Critical overview of image-based 3D modeling

F. Remondino
Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland

S. El-Hakim
Visual Information Technology (VIT), National Research Council, Ottawa, Canada

ABSTRACT: In this paper we address the main problems and available solutions for the generation of 3D models from terrestrial images. Close range photogrammetry deals since many years with manual or automatic image measurements for precise 3D modeling. Nowadays 3D scanners are also becoming a standard source for input data in many application areas, but image-based modeling still remain the most complete, cheap, portable, flexible and widely used approach. After reviewing the different 3D shape techniques for surface reconstruction, we will report the full pipeline for 3D modeling from image data. Some modeling methods are also described and different examples are presented.

1 INTRODUCTION

Nowadays the generation of a 3D model is achieved using range or image data or a combination of both methods. Range data contains already the three-dimensional coordinates necessary for the modeling phase while images needs a mathematical model to derive the object coordinates. Image-based approaches require information that can often be extracted automatically from the images. In some cases, automated procedures are not satisfactory or can not recover all the required features, therefore manual or semi-automated measurements must be performed. After the measurements, the data must be structured and a consistent polygonal surface generated to build a realistic representation of the imaged scene.

We can actually distinguish 4 alternative methods for scene modeling:

- Image-based rendering (IBR): it does not include the generation of a geometric 3D model, but for particular objects and under specific camera motions and scene conditions, it might be considered a good technique for the generation of virtual views (Kang, 1999). The technique relies on either accurately knowing the camera positions or automatic stereo matching that, in the absence of geometric data, requires a large number of closely spaced images to succeed. Object occlusions and discontinuities particularly in large-scale and geometrically complex environments will also affect the output. The ability to move freely into the scene and viewing objects from any position may be limited depending on the method used. It is therefore unlikely that IBR will be the approach of choice for purposes other than limited visualization.
- Image-based modeling (IBM): it is the widely used method for geometric surfaces of architectural objects (Streilein 1994, Debevec et al. 1996, Van den Heuvel 1999, El-Hakim 2002) or for terrain and city modeling (Gruen 2000). In most of the cases, the most impressive and accurate results remain those achieved with interactive approaches.
- Range-based modeling: this method can directly capture the 3D geometric information of an object. It is based on expensive active sensor (e.g. Breuckmann™, Cyberware™, ShapeGrabber™) that often lack in texture information. However, texture or color infor-

mation can be added, either from the scanner itself via a color channel or from a separate digital camera (Beraldin et al. 2002, Guidi et al. 2003).
- Combination of image- and range-based modeling: particularly for large structures and scenes no one technique by itself can efficiently provide the complete model. Therefore, we believe that combining techniques where the basic shapes are determined by image-based methods and fine details by laser scanning is the logical solution. El-Hakim et al. (2004) used aerial and terrestrial images for the main shapes, and laser scanning for fine geometric details to fully model the abbey of Pomposa in Italy. Beraldin et al. (2005) combined the 3D technologies to model the heritage site of Selinunte and a Byzanthine Crypt near Lecce (Italy).

For sake of completeness, 3D modeling and animation software (Lightwave™, Maya™, 3DMax™) are also used for the generation of 3D models from scratch. Starting from simple elements like polygonal boxes, they generally subdivide and smooth them using 3D splines, and provide for realistic results. They are mainly used for movie production, architectural and object design.

A wide range of applications requires 3D reconstruction of real world objects and scenes. In general, most applications specify a number of requirements (El-Hakim 2002): high geometric accuracy, capturing all details, and photo-realism. In addition, full automation, low cost, portability, flexibility in applications, and efficiency in model size are also desirable features. The order of importance of these requirements depends on the application, but in many all are important. A single system that satisfies all requirements is still in the future. In particular, accurately covering all the details with a fully automated low-cost system for a wide range of objects and scene remains elusive.

In this work, after a short overview on the 3D reconstruction systems, only the terrestrial image-based 3D modeling problem for close-range applications will be discussed in detail.

2 3D SURFACE RECONSTRUCTION METHODS

The most general classification of 3D object measurement and reconstruction techniques can be done in contact (CMM, rulers, bearing, etc) and non-contact (laser range, patter projection, X-Ray, SAR, photogrammetry, etc) methods. Non-contact measurement techniques are widely used, in particular in industrial application, heritage modeling and documentation or scene reconstruction. Non-contact measurement techniques based on light waves (e.g. Chen et al. 2000) can be generally divided into two classes:

1. Systems based on active sensors: these projector-based methods rely on artificial lights or patter projection. For many years structured light (Maas 1992, Gartner et al. 1995, Sablatnig & Menard 1997, Breukmann™), coded light (Wahl 1984) or laser light (Sequeira et al. 1999, Cyberware™, ShapeGrapper™, Riegl™) are used for the measurements of objects. Currently many commercial solutions are available, based on triangulation (with laser light or stripes projection), time-of-flight, continuous wave or reflectivity measurements. 3D active sensors are becoming a common approach for objects or scene recording and a standard source for geometric data. They provide for millions of points and recently also for the associated color information. These sensors can be quite expensive, designed for specific applications and depend on the reflective characteristics of the surface. Based on these ideas, in the last twenty-five years many progresses have been made in the field of solid-state electronics and photonics and different 3D sensing have been developed (Beraldin et al. 2000, Blais 2004).
2. Systems based on passive sensor (image-based approaches with ambient light): they use 2D image measurements to recover 3D object information (e.g. photogrammetry) or they estimate surface normals instead of 3D data like shape from shading (Horn & Brooks 1989), shape from texture (Kender, 1978), shape from specularity (Healey & Binford 1987), shape from contour (medical applications) (Asada 1993, Ulipinar & Nevatia 1995), shape from 2D edge gradients (Winkelback & Wahl 2001). Passive image-based methods acquire 3D

measurements from single or multi-stations; they use projective geometry or perspective camera model; the sensors are very portable and low cost.

3 TERRESTRIAL IMAGE-BASED 3D MODELING

Recovering a complete, accurate and realistic 3D model from images is a very difficult task, in particular if uncalibrated or widely separated images are used. Firstly because the wrong recovery of the parameters could lead to inaccurate and deformed results, and secondly because a wide baseline between images always requires the user interaction for points measurements.

Image-based modeling, by definition, obtains measurements and 3D models from photographs. Deterministic or probabilistic methods are generally employed. In particular, for many years photogrammetry dealt with the 3D reconstruction of objects from images and currently several commercial packages are available. They are all based on manual or semi-automated measurements (Australis™, Canoma™, ImageModeler™, iWitness™, PhotoModeler™, ShapeCapture™); they allow, after an orientation and bundle adjustment phase, to obtain sensor calibration data and three-dimensional object point coordinates from multi-sensor or multi-image networks, and most provide textured 3D models.

The research activities in terrestrial image-based modeling can be classified as:

1. Approaches that try to get a 3D model of the scene having just the images as input (also called ‘shape from video’ or ‘VHS to VRML’). Many efforts to completely automate the process of taking images, calibrate and orient them, recover the 3D coordinates of the imaged scene and model them have been done, but while promising, the methods are thus far not always successful. The full-automated procedure, which is widely reported in computer vision (Fitzgibbon & Zisserman 1998, Pollefeys et al. 1999, Liebowitz et al. 1999, Roth & Whitehead 2000, Nister 2004), starts with a sequence of images taken by an uncalibrated camera. The system automatically extracts interest points, like corners, sequentially matches them across views, then computes camera parameters and 3D coordinates of the matched points using robust techniques. The key to the success of this fully automatic procedure is that successive images cannot vary significantly, thus the images must be taken at short intervals. The first two images are generally used to initialize the sequence. This is done in a projective geometry basis and is usually followed by a bundle adjustment. A “self-calibration” to compute the intrinsic camera parameters (usually only the focal length), is generally used in order to obtain metric reconstruction, up to a scale, from the projective one. The 3D model is then automatically generated in case of simple geometry by means of dense depth maps. Few approaches have been also presented for the automated extraction of image correspondences between wide-baseline images (Pritchett et al. 1997, Ferrari et al. 2003, Xiao et al. 2003), but their reliability and applicability for automated image-based modeling of complex scenes is still not satisfactory. Automated image-based methods rely on features that can be extracted and matched, therefore occlusions, illumination changes, limited locations for the image acquisition and un-textured surfaces are problematic. Furthermore, it is very common that an automated process ends up with areas containing too many features that are not all required for modeling while there are areas without any or with a minimum number of features that cannot produce a complete 3D model. Automated processes require highly structured closely spaced images with good texture, and uniform camera motion, otherwise they will inevitably fail. The level of automation is also strictly related to the quality (accuracy) of the required 3D model. ‘Nice-looking’ 3D models, used for visualization purposes can certainly be generated with automated processes, while for documentation, high accuracy and photo-realistic models, user interaction is required. For all these reasons, more emphasis has been put on semi-automated and interactive procedures, combining the human ability of image understanding with the powerful capacity and speed of computers. This has led to a number of promising approaches for the modeling of architectures and complex geometric objects.
2. Approaches that interactively or automatically orient and calibrate the images and afterwards perform a semi-automated 3D reconstruction and modeling of the scene. Semi-automated modeling approaches that rely on the human operator are much more common.

In particular, in case of complex geometric objects, the interactive creation of a 3D model is done to define the topology and then to edit or post-process the 3D data. An output model based only on the measured points will usually consist of surface boundaries that are irregular and overlapping and need some assumption to be corrected using for example planes and plane intersections. For large structures and scenes, since the technique may require a large number of images, the creation of the model requires a significant human interaction even if image registration and a large number of 3D points were computed fully automatically. The degree of modeling automation increases when certain assumptions about the object, such as architectures, can be made. Debevec et al. (1996) developed a hybrid easy to use system to create 3D models of architectures from a small number of photographs. It is the well-known Façade, afterwards included in the commercial software Canoma^{TM}. The basic geometric shape of a structure is first recovered using models of polyhedral elements. In this interactive step, the actual size of the elements and camera pose are captured assuming that the camera intrinsic parameters are known. The second step is an automated matching procedure, constrained by the now known basic model to add geometric details. The approach proved to be effective in creating geometrically accurate and realistic models. The drawback is the high level of interaction. Also since assumed shapes determine camera poses and all 3D points, the results are as accurate as the assumption that the structure elements match those shapes. Van den Heuvel (1999) uses a line-photogrammetric mathematical model and geometric constraints to recover the 3D shapes of polyhedral objects. Using lines, occluded object points can also be reconstructed and part of occluded objects can be modeled due to the introduction of coplanarity constraints. El-Hakim (2002) developed a semi-automatic technique (partially implemented in ShapeCapture^{TM}) able to recover 3D model of simple as well as complex objects, as will be described in section 3.3.2.

3. Approaches that interactively or automatically orient and calibrate the images and afterwards fully automatically recover the geometry of the scene using constraints. These techniques are employed for architectural objects with very simple geometry. Dick et al. (2001) employ model-based recognition technique to extract high-level models in a single image and then use their projection into other images for verification. The method requires parameterized building blocks with a priori distribution defined by the building style. The scene is modeled as a set of base planes corresponding to walls or roofs, each of which may contain offset 3D shapes that model common architecture elements such as windows and columns. Again, the full automation necessitates feature detection and projective geometry approach, however the technique used here also employs planner constraints and perpendicularity between planes to improve the matching process. Werner & Zisserman (2002) proposed a fully automated façade-like approach: instead of the basic shapes, the principal planes of the scene are created automatically to assemble a coarse model. Wilczkowiak et al. (2003) proposed a similar approach, based on three dominating directions that are assumed to be perpendicular to each other: the coarse model guides a more refined polyhedral model of details such as windows, doors, and wedge blocks. Since this is a fully automated approach, it requires feature detection and closely spaced images for the automatic matching and camera pose estimation using projective geometry.

The overall image-based 3D modeling process (Figure 1) consists of few well-known steps:

1. Design (sensor and network geometry)
2. Measurements (point clouds, lines, etc.)
3. Structuring/Modeling (geometry, texture)
4. Visualization/Analysis

Currently, the recovery of the sensor and network geometry and the measurement phase are often separated from the modeling and visualization part (Figure 1). But in many applications this gap has to be bridged in order to perform correct measurements and recover realistic 3D models.

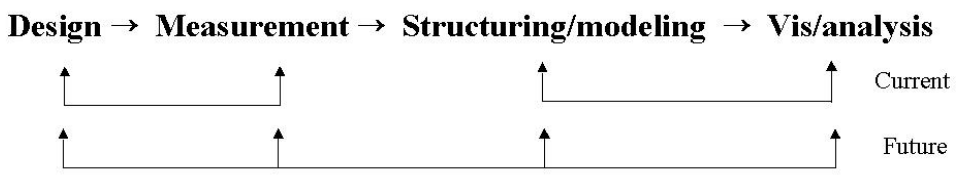


Figure 1. Photogrammetric modeling and visualization process.

3.1 *Design and recovery of the network geometry*

In order to optimize the measurement operations in terms of accuracy and reliability, particular attention must be given to the design of the network of images. This includes, among other things, deciding on suitable sensor and image measurements scheme, how many camera stations are needed, and their locations to have good imaging geometry. The network configuration determines the quality of the calibration and defines the imaging geometry. Unfortunately, in practice, the network design phase is often not considered, or impossible to apply in the actual object setting, or the images are obtained from existing videos, leading to poor imaging geometry.

3.2 *Surface measurement*

The measurement step can be performed with manual, semi- or fully-automated procedures. Automated photogrammetric matching algorithms (e.g. D'Apuzzo 2003), usually performed by means of template least squares matching (Förstner 1982, Ackermann 1983, Gruen 1985) can produce very dense point clouds, but mismatches, irrelevant points and missing parts could be present in the results, requiring post-processing. Furthermore, these automated procedures usually do not take into consideration the geometrical conditions of the object's surface when applying smoothing constraints: therefore it is often difficult to correctly turn randomly generated point clouds into polygonal structures of high quality without losing important information. The smoothing effects of automated template matching algorithms are mainly due to:

- The image patches of the matching algorithm are assumed to correspond to planar object surface patches: along small objects or corners, this assumption is not valid anymore, therefore these features are smoothed out (Figure 2);
- Smaller image patches could theoretically avoid or reduce the smoothing effects, but may not be suitable for the correct determination of the matching reshaping parameters because a small patch may not include enough image signals content.

In the vision community, mainly two-frames stereo correspondence algorithms are used [Dhond and Aggarwal 1989; Brown 1992; Scharstein and Szeliski 2002], producing a dense disparity map, i.e. a parallax estimate at each pixel. Having the exterior orientation of the images, the 3D coordinates can be computed. Often the second image is resampled according to the epipolar geometry, to have parallax value in only one direction. A large number of algorithms have been developed and the dense output is generally used for view synthesis, image-based rendering or modeling of complete regions. Despite a dense output, these methods declared errors between 3% and 5% (i.e. 1:20), limiting therefore their use in applications that require nice-looking 3D models (Pollefeys et al. 1999).

On the other hand, if the measurements are done in manual (mono- or stereoscopically) or semi-automatic mode, the reliability and accuracy of the measures are higher but a smaller number of points describing the object. Moreover, in case of manual stereo-measurements, it is very important for the operator to understand the functional behavior of the 3D modeling software to perform correct measurements. In fact, as the number of measurements is small compared to automated procedures, the points should identify and describe the salient features of the object. In this context, the on-line modeler should be able to project onto the images the generated mesh to control the agreement between measurements and structure of the generated object surface.

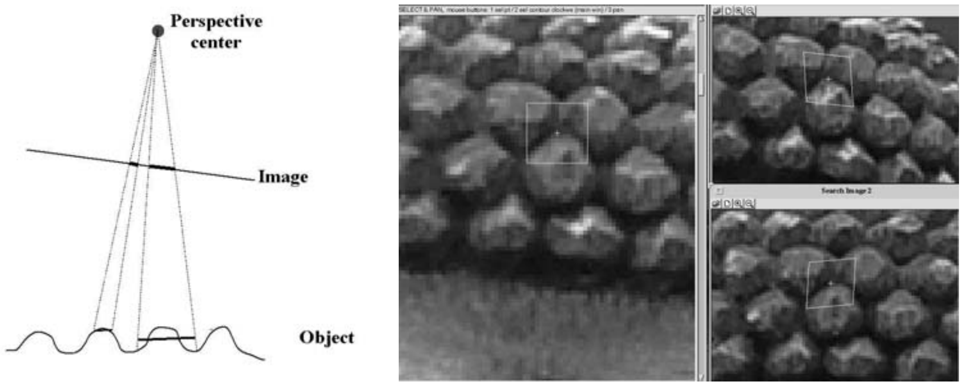


Figure 2. Patch definition in template least squares matching (left). Triplet of images where patches assumed to correspond to planar object surface patches where the assumption is not valid (right).

3.3 *From 3D point cloud to surface model*

For some modeling applications, like buildings reconstruction, where the object is mainly described with planar patches, the measured points are already structured and the surface generation is often achieved with few triangular surfaces or fitting particular surfaces to the data (section 3.3.2). For other applications, like statues, human body or complex objects, the surface generation from the measured points is much more difficult and requires smart algorithms to triangulate all the measured points, in particular in case of uneven and sparse point clouds.

The goal of surface reconstruction can be stated as follows: given a set of sample points P_i assumed to lie on or near an unknown surface S, create a surface model S' approximating S. A surface reconstruction procedure (also called surface triangulation) cannot guarantee the recovering of S exactly, since we have information about S only through a finite set of sample points P_i. Sometimes additional information of the surface (e.g. breaklines) are available and generally, as the sampling density increases, the output result S' is more likely topologically correct and converges to the original surface S. A good sample should be dense in detailed area and sparse otherwise. But usually the measured points are unorganized and often noisy; moreover the surface can be arbitrary, with unknown topological type and with sharp details. Therefore the reconstruction method must infer the correct geometry, topology and all details just from a finite set of sample points. Usually if the input data does not satisfy certain properties required by the triangulation algorithm (like good points distribution and density, small noise, etc.), the program will produce incorrect results or may fail altogether.

To classify all the triangulation methods is beyond the scope of this paper. There is a large amount of published work and in the following they are reported according to some categories, like ’used approach’, ’type of data’ or ’representation’. Some algorithms could belong to different groups even though they are listed only once.

A first and very general classification is done according to the *quality (type) of the input data*:

- Unorganized point clouds: algorithms working on unorganized data have no other information on the input data except their spatial position. They do not use any assumption on the object geometry and therefore, before generating a polygonal surface, they usually structure the points according to their coherence. They need a good distribution of the input data and if the points are not uniformly distributed they easily fail.
- Structured point clouds: algorithms based on structured data can take into account additional information of the points (e.g. breaklines, coplanarity. etc.).

A further distinction can be done according to their *spatial subdivision*:

- Surface oriented algorithms do not distinguish between open and closed surfaces. Most of the available algorithms belong to this group (Hoppe et al. 1992, Mencl 2001).
- Volume oriented approaches work in particular with closed surfaces and generally are based on the Delaunay tetrahedrization of the given set of sample points (Boissonnat 1984, Curles & Levoy 1996, Isselhard 1997). Voxel-based mesh generation techniques

also belong to this category. The marching cubes algorithm (Lorensen & Cline 1987) is usually used to create the triangles for each voxel. Although this is a very efficient and easy to implement technique, the accuracy of the generated mesh is an approximation of the original 3D data and the larger the voxel size the less accurate, and more smoothed out, the final model will be.

Another classification is based on the *type of representation* of the surface:

- Parametric representation: these methods represent the surface as a number of parametric surface patches, described by parametric equations. Multiple patches may then be pieced together to form a continuous surface. Examples of parametric representations include B-spline, Bezier curves, NURBS, and Coons patches (Terzopoulos 1988).
- Implicit algebraic representation: these methods try to find a smooth function that passes through all positions where the implicit function evaluates to some specified value (usually zero) (Gotsman & Keren 1998).
- Simplicial representation: in this representation the surface is a collection of simple entities including points, edges and triangles. This group includes Alpha shapes (Edelsbrunner & Mucke, 1994) and the Crusts algorithm (Amenta et al. 1998).
- Approximated surfaces: they do not always contain all the original points, but points as near as possible to them. They can use a distance function (shortest distance of a point in space from the generated surface) to estimate the correct mesh (Hoppe et al. 1992). In this group we can also insert the warping-based surface reconstruction (they deform an initial surface so that it gives a good approximation of the given set of points) (Muraki 1991) and the implicit surface fitting algorithms (they fit e.g. piecewise polynomial functions to the given set of points) (Moore & Warren 1990).
- Interpolated surfaces: these algorithms are used when precise models are needed: all the input data is used and a correct linkage of them is necessary (Dey & Giesen 2001).

Finally, the reconstruction methods can be divided according to their *assumptions*:

- Algorithms assuming fixed shapes: they usually assume that the surface shape is known a priori (e.g. plane, cylinder or sphere) (Brinkley 1985, Hastie & Stuetzle 1989).
- Algorithms exploiting structure or orientation information: many surface reconstruction algorithms exploit the structure of the data for the surface reconstruction. For example, in case of multiple scans, they can use the adjacency relationship of the data within each range image (Merrian 1992). Other reconstruction methods instead use knowledge of the orientation of the surface that is supplied with the data. For example, if the points are obtained from volumetric data, the gradient of this data can provide orientation information useful for the reconstruction (Miller et al. 1991).

Conversion of measured data into consistent polygonal surface is usually based on four steps:

1. Pre-processing: in this phase erroneous data are eliminated or point clouds are sampled to reduce the computation time;
2. Determination of the global topology of the object's surface: the neighborhood relations between adjacent parts of the surface have to be derived. This operation typically needs some global sorting step and the consideration of possible 'constraints' (e.g. breaklines), mainly to preserve special features (e.g. edges);
3. Generation of the polygonal surface (section 3.3.1): triangular (or tetrahedral) meshes are created satisfying certain quality requirements, e.g. limit on the meshes element size, no intersection of breaklines, etc.;
4. Post-processing: when the model is created, editing operations are commonly applied to refine or correct the generated polygonal surface.

3.3.1 *Triangulation or mesh generation*

This is the core part of almost all reconstruction programs (Edelsbrunner 2001) gives a good introduction to the topic). Triangulation converts a given set of points into a consistent polygonal mesh. This operation partitions the input data into simples and usually generates vertices, edges and faces (representing the analyzed surface) that meet only at shared edges. Finite element methods are used to discretize the measured domain by dividing it into many small 'elements', typically triangles or quadrilaterals in two dimensions and tetrahedra in three dimensions. An optimal triangulation is defined by measuring angles, edge lengths, height or area of the elements while the error of the finite element approximations is usually related to the minimum angle of the elements. The vertices of the triangulation can be exactly the input points or extra points, called Steiner points, which are inserted to create a more optimal mesh (Bern & Eppstein 1992) Triangulation can be performed in 2D or in 3D, according to the input data:

- 2D Triangulation: the input domain is a polygonal region of the plane and, as result, triangles that intersect only at shared edges and vertices are generated. A well known construction method is the Delaunay Triangulation (DT) that simultaneously optimize several quality measures, like edge lengths, height or area of the elements. Delaunay criterion ensures that no vertex lies within the interior of any of the circumcircles of the triangles in the network. DT of a given set of point is the dual of the Voronoi diagram (also called the Thiessen or Dirichlet tessellation). In Voronoi diagram, each region consists of the part of the plane nearest to that node: connecting the nodes of the Voronoi cells that have common boundaries forms the Delaunay triangles.
- 2.5D Triangulation: the input data is a set of points P in a plane along with a real and unique elevation function f(x,y) at each point (x,y). A 2.5D triangulation creates a linear function F interpolating P and defined on the region bounded by the convex hull of P. For each point p in P, F(p) is the weighted average of the elevation of the vertices of the triangle that contains p. Usually Delaunay triangulation is used as interpolation function. According to the data structure, regularly or almost randomly distributed, the generated surface is called elevation grid or TIN (Triangulated Irregular Network) model.
- Surfaces for 3D models: the input data is always a set of point P in R^3, but no more restricted on a plane; therefore the elevation function *f(x,y)* is no more unique. The input set is also called unorganized point cloud.
- 3D Triangulation: the triangulation in 3D is called tetrahedralization or tetrahedrization. A tetrahedralization is a partition of the input domain into a collection of tetrahedra that meet only at shared faces (vertices, edges or triangles). Tetrahedralization results are much more complicated than a 2D triangulation. The types of input domains could be simple polyhedron (sphere), non-simple polyhedron (torus) or point clouds.

3.3.2 *Generation of 3D points on basic geometric surfaces*

The image-based modeling approach described here is the semi-automatic technique initially described in El-Hakim, 2002. The method, although similar in philosophy to some existing approaches like Façade, interactively measures only a small number of seed points in multiple images while most of the points are determined automatically. This achieves more flexibility and gives higher levels of detail than most other methods. In addition, the camera poses and 3D coordinates of seed points are determined without any assumption of the shapes but instead by bundle adjustment, with or without self-calibration depending on the given configuration. This achieves higher geometric accuracy regardless of the shape of the object. Here, we describe a specially developed procedure aimed at automatically adding large numbers of 3D points to surfaces of basic shapes that can be represented by an algebraic implicit function: *F(X, Y, Z) = 0*, such as groin vault ceiling. A number of seed points (at least 9 in case of quadric or cylindrical surface) are measured manually in multiple images on each section (Figure 3). A quadric is fitted to the seed points in each section. The section can be either rectangle or triangle in boundary. Using the parameters of the quadric and the known camera internal and external parameters for a given image, any number of 3D points can be added automatically within the boundary of the section (Figure 4). In order to achieve smooth surface model, we may need to generate hundreds of thousands of points per shape. Therefore, efficient modeling software, such as PolyWorks™, is used to create the triangular mesh.

The general rule for adding points on an element and for generating points in occluded or symmetrical parts is to do the work in the 3D space to find new points then project them on the images using the known camera parameters. A cylinder is constructed after its direction and approximate radius and position have been automatically determined from four seed points (

Figure 5a). The ratio between the upper and the lower circle can be set in advance, for example to 0.85 to create a tapered column. From this information, 3D points on the top and bottom circle of the column (

Figure 5b) are automatically generated resulting in a complete model (Fig. 5c, d).

Reconstructing arches is similar to the approach used in Façade except that this approach uses seed points instead of blocks and the arch points are extracted automatically. First a plane is fitted to seed points on the wall (Figure 6a). An edge detector is applied to the region (Figure 6b) and points at constant interval along the arch are automatically sampled. Using image coordinates of these points (in one image), the known image parameters, and the equation of the plane, the 3D coordinates are computed and projected on the images (Figure 6c). Blocks, even when partially invisible, can also be modeled. For example in figure 6-d the part of the middle block where it meets the base is not visible and needs to be measured in order to reconstruct the whole block. To solve this problem, we first extract the visible corners on all blocks from several images and compute their 3D coordinates. We then fit a plan to the top of the base block, using the gray points in figure 8, then project a normal to this plane from each of the corners of the block attached to it (the white points). The intersections of each normal will produce a new point (a black point in Figure 6d) automatically. Using symmetry, we can fully construct the block.

For windows and doors we need four corner points and one point on the main surface. By fitting a plane to the corner points, and a plane parallel to it at the surface point, the complete window or door can be reconstructed. To model steps, sufficient seed points to compute the two side planes, plus one point on either side of each step are needed. A variety of shapes or architectures can be modeled with this procedure.

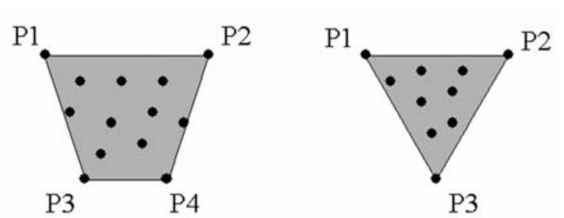
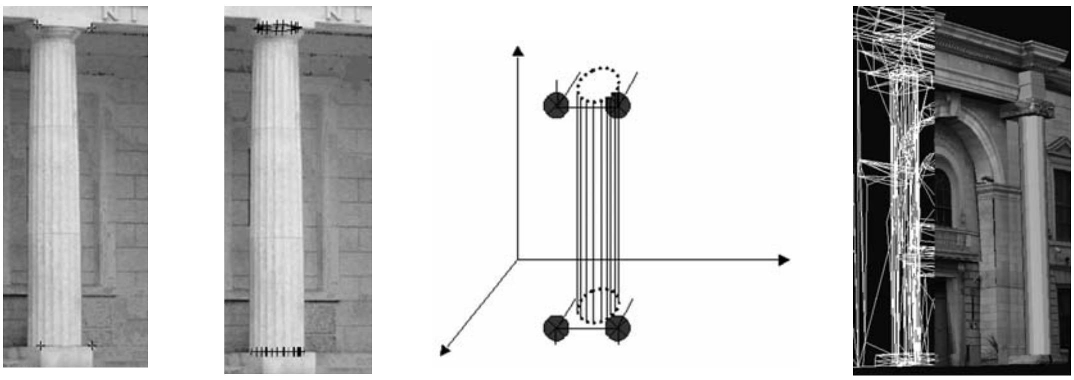


Figure 3. Seed points for ceiling sections.

Figure 4. Automatically generated 3D points.



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

Figure 5. (a) Four seed points are extracted on the base and crown, (b and c) column points are added automatically, (d) column model.

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

Figure 6. Automatic point extraction on arches (a). Constructing blocks (b). Seed points (c). Detected edge (d).

3.4 *Texturing and visualization*

In many applications like particle tracking, fog, clouds or water visualization and with large amount of points, the data can be visualized by just drawing all the samples. However, for some objects (and not very dense point clouds) this technique does not give good results and does not provide realistic visualization. Moreover the visualization of a 3D model is often the only product of interest for the external world and remains the only possible contact with the model. Therefore a realistic and accurate visualization is often required.

Generally, after the creation of a triangular mesh, the results are visualized, according to the used package and the requirements, in wireframe, shaded or textured mode. In case of digital terrain model (DTM) other common methods of representation are the contour maps, the color shaded models (hypsometric shading) or the slope maps.

In the photogrammetric community, the first attempts in the visualization of 3D models were done at the beginning of the '90. Small objects (e.g. architectural models, cars, human faces) were displayed in wireframe format, or using CAD packages, while terrain models were visualized in perspective wireframe models with draping of orthophotos or orthophotomaps. Currently, with the increase in size of computer memory, shade and texture are added, but to visualize large data sets, the model has to be simplified. The consequences are that the accuracy of the data is lost (many tools use single precision files) as well as the geo-referencing (most of the software have their own coordinate systems) and that high-resolution textures are discarded (because of the control on the Level of Detail). However, low-quality visualization does not attract the end-user and cannot justify the high cost of producing 3D models. Thus, advanced visualization techniques must be able to visualize large models in real time without apparent loss of details. For example, hierarchical levels of detail (LOD), where objects far away from the viewpoint are rendered with pre-sampled lower resolution representations, is a standard technique that is supported by most scene graph libraries. Hoek & Damon, 2004, review automatic model simplification and run-time LOD techniques. They discuss LOD framework, LOD management, LOD simplification models and metrics for simplification and error evaluation. Another effective performance enhancement technique for both geometry and texture is occlusion culling (e.g. Zhang 1998), which skips objects that are occluded by other objects or surfaces.

4 EXAMPLES

In the following, two image-based 3D modeling examples are presented, firstly to compare automated with manual measurements, secondly to present very accurate and detailed 3D models obtained with interactive modeling methods.

4.1 *Image-based modeling of complex architectures*

We show here examples of relatively complex image-based models. The images were taken during routine tours without any advanced planning of where to take the images. The images were taken by walking around the monument and getting the best view under real conditions, such as presence of tourists, vehicles, and other buildings and objects. Each model usually required 1-2 days of work by one person. The number of points and level of interaction and automation obviously varied significantly from one model to another, however at least 80% of the points were generated automatically. Examples to illustrate different types of details are presented in wireframe, solid model without texture, and solid model with texture (Fig. 7, 8, 9).

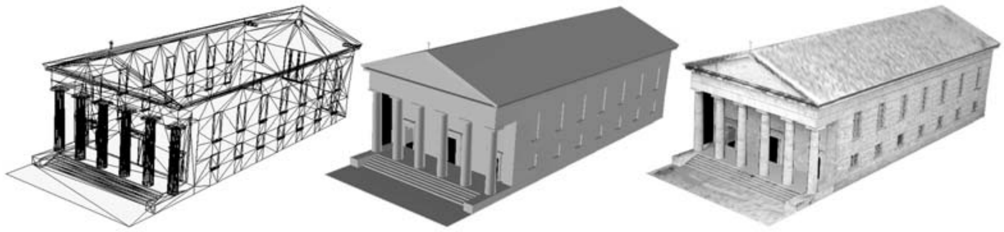


Figure 7. Model of the church in the old fortress in Corfu City, Greece.

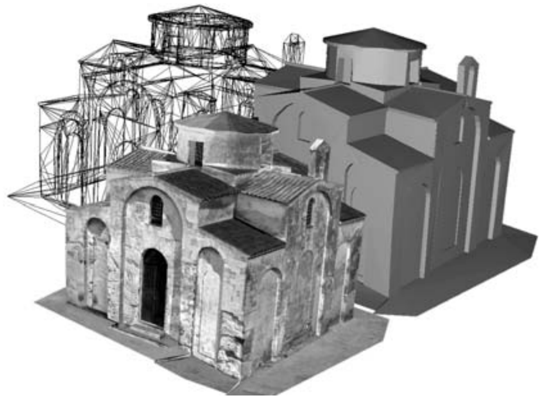


Figure 8. St. Pietro Church, Otranto, Italy.

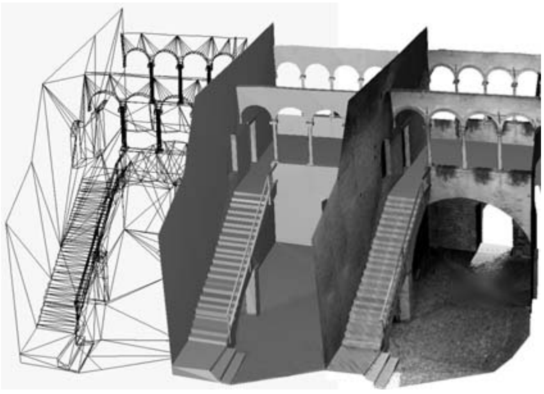


Figure 9. The old loggia at Stenico castle, Italy.

4.2 *Image-based modeling of the big Buddha of Bamiyan*

The modeling of the cultural heritage area of Bamiyan, Afghanistan, is a good example showing the capabilities and achievements of a photogrammetric image-based modeling. The project is a combination of large site landscape modeling with highly detailed modeling of terrestrial objects (the Buddha statues) (Gruen et al. 2004a, Gruen et al. 2004b). In particular, the Buddha statues, demolished in 2001, could be virtually reconstructed only using old images and no other modeling technique could be employed. The Great Buddha was firstly modeled applying automated measurements techniques on a set of available images (Fig. 10). The statue as well as the rock around it are well modeled, but due to the smoothness constraint and grid-point based matching, the small folds on the body of the Buddha were filtered or skipped and they are not visible in the 3D modeled (Fig. 11). Other automated matching algorithms were tested (Gruen et al. 2004a) but none of the procedures could recover the small details of the dress. Therefore only precise manual measurements could correctly reconstruct the exact shape and curvature of the folds of the dress. The recovered point cloud was uneven and sparse, therefore smart algorithms and a lot of manual editing were necessary for the final 3D model (Fig. 12).

5 CONCLUSIONS

Although efforts to automate image-based modeling is continuing, semi-automatic approaches that is designed specifically to take advantage of properties and arrangements common to man-made objects such as architectures are the most effective. In those approaches, parts of the process that can straightforwardly be performed by humans, such as registration, extracting seed points, and topological surface segmentation, remain interactive while those best performed by the computer, such as feature extraction, point correspondence, and modeling of segmented regions should be automated. When conditions allow, the steps of initial point extraction and image registration can be fully automated, although this still requires numerous closely-spaced images. In the mean time, to achieve immediate and useful results, parts of the process necessitate human interaction.

Figure 10. Three metric images used for the image-based modeling of the Great Buddha of Bamiyan and the small details of the folds on the statue (right).

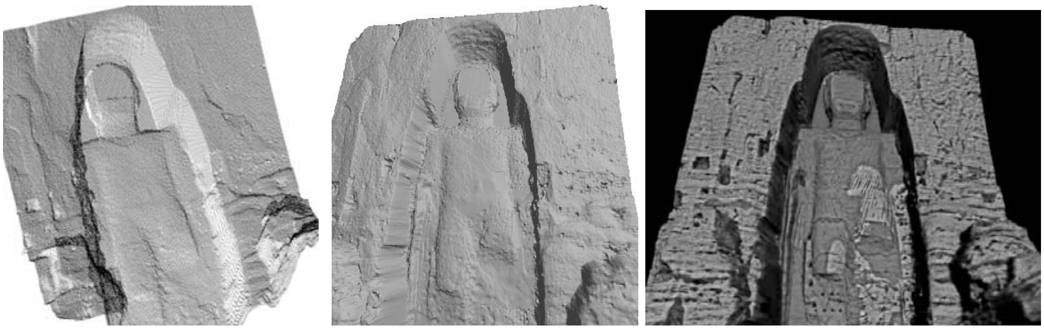


Figure 11. 3D model of the Great Buddha of Bamiyan (displayed as point cloud, shaded and textured model) generated with automated measurements. The automated matching procedure smoothed out all the small details of the folds present on the statue.

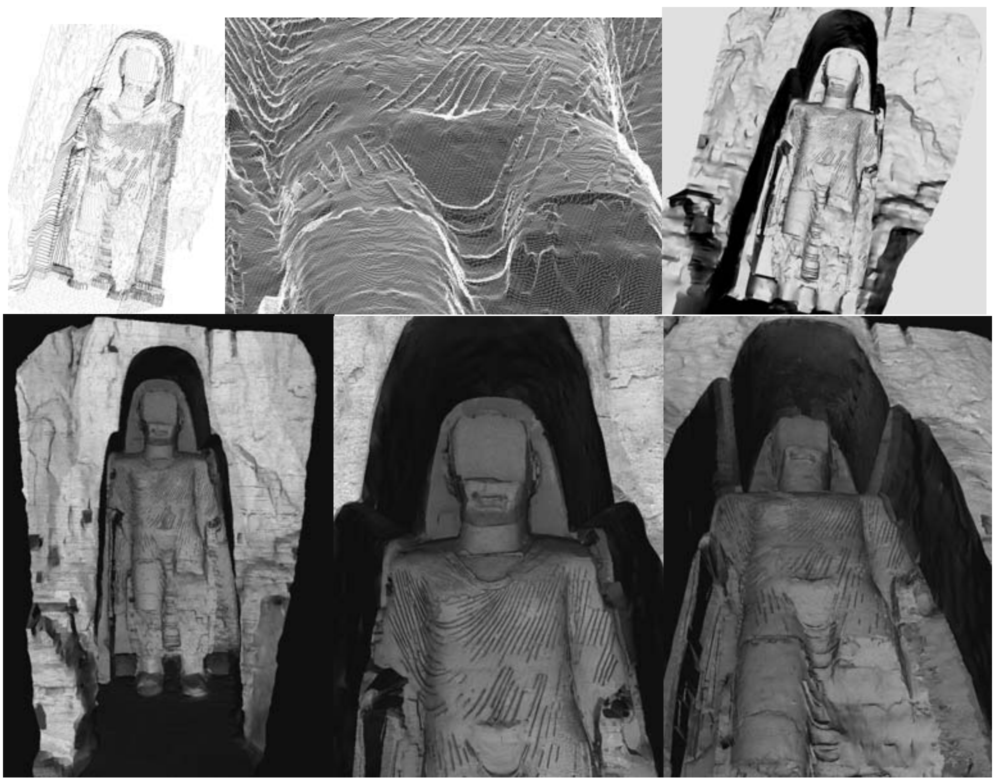


Figure 12. 3D model of the Great Buddha of Bamiyan recovered with manual measurements. The model is shown in point cloud, wireframe (closed view of the folds), shaded and textured mode (lower images).

REFERENCES

Ackermann, F. 1983. High precision digital image correlation. *Photogrammetric Week 1983, Institut für Photogrammetrie, Stuttgart, Schriftenreihe, Heft* 9: 231-242.

Amenta, N., Bern, M. & Kamvyselis, M. 1998. New Voronoi Based Surface Reconstruction Algorithm. In *Proc. of ACM of SIGGRAPH'98, 415-422, Orlando, Florida, 19-24 July*.

Asada, M. 1993. Cylindrical Shape from Contour and Shading without Knowledge of Lighting Conditions or Surface Albedo. *Information Processing Society of Japan Journal* 34(5).

Beraldin, J.A & Blais, F. 2000. Active 3D sensing. *Scuola Normale Superiore Pisa, Centro di Ricerche Informatiche per I Beni Culurali,* Quaderni 10 : 22-46.

Beraldin, J.-A., Picard, M., El-Hakim, S., Godin, G., Latouche, C., Valzano, V. & Bandiera, A. 2002. Exploring a Byzantine crypt through a high-resolution texture mapped 3D model: combining range data and photogrammetry. *Proc. of the ISPRS/CIPA International Workshop Scanning for Cultural Heritage Recording, 65-72, Corfu, Greece, 1-2 September*.

Beraldin, J.A., Picard, M., El-Hakim, S., Godin, G., Valzano, V. & Bandiera, A. 2005. Combining 3D technologies for cultural heritage interpretation and entertainment. Beraldin, J.A., El-Hakim, S., Gruen, A. & Walton, J. (eds), *Videometrics VIII, SPIE* Vol. 5665: 108-118.

Bern, M. & Eppstein D. 1992. Mesh Generation and Optimal Triangulation. *Computing in Euclidean Geometry*, Du/Hwang, (eds.), *World Scientific, Lecture Notes Series on Computing*, Vol. 1: 23-90.

Blais, F. 2004. Review of 20 years of range sensor development, *J. Electronic Imaging* 13(1): 232-240.

Boissonnat, J.D. 1984. Geometric structures for three-dimensional shape representation. *ACM Transactions on Graphics* 3(4): 266–286.

Brinkley, J.F. 1985. Knowledge-driven ultrasonic three-dimensional organ modeling. *IEEE Transaction on PAMI* 7(4): 431–441.

Brown, L.G., 1992: A survey of image registration techniques. *Computing survey*, 24(4): 325-376.

Chen, F., Brown, G.M. & Song, M. 2000. Overview of three-dimensional shape measurement using optical methods. *Optical Engineering* 39: 10-22.

Curless, B. & Levoy, M. 1996. A volumetric method for building complex models from range images. In *Proc. of ACM SIGGRAPH'96, 303-312, New Orleans, USA, 4-9 August.*

D'Apuzzo, N. 2003. *Surface Measurement and Tracking of Human Body Parts from Multi Station Video Sequences*. Ph.D. Thesis, Department of Civil, Environmental and Geomatics Engineering, ETH Zurich, Switzerland, Mitteilungen No. 81.

Debevec, P.E., Taylor, C.J. & J. Malik, 1996. Modelling and rendering architecture from photographs: A hybrid geometry and image-based approach. *Proc. of ACM SIGGRAPH'96,11–20, New Orleans, USA, 4-9 August.*

Dey, T.K. & Giesen, J. 2001. Detecting undersampling in surface reconstruction. In *Proc. of the 17th Symposium of Computational Geometry, 257-263, Medford, USA, 3-5 June.*

Dick, A.R., Torr, P.H., Ruffle, S.J. & Cipolla, R. 2001. Combining single view recognition and multiple view stereo for architectural scenes. In *Proc. of the 8th IEEE International Conference on Computer Vision (ICCV 01), 268-274, Vancouver, Canada, 9-12 July.*

Dhond, U. and Aggarwal, J.K., 1989. Structure from stereo - a review. *IEEE Transaction on System, Man and Cybern.*, 19(6): 1489-1510.

Edelsbrunner H. & Mucke, E. 1994. Three Dimensional Alpha Shapes. *ACM Transactions on Graphics* 13(1): 43-72.

Edelsbrunner, H. 2001. *Geometry and Topology for Mesh Generation,* Cambridge Monographs on Applied and Computational Mathematics, Vol. 6, Cambridge University Press, UK.

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

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

Ferrari, V., Tuytelaars, T. & Van Gool, L. 2003. Wide-baseline muliple-view correspondences. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 718-728, Madison, USA, 16-22 June.*

Fitzgibbon, A. & Zisserman, A. 1998. Automatic 3D model acquisition and generation of new images from video sequence. In *Proc. of the European Signal Processing Conference, 1261-1269, Rhodes, Greece, 8 – 11 September.*

Förstner, W. 1982. On the geometric precision of digital correlation. *International Archives of Photogrammetry and Remote Sensing*, Vol. 26, Part 3: 150-166.

Gartner, H., Lehle, P. & Tiziani, H.J. 1995. New, high efficient, binary codes for structured light methods. *SPIE Proceedings*, Vol. 2599: 4-13.

Gotsman, C. & Keren, D. 1998. Tight Fitting of Convex Polyhedral Shapes. *International Journal of Shape Modeling* 4(3-4): 111-126.

Grün, A. 1985. Adaptive least square correlation: a powerfull image matching technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography* 14(3): 175-187.

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

Guidi, G., Beraldin, J-A., Ciofi, S. & Atzeni, C. 2003. Fusion of range camera and photogrammetry: a systematic procedure for improving 3D models metric accuracy, *IEEE Trans. on Systems, Man and Cybernetics* 33(4): 667-676.

Hastie, T. & Stuetzle W. 1989. Principal curves. *Journal of the American Statistical Association* 84: 502–516.

Healey, G. & Binford, T.O. 1987. Local Shape from Specularity. In *Proc. of the International Conference on Computer Vision, 151-160, June, London, UK.*

Heok, T.K. & Damen, D. 2004. A review of level of detail. In *Proc. of the IEEE International Conference on Computer Graphics, Imaging and Visualization, 70-75, Penang, Malaysia, 26–29 July*.

Hoppe, H., DeRose, T., Duchamp, T., McDonald, J. & Stuetzle, W. 1992. Surface reconstruction from unorganized points. In *Proc. ACM of SIGGRAPH'92, 71-78, Chicago, USA, 26 - 31 July*.

Horn, B.K.P. & Brooks, M.J. 1989. *Shape from Shading*. MIT Cambridge.

Isselhard, F., Brunnett, G., Schreiber, T., 1997. Polyhedral reconstruction of 3D objects by tetrahedra removal. Technical report No. 288/97, Fachbereich Informatik, University of Kaiserslautern, Germany.

Kender, J.R. 1978. Shape from Texture. In *Proc. DARPA Image Understanding Workshop: 134-138.*

Kersten, T., Pardo, C.A. & Lindstaedt, M. 2004. 3D acquisition modeling and visualization of north German castles by digital architectural Photogrammetry. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 35, Part B2: 126-132.

Lee, S.C. & Nevatia, R. 2003. Interactive 3D Building Modeling Using a Hierarchical Representation, *IEEE Workshop on Higher-Level Knowledge in 3D Modeling and Motion (HLK) part of the 9^{th} International Conference on Computer Vision:* 58-65.

Liebowitz, D., A. Criminisi, A. & Zisserman, A. 1999. Creating Architectural Models from Images, *EUROGRAPHICS '99 , Vol. 18, 39-50, Milano, Italy, 7-11 September*.

Lorensen, W.E. & Cline, H.E. 1987. Marching cubes: a high resolution 3d surface reconstruction algorithm. Computer Graphics 21(4):163-169.

Maas, H.G. 1992. Robust Automatic Surface Reconstruction with Structured Light. *International Archives of Photogrammetry and Remote Sensing*, Vol. 24, Part B5: 102-107.

Mencl, R. 2001. *Reconstruction of Surfaces from Unorganized 3D Points Clouds*. PhD Thesis, University of Dortmund, Germany.

Merriam, M. 1992. Experience with the cyberware 3D digitizer. *Proc. of the National Computer Graphics Association*: 125–133, March.

Miller, J.V., Breen, D., Lorensen, W., O'Bara, R. & Wozny, M. 1991. Geometrically deformed models: A method for extracting closed geometric models from volume data. *Computer Graphics* 25(4): 217–226.

Moore D. & Warren, J. 1990. Approximation of dense scattered data using algebraic surfaces. Technical Report: 90-135, Rice University, USA.

Muraki, S. 1991. Volumetric shape description of range data using "blobby model". *Computer Graphics* 25(4): 217–226.

Nister, D. 2004. Automatic Passive Recovery of 3D from Images and Video. IEEE *Proceedings of the 2nd Intl Symp 3D Data Processing, Visualization, and Transmission (3D Data Processing, Visualization and Transmission 2004), 438-445, 6-9 September, Thessaloniki, Greece.*

Pollefeys, M., Koch, R. & Van Gool, L., 1999. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *Int. Journal of Computer Vision*, 32(1): 7-25

Pritchett, P. & Zisserman, A. 1998. Matching and Reconstruction from Widely Separated Views. Koch, R. & Van Gool, L. (eds.), *3D Structure from Multiple Images of Large-Scale Environments*, LNCS 1506: 78-92.

Roth, G. & Whitehead, A. 2000. Using projective vision to find camera positions in an image sequence. In *Proc. of 13th Vision Interface Conference, 87-94, Montreal, Canada, May.*

Sablatnig, R. & Menard, C. 1997. 3D Reconstruction of Archaeological Pottery using Profile Primitives. Sarris N. & Strintzis M.G. (eds.), In *Proc. of the International Workshop on Synthetic-Natural Hybrid Coding and Three-Dimensional Imaging*, 93-96.

Scharstein, D. & Szeliski, R., 2002: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. Journal of Computer Vision*, 47(1/2/3):7-42.

Schindler, K. & Bauer, J. 2003. A model-based method for building reconstruction. In *Proc. of the International Conference on Computer Vision workshop on Higher-Level Knowledge in 3D Modeling and Motion (HLK'03), 74-82, Nice, France.*

Sequeira V. & Ng K. 1999. Automated reconstruction of 3D models from real environments. *ISPRS Journal for Photogrammetry and Remote Sensing* 54(1): 1-22.

Streilein, A. 1994. Towards Automation in Architectural Photogrammetry: CAD-Based 3D-Feature Extraction. *ISPRS Journal of Photogrammetry & Remote Sensing* 49(5): 4-15.

Terzopoulos, D. 1988. The Computation of Visible Surface Representation. *IEEE Transactions on PAMI* 10(4): 417-438.

Ulupinar F. & Nevatia R. 1995. Shape from Contour: Straight Homogeneous Generalized Cylinders and Constant Cross Section Generalized Cylinders. *IEEE Transaction on Pattern Analysis Machine Intelligence* 17(2): 120-135.

Van den Heuvel, F.A., 1999a. A line-photogrammetric mathematical model for the reconstruction of polyhedral objects. In El-Hakim, Gruen (eds.), *SPIE Proceedings, Videometrics VI*, Vol. 3641: 60-71.

Xiao, J. & Shah, M. 2003. Two-Frame Wide Baseline Matching. In *Proc. of 9th IEEE International Conference on Computer Vision*, Vol. 1: 603-610.

Wahl, F.M. 1984. A Coded Light Approach for 3-Dimensional Vision. *IBM Research Report*, RZ 1452.

Werner, T. & Zisserman, A. 2002. New technique for automated architectural reconstruction from photographs. In *Proc. of the 7th European Conference Computer Vision, Vol. 2, 541-555, Copenhagen, May*.

Wilczkowiak, M., Trombettoni, G., Jermann, C., Sturm, P. & Boyer, E. 2003. Scene modeling based on constraint system decomposition techniques. In *Proc. of the 9th IEEE International Conference on Computer Vision (ICCV 03), 1004-1010, Nice, France, 14-17 October.*

Winkelbach, S. & Wahl, F.M. 2001. Shape from 2D Edge Gradient. *Pattern Recognition*, Lecture Notes in Computer Science 2191. Berlin: Springer.

Zhang, H. 1998. *Effective occlusion culling for the interactive display of arbitrary models*. Ph.D. Thesis, Department of Computer Science, UNC Chapel Hill.

WEB LINKS

3DMax: http://www.discreet.com [May 2005]
Breuckmann: http://www.breuckmann.com [May 2005]
Canoma: http://www.metacreations.com/products/canoma [May 2005]
Cyberware: http://www.cyberware.com [May 2005]
ImageModeler: http://www.realviz.com [May 2005]
IWitness: http://www.iwitnessphoto.com [May 2005]
Lightwave: http://www.newtek.com [May 2005]
Maya: http:// www.aliaswavefront.com/ [May 2005]
PhotoModeler: http://www.photomodeler.com [May 2005]
PolyWorks: http://www.innovmetric.com [May 2005]
ShapeCapture: http://www.shapecapture.com [May 2005]
ShapeGrabber: http://www.shapegrabber.com [May 2005]