, R. (2008). Mutual information for lucas-kanade tracking (milk): An inverse compositional formulation. *IEEE T Pattern Anal*, 30(1):180–185.

Drauschke, M., Schuster, H.-F., and Förstner, W. (2006). Detectibility of buildings in aerial images over scale space. In Förstner, W. and Steffen, R., editors, *Symposium of ISPRS Commission III: Photogrammetric Computer Vision*, volume XXXVI/Part 3, pages 7–12, Bonn. ISPRS, ISPRS.

Drewniok, C. and Rohr, K. (1997). Model-based detection and localization of circular landmarks in aerial images. *Int J Comp Vis*, 24(3):187–217.

Duan, C., Meng, X., and Wang, L. (2008). 3d reconstruction from uncalibrated images taken from widely separated views. In *Cybernetics and Intelligent Systems, 2008 IEEE Conference on*, pages 58–62.

Eade, E. and Drummond, T. (2009). Edge landmarks in monocular slam. *Image Vis Comput*, 27(5):588–596.

Ferrari, V., Tuytelaars, T., and Gool, L. V. (2006). Simultaneous object recognition and segmentation from single or multiple model views. *Int J Comp Vis*, 67(2):159–188.

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

Förstner, W. (1986). A feature based correspondence algorithm for image matching. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 26(3/3):150–166. Rovaniemi.

Förstner, W. (1993). *Image Matching*, volume II, chapter 16, pages 289–379. Addison-Wesley.

Förstner, W. (1994). A framework for low level feature extraction. In Eklundh, J.-O., editor, *Computer Vision-ECCV'94*, number 801 in Lecture Notes in Computer Science, pages 383–394. Springer Verlag.

Förstner, W. and Gülch, E. (1987). A fast operator for detection and precise location of distinct points, corners and circular features. In *Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken.

Frahm, J.-M. and Pollefeys, M. (2006). RANSAC for (quasi-)degenerate data (QDEGSAC). In *Proc. CVPR 2006*, volume 1, pages 453–460. IEEE.

Gevrekci, M. and Gunturk, B. K. (2009). Illumination robust interest point detection. *Comp Vis Imag Under*, 113(4):565 – 571.

Goshen, L. and Shimshoni, I. (2008). Guided sampling via weak motion models and outlier sample generation for epipolar geometry estimation. *Int J Comp Vis*, 80(2):275–288.

Gruen, A. (1996). Least squares matching: a fundamental measurement algorithm. In Atkinson, K. B., editor, *Close Range Photogrammetry and Machine Vision*, chapter 8, pages 217–255. Whittles, Caithness, Scotland.

Gruen, A. W. (1985). Adaptive least squares correlation: A powerful image matching technique. *S Afr J of Photogrammetry*, 14(3):175–187.

Harris, C. J. and Stephens, M. (1988). A combined corner and edge detector. In *4th Alvey Vision Conference*, pages 147–151, Manchester.

Heitger, F., Rosenthalter, L., von der Heydt, R., Peterhans, E., and Kuebler, O. (1992). Simulation of neuronal contour mechanisms: from simple to end-stopped cells. *Vision Res*, 32(5):963–981.

Jazayeri, I. and Fraser, C. (2008). Interest operators in close-range object reconstruction. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVII(B5):69–74.

Kadir, T., Zisserman, A., and Brady, M. (2004). An affine invariant salient region detector. In *ECCV'04*, volume 1, pages 228–241.

Ke, Y. and Sukthankar, R. (2004). Pca-sift: A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition*, volume 2, pages 506–513, Washington, DC, USA. IEEE Computer Society.

Kim, S. J. and Pollefeys, M. (2008). Robust radiometric calibration and vignetting correction. *IEEE T Pattern Anal*, 30(4):562–576.

Köthe, U. (2003a). Edge and junction detection with an improved structure tensor. In Michaelis, B. and Krell, G., editors, *Proceedings of 25th DAGM Symposium*, number 2781 in Lecture Notes in Computer Science, pages 25–32, Magdeburg, Germany. DAGM - German Pattern Recognition Society, Springer.

Köthe, U. (2003b). Integrated edge and junction detection with the boundary tensor. In *Proc of 9th Intl Conf on Computer Vision*, volume 1, pages 424–431, Nice. IEEE Computer Society.

Köthe, U. (2006). Low-level feature detection using the boundary tensor. In Weichert, J. and Hagen, H., editors, *Visualization and Processing of Tensor Fields*, Series on Mathematics and Visualization, pages 63–79. Springer, Berlin.

Laptev, I. (2005). On space-time interest points. *Int J Comp Vis*, 64(2-3):107–123.

Leung, C., Appleton, B., and Sun, C. (2008). Iterated dynamic programming and quadtree subregioning for fast stereo matching. *Image Vis Comput*, 26(10):1371–1383.

Li, Z., Liu, J., and Tang, X. (2007). A closed-form solution to 3d reconstruction of piecewise planar objects from single images. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–6.

Lindeberg, T. (1993). Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *Int J Comp Vis*, 11(3):283–318.

Lindeberg, T. (1994). Scale-space theory: A basic tool for analysing structures at different scales. *J Appl Stat*, 21(2):225–270.

Lindeberg, T. (1998). Feature detection with automatic scale selection. *Int J Comp Vis*, 30(2):79–116.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proc Intl Conf on Computer Vision*, pages 1150–1157, Corfu, Greece.

Lowe, D. G. (2001). Local feature view clustering for 3d object recognition. In *Proc IEEE Conf on Computer Vision and Pattern Recognition*, pages 682–688, Kauai, Hawaii. IEEE.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int J Comp Vis*, 60(2):91–110.

Lucas, B. D. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proc Intl Joint Conf Artificial Intelligence*, pages 674–679, Vancouver, British Columbia.

Matas, J., Chum, O., Urban, M., and Pajdla, T. (2002). Robust wide baseline stereo from maximally stable extremal regions. In Marshall, A. D. and Rosin, P. L., editors, *Proc British Machine Vision Conference*, pages 384–393, Cardiff, UK. British Machine Vision Association.

Matthews, L., Ishikawa, T., and Baker, S. (2004). The template update problem. *IEEE T Pattern Anal*, 26(6):810–815.

Mikolajczyk, K. and Schmid, C. (2002). An affine invariant interest point detector. In Heyden, A., Sparr, G., Nielsen, M., and Johansen, P., editors, *Proc. European Conference on Computer Vision*, number 2350 in Lecture Notes in Computer Science, pages 128–142, Copenhagen, Denmark. Springer.

Mikolajczyk, K. and Schmid, C. (2003). A performance evaluation of local descriptors. In *Computer Vision and Pattern Recognition*, volume 2, pages 257–263, Madison, WI, USA. IEEE Computer Society.

Mikolajczyk, K. and Schmid, C. (2004). Scale & affine invariant interest point detectors. *Int J Comp Vis*, 60(1):63–86.

Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE T Pattern Anal*, 27(10):1615–1630.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., and van Gool, L. (2005). A comparison of affine region detectors. *Int J Comp Vis*, 65(1-2):43–72.

Moisan, L. and Stival, B. (2004). A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *Int J Comp Vis*, 57(3):201–218.

Mokhtarian, F. and Soumela, R. (1998). Robust image corner detection through curvature scale space. *IEEE T Pattern Anal*, 20(12):1376–1381.

Moravec, H. P. (1981). Rover visual object avoidance. In Hayes, P. J., editor, *Proc Intl Joint Conf Artificial Intelligence*, pages 785–790, Vancouver, Canada. IJCAI, William Kaufmann.

Moreels, P. and Perona, P. (2007). Evaluation of features detectors and descriptors based on 3D objects. *Int J Comp Vis*, 73(3):263–284.

Qiang, J. and Haralick, R. M. (1998). Breakpoint detection using covariance propagation. *IEEE T Pattern Anal*, 20(8):845–851.

Remondino, F. (2006). Detectors and descriptors for photogrammetric applications. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVI(3):49–54.

Rohr, K. (1992a). Modelling and identification of characteristic intensity variations. *Image Vis Comput*, 10(2):66–76.

Rohr, K. (1992b). Recognizing corners by fitting parametric models. *Int J Comp Vis*, 9(3):213–230.

Rohr, K. (1994). Localization properties of direct corner detectors. *J Math Imaging Vis*, 4(2):139–150.

Sato, T. and Yokoya, N. (2005). New multi-baseline stereo by counting interest points. In *Proceedings of the 2nd Canadian conference on Computer and Robot Vision*, pages 96–103. IEEE.

Schmid, C. and Mohr, R. (1997). Local grayvalue invariants for image retrieval. *IEEE T Pattern Anal*, 19(5):530–543.

Shi, J. and Tomasi, C. (1994). Good features to track. In *Proc IEEE CVPR*, pages 593–600, Seattle. IEEE.

Shimizu, M. and Okutomi, M. (2005). Sub-pixel estimation error cancellation on area-based matching. *Int J Comp Vis*, 63(3):207–224.

Smith, S. M. and Brady, J. M. (1997). SUSAN — A new approach to low level image processing. *Int J Comp Vis*, 23(1):45–78.

Todorovic, S. and Ahuja, N. (2008). Region-based hierarchical image matching. *Int J Comp Vis*, 78(1):47–66.

Tomasi, C. and Kanade, T. (1991). Detection and tracking of point features. Technical Report CMU-CS-91-132, Carnegie Mellon University.

Torr, P. H. and Davidson, C. (2003). Impsac: Synthesis of importance sampling and random sample consensus. *IEEE T Pattern Anal*, 25(3):354–364.

Trajković, M. and Hedley, M. (1998). Fast corner detection. *Image Vis Comput*, 16(2):75–87.

Tuytelaars, T. and Gool, L. V. (2000). Wide baseline stereo matching based on local, affinely invariant regions. In Mirmehdi, M. and Thomas, B. T., editors, *Proc British Machine Vision Conference*, pages 412–422, Bristol, UK. British Machine Vision Association.

Tuytelaars, T. and Gool, L. V. (2004). Matching widely separated views based on affine invariant regions. *Int J Comp Vis*, 59(1):61–85.

Tuytelaars, T. and Mikolajczyk, K. (2007). Local invariant feature detectors: A survey. *Found Trend Comput Graphics Vis*, 3(3):177–280.

Wang, Z. and Quan, Y. (2008). An improved method for feature point matching in 3d reconstruction. In *Information Science and Engieering, 2008. ISISE '08. International Symposium on*, volume 1, pages 159–162.

Yang, Q., Wang, L., Yang, R., Stewenius, H., and Nister, D. (2009). Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. *IEEE T Pattern Anal*, 31(3):492–504.

Yoon, K.-J. and Kweon, I. S. (2008). Distinctive similarity measure for stereo matching under point ambiguity. *Comp Vis Imag Under*, 112(2):173 – 183.

Zhang, H. and Negahdaripour, S. (2008). Epiflow–a paradigm for tracking stereo correspondences. *Comp Vis Imag Under*, 111(3):307 – 328.

Zhu, Q., Wu, B., and Tian, Y. (2007). Propagation strategies for stereo image matching based on the dynamic triangle constraint. *ISPRS J Photogramm*, 62(4):295 – 308.

B.2 Camera calibration, bundle adjustment, and optimization

Abraham, S. and Förstner, W. (2005). Fish-eye-stereo calibration and epipolar rectification. *ISPRS J Photogramm*, 59(5):278 – 288.

Armstrong, M., Zisserman, A., and Hartley, R. I. (1996). Self-calibration from image triplets. In Buxton, B. F. and Cipolla, R., editors, *Computer Vision-ECCV'96*, number 1064 in Lecture Notes in Computer Science, pages 3–16. Springer Verlag.

Borghese, N. A., Colombo, F. M., and Alzati, A. (2006). Computing camera focal length by zooming a single point. *Pattern Recogn*, 39(8):1522–1529.

Börlin, N., Grussenmeyer, P., Eriksson, J., and Lindström, P. (2004). Pros and cons of constrained and unconstrained formulation of the bundle adjustment problem. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXV(B3):589–594.

Börlin, N., Lindström, P., and Eriksson, J. (2003). A globally convergent Gauss-Newton algorithm for the bundle adjustment problem with functional constraints. In Gruen, A. and Kahmen, H., editors, *Optical 3-D Measurement Techniques VI*, volume 2, pages 269–276. Wichmann-Verlag, Zürich, Switzerland.

Brown, D. C. (1966). Decentering distortion of lenses. *Photogramm Rec*, 32(3):1272–1292.

Brown, D. C. (1971). Close-range camera calibration. *Photogrammetric Engineering*, 37(8):855–866.

Brown, M. and Lowe, D. G. (2007). Automatic panoramic image stitching using invariant features. *Int J Comp Vis*, 74(1):59–73.

Chen, P. (2008). Optimization algorithms on subspaces: Revisiting missing data problem in low-rank matrix. *Int J Comp Vis*, 80(1):125–142.

Chen, P. and Suter, D. (2006). An analysis of linear subspace approaches for computer vision and pattern recognition. *Int J Comp Vis*, 68(1):83–106.

Chesi, G. (2009). Camera displacement via constrained minimization of the algebraic error. *IEEE T Pattern Anal*, 31(2):370–375.

Chojnacki, W., Brooks, M. J., and Hengel, A. V. D. (2001). Rationalising the renormalisation method of kanatani. *J Math Imaging Vis*, 14(1):21–38.

Devernay, F. and Faugeras, O. (2001). Straight lines have to be straight. *Mach Vision Appl*, 13(1):14–24.

Förstner, W. (2001). Generic estimation procedures for orientation with minimum and redundant information. In Gruen, A. and Huang, T. S., editors, *Calibration and Orientation of Cameras in Computer Vision*, volume 34 of *Springer Series in Information Sciences*. Springer.

Fraser, C. S. (1992). Photogrammetric camera component calibration — a review of analytical techniques. In *Workshop on Calibration and Orientation of Cameras in Computer Vision (TU-1), ISPRS XVII Congress*, Washington, D. C. ISPRS. 22 pages.

Fraser, C. S. (1997). Digital camera self-calibration. *ISPRS J Photogramm*, 52(4):149–159.

Hartley, R. and Kahl, F. (2009). Global optimization through rotation space search. *Int J Comp Vis*, 82(1):64–79.

Hartley, R. and Kang, S. B. (2007). Parameter-free radial distortion correction with center of distortion estimation. *IEEE T Pattern Anal*, 29(8):1309–1321.

Hasinoff, S. W. and Kutulakos, K. N. (2009). Confocal stereo. *Int J Comp Vis*, 81(1):82–104.

Heikkilä, J. and Silvén, O. (1997). A four-step camera calibration procedure with implicit image correction. In *Proc CVPR*, pages 1106–1112, San Juan, Puerto Rico. IEEE.

Helmke, U., Hüper, K., Lee, P. Y., and Moore, J. (2007). Essential matrix estimation using Gauss-Newton iterations on a manifold. *Int J Comp Vis*, 74(2):117–136.

Jianchar, Y. and Chern, C. T. (2001). Comparison of Newton-Gauss with Levenberg-Marquardt algorithm for space resection. In *ACRS 2001 — 22nd Asian Conference on Remote Sensing*, volume 1, pages 256–261, Singapore. Asian Association on Remote Sensing.

Jiang, H., Drew, M., and Li, Z.-N. (2007). Matching by linear programming and successive convexification. *IEEE T Pattern Anal*, 29(6):959–975.

Kahl, F., Agarwal, S., Chandraker, M. K., Kriegman, D., and Belongie, S. (2008). Practical global optimization for multiview geometry. *Int J Comp Vis*, 79(3):271–284.

Kahl, F. and Hartley, R. (2008). Multiple-view geometry under the L_infty-norm. *IEEE T Pattern Anal*, 30(9):1603–1617.

Kanatani, K., Nakatsuji, A., and Sugaya, Y. (2006). Stabilizing the focal length computation for 3-d reconstruction from two uncalibrated views. *Int J Comp Vis*, 66(2):109–122.

Ke, Q. and Kanade, T. (2007). Quasiconvex optimization for robust geometric reconstruction. *IEEE T Pattern Anal*, 29(10):1834–1847.

Lenz, R. K. and Tsai, R. Y. (1988). Techniques for calibration of the scale factor and image centre for high accuracy 3-d machine vision metrology. *IEEE T Pattern Anal*, 10(5):713–720.

Liu, B., Yu, M., Maier, D., and Männer, R. (2005). An efficient and accurate method for 3d-point reconstruction from multiple views. *Int J Comp Vis*, 65(3):175–188.

Lourakis, M. I. A. and Argyros, A. A. (2005). Is Levenberg-Marquardt the most efficient optimization algorithm for implementing bundle adjustment? In *Proceedings of IEEE International Conference on Computer Vision ICCV'05*, volume 2, pages 1526–1531, Beijing, China. IEEE.

Lourakis, M. I. A. and Argyros, A. A. (2009). Sba: A software package for generic sparse bundle adjustment. *ACM TOMS*, 36(1):30 pp.

Ma, L., Chen, Y., and Moore, K. L. (2003a). A family of simplified geometric distortion models for camera calibration.

Ma, L., Chen, Y., and Moore, K. L. (2003b). A new analytical radial distortion model for camera calibration.

Ma, Y., Huang, K., René Vidal, J. K., and Sastry, S. (2004). Rank conditions on the multiple-view matrix. *Int J Comp Vis*, 59(2):115–137.

Matei, B. and Meer, P. (2000). A general method for errors-in-variables problems in computer vision. In *Computer Vision and Pattern Recognition Conference*, volume II, pages 18–25. IEEE.

Maybank, S. J. and Faugeras, O. D. (1992). A theory of self-calibration of a moving camera. *Int J Comp Vis*, 8(2):123–151.

Nayak, A., Trucco, E., and Thacker, N. A. (2006). When are simple LS estimators enough? an empirical study of LS, TLS, and GTLS. *Int J Comp Vis*, 68(2):203–216.

Okatani, T. and Deguchi, K. (2007). On the wiberg algorithm for matrix factorization in the presence of missing components. *Int J Comp Vis*, 72(3):329–337.

Oliensis, J. and Hartley, R. (2007). Iterative extensions of the sturm/triggs algorithm: Convergence and nonconvergence. *IEEE T Pattern Anal*, 29(12):2217–2233.

Remondino, F. and Börlin, N. (2004). Photogrammetric calibration of image sequences acquired with a rotating camera. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXIV(5/W16). Panoramic Photogrammetry Workshop, Dresden, Germany.

Remondino, F. and Fraser, C. (2006). Digital camera calibration methods: Considerations and comparisons. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVI(5):266–272.

Rieke-Zapp, D., Tecklenburg, W., Peipe, J., Hastedt, H., and Haig, C. (2009). Evaluation of the geometric stability and the accuracy potential of digital cameras - comparing mechanical stabilisation versus parameterisation. *ISPRS J Photogramm*, 64(3):248 – 258.

Sengupta, S. and Das, S. (2008). Modified auto-calibration for 3d reconstruction from multiple views of an object. In *TENCON 2008 - 2008, TENCON 2008. IEEE Region 10 Conference*, pages 1–6.

Triggs, B., McLauchlan, P., Hartley, R., and Fitzgibbon, A. (2000). Bundle adjustment — A modern synthesis. In *Vision Algorithms: Theory and Practice, Proceedings of the International Workshop on Vision Algorithms*, volume 1883 of *Lecture Notes in Computer Science*, pages 298–372. Spring Verlag.

Tsai, R. Y. (1986). An efficient and accurate camera calibration technique for 3d machine vision. In *Proceedings of IEEE Computer Vision and Pattern Recognition Conference*, pages 364–374, Miami Beach. IEEE.

Weng, J., Cohen, P., and Herniou, M. (1992). Camera calibration with distortion models and accuracy evaluation. *IEEE T Pattern Anal*, 14(10):965–980.

Wu, F. C., Wang, Z. H., and Hu, Z. Y. (2009). Cayley transformation and numerical stability of calibration equation. *Int J Comp Vis*, 82(2):156–184.

Wu, Y., Li, Y., and Hu, Z. (2008). Detecting and handling unreliable points for camera parameter estimation. *Int J Comp Vis*, 79(2):209–223.

Zhen, C., Ming, L., Shao-feng, J., Shui-gen, W., and Yu, W. (2009). A new non-linear algorithm for 3d reconstruction from straight-line optical flow. In *Informatics in Control, Automation and Robotics, 2009. CAR '09. International Asia Conference on*, pages 311–316.

B.3 Relative and absolute orientation, 3D reconstruction, co-registration

Akca, D. (2003). Full automatic registration of laser scanner point clouds. In *Proc. of Optical 3-D Measurement Techniques VI*, volume I, pages 330–337, Zurich, Switzerland. ISPRS.

Akca, D. (2007). Matching of 3d surfaces and their intensities. *ISPRS J Photogramm*, 62(2):112 – 121.

Akca, D. and Gruen, A. (2007). Generalized least squares multiple 3d surface matching. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVI(3/W52):1–7.

Arun, K. S., Huang, T. S., and Blodstein, S. D. (1987). Least-squares fitting of two 3-d point sets. *IEEE T Pattern Anal*, 9(5):698–700.

Bae, K.-H. and Lichti, D. D. (2008). A method for automated registration of unorganised point clouds. *ISPRS J Photogramm*, 63(1):36 – 54.

Barnea, S. and Filin, S. (2008). Keypoint based autonomous registration of terrestrial laser point-clouds. *ISPRS J Photogramm*, 63(1):19 – 35.

Barrois, B. and Wöhler, C. (2007). 3d pose estimation based on multiple monocular cues. In *CVPR'07*, page 8 pp.

Bartoli, A. and Sturm, P. (2004). Nonlinear estimation of the fundamental matrix with minimal parameters. *IEEE T Pattern Anal*, 26(3):426–432.

Beder, C. and Förstner, W. (2006). Direct solutions for computing cylinders from minimal sets of 3d points. In Leonardis, A., Bischof, H., and Pinz, A., editors, *Proceedings of the European Conference on Computer Vision*, number 3951 in LNCS, pages 135–146. Springer.

Besl, P. J. and McKay, N. D. (1992). A method for registration of 3-d shapes. *IEEE T Pattern Anal*, 14(2):239–256.

Boström, G., Gonçalves, J. G., and Sequeira, V. (2008). Controlled 3d data fusion using error-bounds. *ISPRS J Photogramm*, 63(1):55 – 67.

Brenner, C., Dold, C., and Ripperda, N. (2008). Coarse orientation of terrestrial laser scans in urban environments. *ISPRS J Photogramm*, 63(1):4 – 18.

Brown, M. and Lowe, D. G. (2005). Unsupervised 3d object recognition and reconstruction in unordered datasets. In *Proc. 5th International Conference on 3-D Digital Imaging and Modeling (3DIM 2005)*, pages 56–63, Ottawa, Canada.

Carballeira, P., Ronda, J., and Valdes, A. (2008). 3d reconstruction with uncalibrated cameras using the six-line conic variety. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 205–208.

David, P., DeMenthon, D., Duraiswami, R., and Samet, H. (2004). SoftPOSIT: Simultaneous pose and correspondence determination. *Int J Comp Vis*, 59(3):259–284.

Faugeras, O. D. and Maybank, S. (1990). Motion from point matches: multiplicity of solutions. *Int J Comp Vis*, 3(4):225–246.

Gao, X.-S., Hou, X.-R., Tang, J., and Cheng, H.-F. (2003). Complete solution classification for the perspective-three-point problem. *IEEE T Pattern Anal*, 25(8):930–943.

González-Aguilera, D., Rodríguez-Gonzálvez, P., and Gómez-Lahoz, J. (2009). An automatic procedure for co-registration of terrestrial laser scanners and digital cameras. *ISPRS J Photogramm*, 64(3):308 – 316.

Gruen, A. and Akca, D. (2005). Least squares 3d surface and curve matching. *ISPRS J Photogramm*, 59(3):151 – 174.

Grussenmeyer, P. and Khalil, O. A. (2002). Solutions for exterior orientation in photogrammetry: A review. *Photogramm Rec*, 17:615–634.

Haralick, R. M., Lee, C., Ottenberg, K., and Nölle, M. (1991). Analysis and solutions of the three point perspective pose estimation problem. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 592–598. IEEE.

Haralick, R. M., Lee, C.-N., Ottenberg, K., and Nölle, M. (1994). Review and analysis of solutions of the three point perspective pose estimation problem. *Int J Comp Vis*, 13(3):331–356.

Hartley, R. I. (1997). In defense of the eight-point algorithm. *IEEE T Pattern Anal*, 19(6):580–593.

Hartley, R. I., Gupta, R., and Chang, T. (1992). Stereo from uncalibrated cameras. In *CVPR'1992*, pages 761–764.

Huang, T. and Faugeras, O. (1989). Some properties of the E matrix in two-view motion estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(12):1310–1312.

Huber, D. F. and Hebert, M. (2003). Fully automatic registration of multiple 3d data sets. *Image Vis Comput*, 21(7):637 – 650.

Hung, Y. S. and and, W. K. T. (2006). Projective reconstruction from multiple views with minimization of 2d reprojection error. *Int J Comp Vis*, 66(3):305–317.

Johnson, A. and Hebert, M. (1999). Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE T Pattern Anal*, 21(5):433–449.

Josephson, K., Byröd, M., Kahl, F., and Åström, K. (2007). Image-based localization using hybrid feature correspondences. In *BenCOS 2007: Towards Benchmarking Automated Calibration, Orientation, and Surface Reconstruction from Images*, page 8 pp, Minneapolis, USA. ISPRS/IEEE.

Kanatani, K. (1993). Unbiased estimation and statistical analysis of 3-d rigid motion from two views. *IEEE T Pattern Anal*, 15(1):37–50.

Kazhdan, M. (2007). An approximate and efficient method for optimal rotation alignment of 3d models. *IEEE T Pattern Anal*, 29(7):1221–1229.

Läbe, T. and Förstner, W. (2006). Automatic relative orientation of images. In *Proc. of the 5th Turkish-German Joint Geodetic Days*, page 6 pp, Berlin, Germany.

Lehmann, S., Bradley, A., Clarkson, I., Williams, J., and Kootsookos, P. (2007). Correspondence-free determination of the affine fundamental matrix. *IEEE T Pattern Anal*, 29(1):82–97.

Liu, S., Zhao, L., Li, J., and Xu, H. (2008). The research of 3d reconstruction from uncalibrated image sequences combined with 3d models. In *Computer Science and Software Engineering, 2008 International Conference on*, volume 2, pages 1117–1119.

Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing scene from two projections. *Nature*, 293:133–135.

Nistér, D. (2004a). An efficient solution to the five-point relative pose problem. *IEEE T Pattern Anal*, 26(6):756–770.

Nistér, D. (2004b). Untwisting a projective reconstruction. *Int J Comp Vis*, 60(2):165–183.

Nistér, D. and Schaffalitzky, F. (2006). Four points in two or three calibrated views: Theory and practice. *Int J Comp Vis*, 67(2):211–231.

Olsson, C., Kahl, F., and Oskarsson, M. (2009). Branch-and-bound methods for euclidean registration problems. *IEEE T Pattern Anal*, 31(5):783–794.

Pottmann, H., Huang, Q.-X., Yang, Y.-L., and Hu, S.-M. (2006). Geometry and convergence analysis of algorithms for registration of 3d shapes. *Int J Comp Vis*, 67(3):277–296.

Rabbani, T., Dijkman, S., van den Heuvel, F., and Vosselman, G. (2007). An integrated approach for modelling and global registration of point clouds. *ISPRS J Photogramm*, 61(6):355 – 370.

Raguram, R., Frahm, J.-M., and Pollefeys, M. (2008). A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus. In *ECCV*, volume 2, pages 500–513.

Remondino, F. and Ressl, C. (2006). Overview and experiences in automated markerless image orientation. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVI(3):248–254.

Schaffalitzky, F. and Zisserman, A. (2002). Multi-view matching for unordered image sets, or "how do i organise my holiday snaps?". In *ECCV'02*, volume 1, pages 414–431, Copenhagen, Denmark.

Schweighofer, G. and Pinz, A. (2006). Robust pose estimation from a planar target. *IEEE T Pattern Anal*, 28(12):2024–2030.

Shashua, A. (1997). Trilinear tensor: The fundamental construct of multiple-view geometry and its applications. In *Proc Int. Workshop on Algebraic Frames for the Perception Action Cycle (AFPAC)*, volume 1315 of *LNCS*, pages 190–206, Kiel, Germany. Springer.

Smith, E. R., King, B. J., Stewart, C. V., and Radke, R. J. (2008). Registration of combined range-intensity scans: Initialization through verification. *Comp Vis Imag Under*, 110(2):226 – 244.

Söderkvist, I. and Wedin, P.-Å. (1993). Determining the movements of the skeleton using well-configured markers. *J Biomech*, 26(12):1473–1477.

Steffen, R. and Förstner, W. (2008). On visual real time mapping for unmanned aerial vehicles. In *21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, volume B3a, pages 57–62, Beijing, China.

Stewénius, H., Engels, C., and Nistér, D. (2006). Recent developments on direct relative orientation. *ISPRS J Photogramm*, 60(4):284–294.

Stewénius, H., Nistér, D., Kahl, F., and Schaffalitzky, F. (2005). A minimal solution for relative pose with unknown focal length. In *Proc. IEEE CVPR*, volume 2, pages 789–794, San Diego. IEEE.

Stewénius, H., Nistér, D., Kahl, F., and Schaffalitzky, F. (2008). A minimal solution for relative pose with unknown focal length. *Image Vis Comput*, 26(7):871–877.

Svoboda, T. and Pajdla, T. (2002). Epipolar geometry for central catadioptric cameras. *Int J Comp Vis*, 49(1):23–37.

Wendt, A. (2007). A concept for feature based data registration by simultaneous consideration of laser scanner data and photogrammetric images. *ISPRS J Photogramm*, 62(2):122 – 134.

Wolfe, W. J., Mathis, D., Sklair, C. W., and Magee, M. (1991). The perspectiev view of three points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(1):66–73.

B.4 Dense stereo

Cech, J. and Sara, R. (2007). Efficient sampling of disparity space for fast and accurate matching. In *CVPR'07*, page 8 pp.

Collins, R. T. (1996). A space-sweep approach to true multi-image matching. In *CVPR'96*, pages 358–363.

Criminisi, A., Blake, A., Rother, C., Shotton, J., and Torr, P. H. S. (2007). Efficient dense stereo with occlusions for new view-synthesis by four-state dynamic programming. *Int J Comp Vis*, 71(1):89–110.

El-Melegy, M. T. and Al-Ashwal, N. H. (2007). A variational technique for 3d reconstruction from multiple views. In *Computer Engineering & Systems, 2007. ICCES '07. International Conference on*, pages 38–43.

Kang, S. B. and Szeliski, R. (2004). Extracting view-dependent depth maps from a collection of images. *Int J Comp Vis*, 58(2):139–163.

Koch, R. (2005). 3-d surface reconstruction from stereoscopic image sequences. In *Proceedings of IEEE International Conference on Computer Vision ICCV'05*, pages 109–114, Beijing, China. IEEE.

Kostlivá, J., J. Cech, J., and Sára, R. (2007). Feasibility boundary in dense and semi-dense stereo matching. In *CVPR'07*, page 8 pp.

Nehab, D., Weyrich, T., and Rusinkiewicz, S. (2008). Dense 3d reconstruction from specularity consistency. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8.

Ogale, A. S. and Aloimonos, Y. (2005). Shape and the stereo correspondence problem. *Int J Comp Vis*, 65(3):147–162.

Okutomi, M. and Kanade, T. (1993). A multiple-baseline stereo. *IEEE T Pattern Anal*, 15(4):353–363.

Pollefeys, M. and Sinha, S. (2004). Iso-disparity surfaces for general stereo configurations. In Pajdla, T. and Matas, J., editors, *Proc of ECCV 2004*, volume 3 of *LNCS 3023*, pages 509–520, Prague, Czech Republic. Springer.

Rav-Acha, A., Engel, G., and Peleg, S. (2008). Minimal aspect distortion (MAD) mosaicing of long scenes. *Int J Comp Vis*, 78(2-3):187–206.

Remondino, F. and Zhang, L. (2006). Surface reconstruction algorithms for detailed close-range object modeling. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVI(3):117–123.

Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int J Comp Vis*, 47(1-3):7–42.

Strecha, C., von Hansen, W., Gool, L. V., Fua, P., and Thoennessen, U. (2008). On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *CVPR'08*, pages 1–8.

Szeliski, R. and Golland, P. (1999). Stereo matching with transparency and matting. *Int J Comp Vis*, 32(1):45–61.

Zhang, G., Jia, J., Wong, T.-T., and Bao, H. (2009). Consistent depth maps recovery from a video sequence. *IEEE T Pattern Anal*, 31(6):974–988.

Zhang, L. and Seitz, S. (2007). Estimating optimal parameters for mrf stereo from a single image pair. *IEEE T Pattern Anal*, 29(2):331–342.

B.5 Interpretation, labelling, and segmentation of 3D data

Baillard, C. and Zisserman, A. (2000). A plane-sweep strategy for the 3D reconstruction of buildings from multiple images. In *19th ISPRS Congress and Exhibition*, Amsterdam.

Bosché, F. and Haas, C. T. (2008). Automated retrieval of 3d cad model objects in construction range images. *Automat Constr*, 17(4):499–512.

Bucksch, A. and Lindenbergh, R. (2008). CAMPINO — a skeletonization method for point cloud processing. *ISPRS J Photogramm*, 63(1):115 – 127.

Cao, L., Liu, J., and Tang, X. (2008). What the back of the object looks like: 3d reconstruction from line drawings without hidden lines. *IEEE T Pattern Anal*, 30(3):507–517.

Chen, J. and Chen, B. (2008). Architectural modeling from sparsely scanned range data. *Int J Comp Vis*, 78(2-3):223–236.

Drauschke, M. and Förstner, W. (2008a). Comparison of adaboost and adtboost for feature subset selection. In *8th International Workshop on Pattern Recognition in Information Systems (PRIS 2008)*, pages 113–122, Barcelona, Spain.

Drauschke, M. and Förstner, W. (2008b). Selecting appropriate features for detecting buildings and building parts. In *21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, volume B3b-1, pages 447–452, Beijing, China.

Farenzena, M., Fusiello, A., Gherardi, R., and Toldo, R. (2009). Automatic structure recovery and visualization. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVIII(5/W1):6 pp.

Förstner, W. and Steffen, R. (2007). Online geocoding and evaluation of large scale imagery without gps. In Fritsch, D., editor, *Photogrammetric Week '07*, pages 243–253, Heidelberg. Wichmann Verlag.

Früh, C. and Zakhor, A. (2002). Data processing algorithms for generating textured 3d building façade meshes from laser scans and camera images. In *1st International Symposium on 3D Data Processing Visualization and Transmission (3DPVT 2002)*, pages 834–849, Padova, Italy. IEEE Computer Society.

Gruen, A. and Wang, X. (1998). Cc-modeler: a topology generator for 3-d city models. *ISPRS J Photogramm*, 53(5):286 – 295.

Kwon, S.-W., Bosché, F., Kim, C., Haas, C. T., and Liapi, K. A. (2004). Fitting range data to primitives for rapid local 3d modeling using sparse range point clouds. *Automat Constr*, 13(1):67–81.

Leibe, B., Schindler, K., Cornelis, N., and van Gool, L. (2008). Coupled object detection and tracking from static cameras and moving vehicles. *IEEE T Pattern Anal*, 30(10):1683–1698.

Mayer, H. (2008). Object extraction in photogrammetric computer vision. *ISPRS J Photogramm*, 63(2):213 – 222.

Mayer, H. and Reznik, S. (2007). Building facade interpretation from uncalibrated wide-baseline image sequences. *ISPRS J Photogramm*, 61(6):371 – 380.

Meidow, J., Beder, C., and Förstner, W. (2009). Reasoning with uncertain points, straight lines, and straight line segments in 2d. *ISPRS J Photogramm*, 64(2):125 – 139.

Müller, P., Zeng, G., Wonka, P., and Gool, L. V. (2007). Image-based procedural modeling of facades. *ACM TOG*, 26(3):85.

Penna, M. and Dines, K. (2007). A simple method for fitting sphere-like surfaces. *IEEE T Pattern Anal*, 29(9):1673–1678.

Rottensteiner, F., Trinder, J., Clode, S., and Kubik, K. (2007). Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis. *ISPRS J Photogramm*, 62(2):135 – 149.

Stamos, I., Liu, L., Chen, C., Wolberg, G., Yu, G., and Zokai, S. (2008). Integrating automated range registration with multiview geometry for the photorealistic modeling of large-scale scenes. *Int J Comp Vis*, 78(2-3):237–260.

Tarscha-Kurdi, F., Landes, T., Grussenmeyer, P., and Koehl, M. (2007). Model-driven and data-driven approaches using lidar data: Analysis and comparison. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 36(3/W49A):6p.

Tarsha-Kurdi, F., Landes, T., and Grussenmeyer, P. (2008). Extended RANSAC algorithm for automatic detection of building roof planes from lidar data. *The Photogrammetric Journal of Finland*, 21(1):97–109.

Terzopoulos, D. (1988). The computation of visible-surface representations. *IEEE T Pattern Anal*, 10(4):417–438.

Troccoli, A. and Allen, P. (2008). Building illumination coherent 3d models of large-scale outdoor scenes. *Int J Comp Vis*, 78(2-3):261–280.

Wenzel, S., Drauschke, M., and Förstner, W. (2008). Detection of repeated structures in facade images. *PattRecImageAnal*, 18(3):406–411.

Wenzel, S. and Förstner, W. (2008). Semi-supervised incremental learning of hierarchical appearance models. In *21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, volume B3b-2, pages 399–404, Beijing, China.

Yu, G., Grossberg, M., Wolberg, G., and Stamos, I. (2008). Think globally, cluster locally: A unified framework for range segmentation. In *Proceedings of 3DPVT'08 — the Fourth Internation Symposium on 3D Data Processing, Visualization and Transmissions*, page 8 pp, Atlanta, GA, USA.

B.6 Error analysis

Adami, A., Guerra, F., and Vernier, P. (2007). Laser scanner and architectural accuracy test. In *Proceedings of XXI Intl CIPA Symposium*, page 5 pp, Athens, Greece. CIPA.

Akca, D., Freeman, M., Gruen, A., and Sargent, I. (2008). Quality assessment of 3d building data by 3d surface matching. In *21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, volume B2-2, pages 771–778, Beijing, China.

Blostein, S. D. and Huang, T. S. (1988). Correction to "Error analysis in stereo determination of 3-d point positions". *IEEE T Pattern Anal*, 10(5):765.

Boehler, W., Vicent, M. B., and Marbs, A. (2003). Investigating laser scanner accuracy. In *Proceedings of CIPA 2003 XIXth Int'l Symposium*, page 9 pp, Antalya, Turkey. CIPA.

Dickscheid, T., Läbe, T., and Förstner, W. (2008). Benchmarking automatic bundle adjustment results. In *21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, volume B3a, pages 7–12, Beijing, China.

Dorst, L. (2005). First order error propagation of the procrustes method for 3d attitude estimation. *IEEE T Pattern Anal*, 27(2):221–229.

Förstner, W. (1982). On the geometric precision of digital correleation. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 24(III):176–189.

Förstner, W. (1987). Reliability analysis of parameter estimation in linear models with applications to mensuration problems in computer vision. *Comput Vision Graphics Image Process*, 40:273–310.

Förstner, W. and Wrobel, B. (2004). *Mathematical Concepts in Photogrammetry*, chapter 2, pages 15–180. IAPRS, 5 edition.

Heuel, S. (2004). *Uncertain Projective Geometry: Statistical Reasoning for Polyhedral Object Reconstruction*. Number 3008 in Lecture Notes in Computer Science. Springer, Berlin.

Kanatani, K. (2008). Statistical optimization for geometric fitting: Theoretical accuracy bound and high order error analysis. *Int J Comp Vis*, 80(2):167–188.

Kraus, K. (1993). *Photogrammetry, Volume 1*. Ferdinand Dümmler, Bonn.

Läbe, T., Dickscheid, T., and Förstner, W. (2008). On the quality of automatic relative orientation procedures. In *21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, volume B3b-1, pages 37–42, Beijing, China.

Liu, X., Kanungo, T., and Haralick, R. (2005). On the use of error propagation for statistical validation of computer vision software. *IEEE T Pattern Anal*, 27(10):1603–1614.

Luhmann, T. (2009). Precision potential of photogrammetric 6dof pose estimation with a single camera. *ISPRS J Photogramm*, 64(3):275 – 284.

Marengoni, M., Hanson, A., Zilberstein, S., and Riseman, E. (2003). Decision making and uncertainty management in a 3d reconstruction system. *IEEE T Pattern Anal*, 25(7):852–858.

Matei, B. and Meer, P. (2006). Estimation of nonlinear errors-in-variables models for computer vision applications. *IEEE T Pattern Anal*, 28(10):1537–1552.

Šegvić, S., Schweighofer, G., and Pinz, A. (2007). Influence of numerical conditioning on the accuracy of relative orientation. In *BenCOS 2007: Towards Benchmarking Automated Calibration, Orientation, and Surface Reconstruction from Images*, page 8 pp, Minneapolis, USA. ISPRS/IEEE.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR'06*, volume 1, pages 519–528.

Shi, W., Cheung, C.-K., and Tong, X. (2004). Modelling error propagation in vector-based overlay analysis. *ISPRS J Photogramm*, 59(1-2):47 – 59.

Thacker, N. A., Clark, A. F., Barron, J. L., Beveridge, J. R., Courtney, P., Crum, W. R., Ramesh, V., and Clark, C. (2008). Performance characterization in computer vision: A guide to best practices. *Comp Vis Imag Under*, 109(3):305 – 334.

B.7 Applications

Abdelhafiz, A. (2009). *Integrating Digital Photogrammetry and Terrestrial Laser Scanning*. PhD thesis, Institute for Geodesy and Photogrammetry, Technical University Braunschweig, Braunschweig, Germany.

Abellán, A., Vilaplana, J., and Martínez, J. (2006). Application of a long-range terrestrial laser scanner to a detailed rockfall study at vall de núria (eastern pyrenees, spain). *Eng Geol*, 88(3-4):136 – 148.

Achille, C., Brumana, R., Fassi, F., and Tuncer, H. (2007). Application of mixed technique for the 3d modeling of the noble floor of the villa reale in monza. In *Proceedings of XXI Intl CIPA Symposium*, page 6 pp, Athens, Greece. CIPA.

Akbarzadeh, A., Frahm, J.-M., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Merrell, P., Phelps, M., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewénius, H., Yang, R., Welch, G., Towles, H., Nistér, D., and Pollefeys, M. (2006). Towards urban 3d reconstruction from video. In *Proc. 3DPVT'06*, pages 1–8, Chapel Hill, North Carolina, USA. IEEE.

Bendea, H., Chiabrando, F., Tonolo, F. G., and Marenchino, D. (2007). Mapping of archaeological areas using a low-cost UAV the augusta bagiennorum test site. In *Proceedings of XXI Intl CIPA Symposium*, page 6 pp, Athens, Greece. CIPA.

Bosché, F., Haas, C. T., and Akinci, B. (2009). Automated recognition of 3d cad objects in site laser scans for project 3d status visualization and performance control. *J Comput Civil Eng*, x(x):x. in press.

Cornelis, N., Leibe, B., Cornelis, K., and Gool, L. V. (2008). 3d urban scene modeling integrating recognition and reconstruction. *Int J Comp Vis*, 78(2-3):121–141.

Debevec, P. E. (1996). *Modeling and Rendering Architecture from Photographs*. PhD thesis, University of California at Berkeley.

Debevec, P. E. (2003). Image-based techniques for digitizing environments and artifacts. In *Proceedings of the 4th International Conference on 3D Digital Imaging and Modeling (3DIM 2003)*, pages 234–242, Banff, Canada. IEEE. Invited paper.

Debevec, P. E., Yu, Y., and Borshukov, G. D. (1998). Efficient view-dependent image-based rendering with projective texture-mapping. In Drettakis, G. and Max, N., editors, *Proc. Eurographics Rendering Workshop*, pages 105–116, Viennea, Austria.

Di, K., Xu, F., Wang, J., Agarwal, S., Brodyagina, E., Li, R., and Matthies, L. (2008). Photogrammetric processing of rover imagery of the 2003 mars exploration rover mission. *ISPRS J Photogramm*, 63(2):181 – 201.

El-Hakim, S. (2002). Semi-automatic 3d reconstruction of occluded and unmarked surfaces from widely separated views. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 34(5):143–148.

El-Hakim, S., Beraldin, J.-A., Picard, M., and Godin, G. (2004). Detailed 3d reconstruction of large-scale heritage sites with integrated techniques. *IEEE Comput Graphics Appl*, 24(3):21–29.

Forberg, A. (2007). Generalization of 3d building data based on a scale-space approach. *ISPRS J Photogramm*, 62(2):104 – 111.

Fors Nilsson, H. and Grundberg, D. (2009). Plane-based close range photogrammetric reconstruction of buildings. Master's thesis, Department of Computing Science, Umeå University.

Frahm, J.-M., Pollefeys, M., Clipp, B., Gallup, D., Raguram, R., Wu, C., and Zach, C. (2009). 3d reconstruction of architectural scenes from uncalibrated video sequences. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVIII(5/W1):7 pp.

Fraser, C. S. and Cronk, S. (2009). A hybrid measurement approach for close-range photogrammetry. *ISPRS J Photogramm*, 64(3):328 – 333.

Früh, C. (2003). *Automated 3D model generation for urban environments*. PhD thesis, Fakultät für Elektrotechnik und Informationstechnik, Universität Karlsruhe, Karlsruhe, Germany.

Fruh, C. and Zakhor, A. (2003). Constructing 3d city models by merging aerial and ground views. *IEEE Comput Graphics Appl*, 23(6):52–61.

Früh, C. and Zakhor, A. (2004). An automated method for large-scale, ground-based city model acquisition. *Int J Comp Vis*, 60(1):5–24.

Gallup, D., Frahm, J.-M., Mordohai, P., Yang, Q., and Pollefeys, M. (2007). Real-time plane-sweeping stereo with multiple sweeping directions. In *Proc. CVPR*, pages 1–8, Minneapolis, Minnesota, USA. IEEE.

Gerth, B., Berndt, R., Havemann, S., and Fellner, D. W. (2005). 3d modeling for non-expert users with the castle construction kit v0.5. In Mudge, M., Ryan, N., and Scopigno, R., editors, *Proc. of 6th Int'l Symposium on Virtual Reality, Archeology and Cultural Heritage (VAST)*, pages 1–9, Pisa, Italy.

Gonzo, L., El-Hakim, S., Picard, M., Girardi, S., and Whiting, E. (2004). Photo-realistic 3-d reconstruction of castles with multiple- sources image-based techniques. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 35(B5):120–125.

Grün, A. (2000). Semi-automated approaches to site recording and modeling. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 33(5/1):309–318. Invited paper.

Grün, A., Remondino, F., and Zhang, L. (2004). Photogrammetric reconstruction of the great buddha of bamiyan, afghanistan. *Photogramm Rec*, 19(107):177–199.

Grussenmeyer, P. and al Khalil, O. (2000). A comparison of photogrammetry software packages for the documentation of buildings. In *International Congress of the International Federation of Surveyors*, Malta.

Grussenmeyer, P., Hanke, K., and Streilein, A. (2002). Architectural photogrammetry. In Kasser, M. and Egels, Y., editors, *Digital Photogrammetry*, pages 300–339. Taylor & Francis.

Grussenmeyer, P., Landes, T., Voegtle, T., and Ringle, K. (2008). Comparison methods of terrestrial laser scanning, photogrammetry and tacheometry data for recording of cultural heritage buildings. In *21st Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS)*, volume B5, pages 213–218, Beijing, China.

Hu, J., You, S., and Neumann, U. (2003). Approaches to large-scale urban modeling. *Computer Graphics and Applications, IEEE*, 23(6):62–69.

Karara, H. M., editor (1989). *Non-Topographic Photogrammetry*. Science & Engineering Series. American Society for Photogrammetry and Remote Sensing, 2nd edition.

Landes, T., Grussenmeyer, P., Voegtle, T., and Ringle, K. (2007). Combination of terrestrial recording techniques for 3d object modelling regarding topographic constraints. example of the castle of haut-andlau, alsace, france. In *Proceedings of XXI Intl CIPA Symposium*, page 6 pp, Athens, Greece. CIPA.

Leibe, B., Leonardis, A., and Schiele, B. (2008). Robust object detection with interleaved categorization and segmentation. *Int J Comp Vis*, 77(1-3):259–289.

Matthies, L., Maimone, M., Johnson, A., Cheng, Y., Willson, R., Villalpando, C., Goldberg, S., Huertas, A., Stein, A., and Angelova, A. (2007). Computer vision on mars. *Int J Comp Vis*, 75(1):67–92.

McGlone, C., Mikhail, E., and Bethel, J., editors (2004). *Manual of Photogrammetry*. ASPRS, 5th edition.

Menci, L. and Rinaudo, F. (2007). New trends in digital photogrammetry teaching and diffusion: The Z-GLIF software. In *Proceedings of XXI Intl CIPA Symposium*, page 4 pp, Athens, Greece. CIPA.

Mikhail, E. M., Bethel, J. S., and McGlone, J. C. (2001). *Introduction to Modern Photogrammetry*. Wiley.

Monserrat, O. and Crosetto, M. (2008). Deformation measurement using terrestrial laser scanning data and least squares 3d surface matching. *ISPRS J Photogramm*, 63(1):142 – 154.

Mordohai, P., Frahm, J.-M., Akbarzadeh, A., Clipp, B., Engels, C., Gallup, D., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewénius, H., Towles, H., Welch, G., Yang, R., Pollefeys, M., and Nistér, D. (2007). Real-time video-based reconstruction of urban environments. In *Proc. 3D-ARCH'2007*, Zürich, Switzerland. ISPRS.

Nistér, D. (2001). *Automatic Dense Reconstruction from Uncalibrated Video Sequences*. PhD thesis, Royal Institute of Technology, Stockholm, Sweden.

Ortin, D. and Remondino, F. (2005). Occlusion-free image generation for realistic texture mapping. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVI(5/W17):7 pp.

Pollefeys, M. (2004). Automatic 3d modeling with hand-held camera images. Tutorial given at ISPRS'04 Congress, Istanbul.

Pollefeys, M., Gool, L. J. V., and Proesmans, M. (1996). Euclidean 3d reconstruction from image sequences with variable focal lenghts. In Buxton, B. F. and Cipolla, R., editors, *Computer Vision-ECCV'96*, number 1064 in Lecture Notes in Computer Science, pages 31–42. Springer Verlag.

Pollefeys, M., Gool, L. V., Vergauwen, M., Cornelis, K., Verbiest, F., and Tops, J. (2003). 3d recording for archeological fieldwork. *IEEE Comput Graphics Appl*, pages 20–27.

Pollefeys, M., Gool, L. V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., and Koch, R. (2004). Visual modeling with a hand-held camera. *Int J Comp Vis*, 59(3):207–232.

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., and Towles, H. (2008). Detailed real-time urban 3d reconstruction from video. *Int J Comp Vis*, 78(2-3):143–167.

Rabbani, T. and van den Heuvel, F. (2004). 3d industrial reconstruction by fitting csg models to a combination of images and point clouds. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, 35(B5):7–12.

Remondino, F. (2006). *Image-based modeling for object and human reconstruction*. PhD thesis, Institute of Geodesy and Photogrammetry, ETH Zürich, ETH Hoenggerberg, Zürich, Swizerland.

Remondino, F. and El-Hakim, S. (2006). Image-based 3D modelling: A review. *Photogramm Rec*, 21(115):269–291.

Remondino, F., El-Hakim, S., Girardi, S., Rizzi, A., Benedetti, S., and Gonzo, L. (2009). 3d virtual reconstruction and visualization of complex architectures - the 3d-arch project. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVIII(5/W1):9 pp.

Remondino, F., El-Hakim, S. F., Gruen, A., and Zhang, L. (2008). Development and performance analysis of image matching for detailed surface reconstruction of heritage objects. *IEEE Signal Proc Mag*, 25(4):55–64.

Remondino, F., Guarnieri, A., and Vettore, A. (2005). 3d modeling of close-range objects: Photogrammetry or laser scanning? In *Proc. SPIE-IS&T Electronic Imaging: Videometrics VIII*, volume 5665, pages 216–225, San Jose, California.

Remondino, F. and Niederoest, J. (2004). Generation of high-resolution mosaic for photo-realistic texture-mapping of cultural heritage 3d models. In Cain, K., Chrysanthou, Y., Niccolucci, F., and Silberman, N., editors, *Proc. of 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST)*, pages 85–92, Brussels, Belgium.

Rottensteiner, F. (2003). Automatic generation of high-quality building models from lidar data. *IEEE Comput Graphics Appl*, 23(6):42–50.

Snavely, N., Seitz, S. M., and Szeliski, R. (2006). Photo tourism: Exploring photo collections in 3d. In *SIGGRAPH Conference Proceedings*, pages 835–846, New York, NY, USA. ACM Press.

Snavely, N., Seitz, S. M., and Szeliski, R. (2008). Modeling the world from internet photo collections. *Int J Comp Vis*, 80(2):189–210.

Stamos, I. (2009). Automated 3d modeling of urban environments. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVIII(5/W1):5 pp.

Tsioukas, V. (2007). Simple tools for architectural photogrammetry. In *Proceedings of XXI Intl CIPA Symposium*, page 4 pp, Athens, Greece. CIPA.

Voltolini, F., Remondino, F., Pontin, M., and Gonzo, L. (2006). Experiences and considerations in image-based modeling of complex architectures. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, XXXVI(5):309–314.

Zalesny, A., Ferrari, V., Caenen, G., and Gool, L. V. (2005). Composite texture synthesis. *Int J Comp Vis*, 62(1/2):161–176.

Zhang, L. and Gruen, A. (2006). Multi-image matching for dsm generation from ikonos imagery. *ISPRS J Photogramm*, 60(3):195 – 211.

C Toolbox

C.1 Project idea

- Input: A set of overlapping images of buildings and optionally 3D data from a laser scanner.
- The buildings are *not* assumed to be free-standing, i.e. the toolbox must be able to handle real-world restrictions such as occlusions, etc.
- The toolbox should be able to handle images taken by uncalibrated or calibrated camera, optionally on a stereo rig.
- The toolbox should support measurements of primitives such as points, lines, and planes.
- It should be possible to combine primitives into e.g. a rectangular house with a saddle-back roof.
- It should be possible to use stored combined primitives for faster measurements of recurring objects.
- The output should be a geometric model *with precision estimates* and texture maps for visualization.

C.2 Toolbox organization

- Key questions:
  - What algorithms should be included?
  - How should the data be organized?
  - What kind of user interface is required? (API/GUI)
  - What about data import/export features?
  - Compatibility with other software?

C.3 Toolbox themes

- The algorithms should be as automatic as possible.
- The user input should be where it is most efficient, i.e. either at the beginning or the end of a calculation pipeline.
- If a choice is necessary, the code should be transparent, i.e. easy to read, rather than optimized.
- The code should contain references to the literature.
- Estimation functions should produce error estimates in addition to the estimated values.

- For non-advanced users, the estimation functions should have traffic-light features (green=ok, red=not ok, yellow=manual attention required).

Quality measures (Fraser) 0th...3rd

C.4 Algorithms

Tentative list of core algorithms:

- Orientation
- Triangulation
- Bundle adjustment
- Camera calibration
- Feature detection
- Feature matching
- Least squares matching
- Lens distortion correction
- Co-registration
- Texture extraction

Other algorithms:

- Coded targets
- ...

C.4.1 Orientation

- Relative orientation
  - Direct solution
    * Eight-point
    * Seven-point
    * Five-point
  - Statistically optimal solution
  - Robust solutions
- Trifocal tensor
  - Direct solution
    * Eighteen-parameters

        * Eleven-parameters
    - Statistically optimal solution
    - Robust solutions

- Absolute orientation
    - Direct
    - Optimal

C.4.2 Triangulation

- Forward intersection
- Bundle adjustment
    - Datum
    - Control points
    - Self-calibration
    - Functional (hard) constraints
    - Soft constraints
    - Robustness (bad initial values)
    - Robustness (outliers)
    - Robust estimation (Huber, Cauchy, ...)

C.4.3 Feature point extraction

- Förstner/Köthe interest operator.
- Harris corner
- SIFT

C.4.4 Least squares matching

- Standard
- Multi-photo constrained

C.4.5 Algorithm validation and simulation

- Simulation
    - Camera networks
    - Image distortion
- Optimization of camera networks
- Algorithm validation

C.5 Data organization

Tentative data organization:

- Project
  - Cameras
  - Images
  - Measurements
  - Objects
  - Uncertainties
  - ...

C.6 Camera models

- Perspective (pin-hole)
  - Outer orientation
    - roll, pitch, yaw
    - azimuth, tilt, swing
    - other
  - Lens distortion models
    - Brown K1, K2, K3, P1, P2
    - Chebychev polynomials
    - Modelling of zoom lenses
    - ...
- Non-perspective cameras
  - Line-cameras
  - Fish-eye
  - ...
- Stereo rig

C.7 Measurement tools

- GUI for
  - Manual measurements
  - "Guided" measurements
  - Input of approximate values

C.8 Visualization

- Points
- Lines
- Cameras
- Residuals
- Error ellipsoids
- ...