The Photogrammetric Record 21(115): 269–291 (September 2006)

IMAGE-BASED 3D MODELLING: A REVIEW

Fabio Remondino (fabio@geod.baug.ethz.ch)
Swiss Federal Institute of Technology (ETH), Zurich

Sabry El-Hakim (sabry.el-hakim@nrc-cnrc.gc.ca)
National Research Council, Ottawa, Canada

Abstract

In this paper the main problems and the available solutions are addressed for the generation of 3D models from terrestrial images. Close range photogrammetry has dealt for many years with manual or automatic image measurements for precise 3D modelling. Nowadays 3D scanners are also becoming a standard source for input data in many application areas, but image-based modelling still remains the most complete, economical, portable, flexible and widely used approach. In this paper the full pipeline is presented for 3D modelling from terrestrial image data, considering the different approaches and analysing all the steps involved.

Keywords: calibration, orientation, visualisation, 3D reconstruction

Introduction

Three-dimensional (3D) modelling of an object can be seen as the complete process that starts from data acquisition and ends with a 3D virtual model visually interactive on a computer. Often 3D modelling is meant only as the process of converting a measured point cloud into a triangulated network (‘‘mesh’’) or textured surface, while it should describe a more complete and general process of object reconstruction. Three-dimensional modelling of objects and scenes is an intensive and long-lasting research problem in the graphic, vision and photogrammetric communities. Three-dimensional digital models are required in many applications such as inspection, navigation, object identification, visualisation and animation. Recently it has become a very important and fundamental step in particular for cultural heritage digital archiving. The motivations are different: documentation in case of loss or damage, virtual tourism and museum, education resources, interaction without risk of damage, and so forth. The requirements specified for many applications, including digital archiving and mapping, involve high geometric accuracy, photo-realism of the results and the modelling of the complete details, as well as the automation, low cost, portability and flexibility of the modelling technique. Therefore, selecting the most appropriate 3D modelling technique to satisfy all requirements for a given application is not always an easy task.

Digital models are nowadays present everywhere, their use and diffusion are becoming very popular through the Internet and they can be displayed on low-cost computers. Although it seems easy to create a simple 3D model, the generation of a precise and photo-realistic computer model of a complex object still requires considerable effort.

The most general classification of 3D object measurement and reconstruction techniques can be divided into *contact* methods (for example, using coordinate measuring machines, callipers, rulers and/or bearings) and *non-contact* methods (X-ray, SAR, photogrammetry, laser scanning). This paper will focus on *modelling from reality* (Ikeuchi and Sato, 2001) rather than computer graphics creation of artificial world models using graphics and animation software such as 3DMax, Lightwave or Maya. Here and throughout this paper, all proprietary names and trade marks are acknowledged; a list of websites providing details of many of these products is provided at the end of the paper. Starting from simple elements such as polygonal boxes, such packages can subdivide and smooth the geometric elements by using splines and thus provide realistic results. This kind of software is mainly used for movie production, games, architectural and object design.

Nowadays the generation of a 3D model is mainly achieved using non-contact systems based on light waves, in particular using active or passive sensors (Fig. 1).

In some applications, other information derived from CAD models, measured surveys or GPS may also be used and integrated with the sensor data. Active sensors directly provide range data containing the 3D coordinates necessary for the network (mesh) generation phase. Passive sensors provide images that need further processing to derive the 3D object coordinates. After the measurements, the data must be structured and a consistent polygonal surface is then created to build a realistic representation of the modelled scene. A photo-realistic visualisation can afterwards be generated by texturing the virtual model with image information.

Considering active and passive sensors, four alternative methods for object and scene modelling can currently be distinguished:

(1) *Image-based rendering (IBR)*. This does not include the generation of a geometric 3D model but, for particular objects and under specific camera motions and scene con-

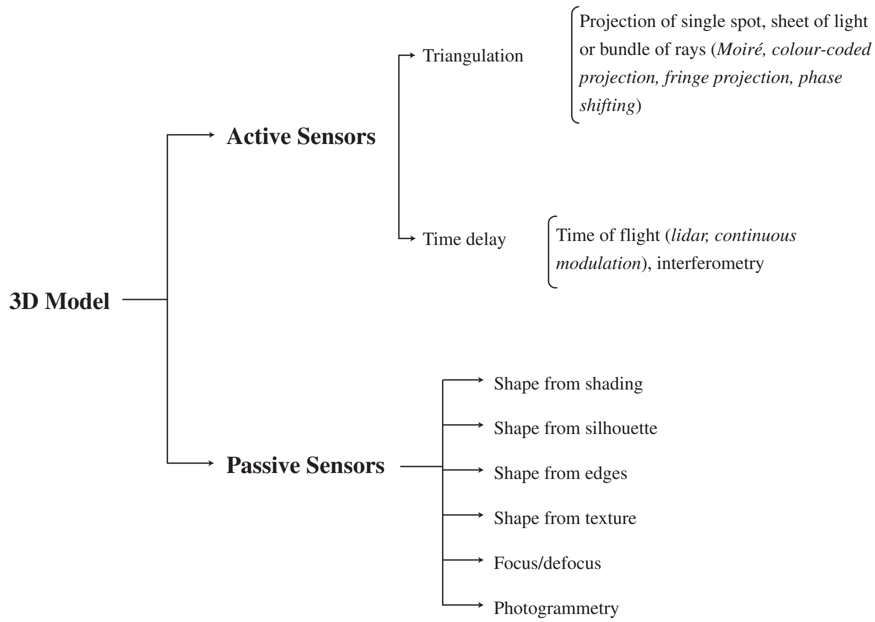

FIG. 1. Three-dimensional acquisition systems for object measurement using non-contact methods based on light waves.

ditions, it might be considered a good technique for the generation of virtual views (Shum and Kang, 2000). IBR creates novel views of 3D environments directly from input images. The technique relies on either accurately knowing the camera positions or performing automatic stereomatching 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 affect the output. The ability to move freely into the scene and view objects from any position may be limited depending on the method used. Therefore, the IBR method is generally only used for applications requiring limited visualisation.

(2) *Image-based modelling (IBM)*. This is the widely used method for geometric surfaces of architectural objects (Streilein, 1994; Debevec et al., 1996; van den Heuvel, 1999; Liebowitz et al., 1999; El-Hakim, 2002) or for precise terrain and city modelling (Grün, 2000). In most cases, the most impressive and accurate results still remain those achieved with interactive approaches. IBM methods (including photogrammetry) use 2D image measurements (correspondences) to recover 3D object information through a mathematical model or they obtain 3D data using methods such as shape from shading (Horn and Brooks, 1989), shape from texture (Kender, 1981), shape from specularity (Healey and Binford, 1987), shape from contour (medical applications) (Asada, 1987; Ulupinar and Nevatia, 1995) and shape from 2D edge gradients (Winkelbach and Wahl, 2001). Passive image-based methods acquire 3D measurements from multiple views, although techniques to acquire three dimensions from single images (van den Heuvel, 1998; El-Hakim, 2001; Al Khalil and Grussenmeyer, 2002; Zhang et al., 2002; Remondino and Roditakis, 2003) are also necessary. IBM methods use projective geometry (Nister, 2004; Pollefeys et al., 2004) or a perspective camera model. They are very portable and the sensors are often low-cost.

(3) *Range-based modelling*. This method directly captures the 3D geometric information of an object. It is based on costly (at least for now) active sensors and can provide a highly detailed and accurate representation of most shapes. The sensors rely on artificial lights or pattern projection (Rioux et al., 1987; Besl, 1988). Over many years, structured light (Maas, 1992; Gaertner et al., 1996; Sablatnig and Menard, 1997), coded light (Wahl, 1984) or laser light (Sequeira et al., 1999) has been used for the measurement of objects. In the past 25 years many advances have been made in the field of solid-state electronics and photonics and many active 3D sensors have been developed (Blais, 2004). Nowadays many commercial solutions are available (including Breuckmann, Cyberware, Cyrax, Leica, Optech, ShapeGrabber, Riegl and Z + F), based on triangulation (with laser light or stripe projection), time-of-flight, continuous wave, interferometry or reflectivity measurement principles. They are becoming a very common tool for the scientific community but also for non-expert users such as cultural heritage professionals. These sensors are still expensive, designed for specific ranges or applications and they are affected by the reflective characteristics of the surface. They require some expertise based on knowledge of the capability of each different technology at the desired range, and the resulting data must be filtered and edited. Most of the systems focus only on the acquisition of the 3D geometry, providing only a monochrome intensity value for each range value. Some systems directly acquire colour information for each pixel (Blais, 2004) while others have a colour camera attached to the instrument, in a known configuration, so that the acquired texture is always registered with the geometry. However, this approach may not provide the best results since the ideal conditions for taking the

images may not coincide with those for scanning. Therefore, the generation of realistic 3D models is often supported by textures obtained from separate high-resolution colour digital cameras (Beraldin et al., 2002; Guidi et al., 2003). The accuracy at a given range varies significantly from one scanner to another. Also, due to object size, shape and occlusions, it is usually necessary to perform multiple scans from different locations to cover every part of the object: the alignment and integration of the different scans can affect the final accuracy of the 3D model. Furthermore, long-range sensors often have problems with edges, resulting in blunders or smoothing effects. On the other hand, for small and medium size objects (up to the size of a human or a statue) range-based methods can provide accurate and complete details with a high degree of automation (Beraldin et al., 1999).

(4) *Combination of image- and range-based modelling*. In many applications, a single modelling method that satisfies all the project requirements is still not available. Different investigations on sensor integration have been performed in El-Hakim and Beraldin (1994, 1995). Photogrammetry and laser scanning have been combined in particular for complex or large architectural objects, where no technique by itself can efficiently and quickly provide a complete and detailed model. Usually the basic shapes such as planar surfaces are determined by image-based methods while the fine details such as reliefs employ range sensors (Flack et al., 2001; Sequeira et al., 2001; Bernardini et al., 2002; Borg and Cannataci, 2002; El-Hakim et al., 2004; Beraldin et al., 2005).

Comparisons between range-based and image-based modelling are reported in Böhler and Marbs (2004), Kadobayashi et al. (2004), Böhler (2005) and Remondino et al. (2005). At the moment it can safely be said that, for all types of objects and sites, 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.

In the next sections, only the terrestrial image-based 3D modelling problem for close range applications will be discussed in detail.

Terrestrial Image-Based 3D Modelling

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

For many years photogrammetry dealt with the precise 3D reconstruction of objects from images. Although precise calibration and orientation procedures are required, suitable commercial packages are now available. They are all based on manual or semi-automated measurements (Australis, Canoma, ImageModeler, iWitness, PhotoGenesis, PhotoModeler, ShapeCapture). After the tie point measurement and bundle adjustment phases, they allow sensor calibration and orientation data and 3D object point coordinates, as well as wire-frame or textured 3D models, to be obtained from multi-image networks.

The overall image-based 3D modelling process consists of several well-known steps: design (sensor and network geometry); 3D measurements (point clouds, lines, etc.); structuring and modelling (segmentation, network/mesh generation, etc.); texturing and visualisation.

In the remainder of the paper, attention is focused on the details of 3D modelling from multiple images.

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

(1) *Approaches that try to obtain a 3D model of the scene from uncalibrated images automatically* (also called "shape from video" or "VHS to VRML" or "Video-To-3D"). Many efforts have been made to completely automate the process of taking images, calibrating and orienting them, recovering the 3D coordinates of the imaged scene and modelling them, but while promising, the methods are thus far not always successful or proven in practical applications. The fully automated procedure, widely reported in the computer vision community (Fitzgibbon and Zisserman, 1998; Nister, 2004; Pollefeys et al., 2004), starts with a sequence of closely separated images taken with an uncalibrated camera. The system automatically extracts points of interest (such as corners), sequentially matches them across views and 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 must not vary significantly, thus the images must be taken at short intervals. The first two images are generally used to initialise the sequence. This is done on a projective geometry basis and it is usually followed by a bundle adjustment. A "self-calibration" (or auto-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 surface model is then automatically generated. In case of complex objects, further matching procedures are applied in order to obtain dense depth maps and a complete 3D model. See Scharstein and Szeliski (2002) for a recent overview of dense stereo-correspondence algorithms. Some approaches have also been presented for the automated extraction of image correspondences between wide baseline images (Pritchett and Zisserman, 1998; Matas et al., 2002; Ferrari et al., 2003; Xiao and Shah, 2003; Lowe, 2004), but their reliability and applicability for automated image-based modelling of complex objects is still not satisfactory as they yield mainly a sparse set of matched feature points. However, dense matching results under wide baseline conditions were reported in Strecha et al. (2003) and Megyesi and Chetverikov (2004). Automated image-based modelling methods rely on features that can be extracted from the scene and automatically matched, therefore occlusions, illumination changes, limited locations for the image acquisition and untextured surfaces are problematic. However, recent invariant point detector and descriptor operators, such as the SIFT operator (Lowe, 2004), proved to be more robust under large image variations. Another problem is that it is very common that an automated process ends up with areas containing too many features that are not all required for modelling while there are areas without any features or with a minimum number that cannot produce a complete 3D model. Automated processes require highly structured images with good texture, high frame rate and uniform camera motion, otherwise they will inevitably fail. Image configurations that lead to ambiguous projective reconstructions have been identified in Hartley (2000) and Kahl et al. (2001) while self-calibration-critical motions have been studied in Sturm (1997) and Kahl et al. (2000). The level of automation is also strictly related to the quality (precision) of the required 3D model. Automated reconstruction methods, even if able to recover the complete 3D geometry of an object, reported errors up to 5% (accuracy of *c.* 1:20) (Pollefeys et al., 2004), limiting their use to applications that require only "nice-looking" partial 3D models. Furthermore, post-processing operations are often required, which means that user interaction is still needed. Therefore, fully automated procedures are generally reliable and limited to finding point correspondences and camera poses

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase® Smarter legal research.