# From images to models (and beyond): a personal retrospective

Richard Szeliski
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052-6399
szeliski@microsoft.com

## Abstract

This paper surveys a number of techniques for extracting 3D geometry and photometry from multiple images. The paper describes the recovery of volumetric models from silhouettes, the extraction of 3D surface curves from profiles, the extraction and integration of 3D surface points from stereo, and a dynamic physically-based surface model for integrating the outputs of these procedures. It also presents recent work in image-based rendering, i.e., the representation of objects and scenes using collections of images which are warped and blended to create novel views.

## 1 Introduction

The reconstruction of 3D models from imagery has long been one of the central topics in computer vision [6, 5, 19]. Initially, it was believed that high-level 3D models were a prerequisite for image and scene understanding, as well as for robotics tasks such as navigation and manipulation. Today, it is widely accepted that visual information can be extracted and used at many levels. For instance, recognition tasks can often be solved using a collection of 2D views or features.

The last few years have seen a renewed interest in the acquisition of photorealistic 3D models for applications such as computer graphics, special effects, and the creation of virtual environments. Active range sensors such as laser scanners provide one way of acquiring such data [10, 43]. Passive image-based reconstruction techniques provide a potentially lower-cost alternative which can be applied to any sized object. These techniques, which reconstruct 3D shape from a number of views of an object or scene, also have the potential to capture the full visual richness of a complex object or scene through the use of view interpolation and view-dependent texture maps [24, 17].

This paper surveys a number of 3D reconstruction techniques which I have developed over the last ten years. My interest in 3D modeling began with an investigation of multiframe stereo algorithms [37]. I then developed a number of full 3D reconstruction algorithms using both surface [54, 65] and volumetric [55] models. I also became interested in the problem of surface reconstruction from sparse data [62, 64], and structure from motion techniques for recovering camera motion [59]. More recently, I have been investigating techniques for building large panoramic scene descriptions [56, 60, 57, 30] using a combination of projective structure from motion and multiframe stereo techniques. Most recently, I have been involved in research into image-based rendering [24].

This paper parallels the chronology of my research into these topics. Section 2 presents a technique for building volumetric models from binary foreground/background silhouettes. Section 3 describes our algorithm for estimating surface shape from tracked contours. Section 4 describes our method for integrating narrow-baseline stereo estimates using a cloud of 3D points with associated uncertainties. Section 5 discusses our technique for extracting and modeling 3D surfaces of arbitrary topology using collections of oriented surface elements. Section 6 discusses some approaches and recent results in extracting 3D surface color distributions (i.e., texture maps) from multiple images. Section 7 presents some recent results on image-based rendering. Section 8 briefly mentions issues related to camera calibration and structure from motion. We close with a discussion of potential applications and topics for future research.

## 2 Volumes from silhouettes

The first three reconstruction algorithms described in this paper use a controlled motion platform and a stationary video camera connected to a frame grabber. The motion platform is an inexpensive spring-loaded microwave turntable, to which we affixed a paper strip with an 8-bit pattern which encodes the turntable's angular position (Figure 1a) [54]. As the object spins on the turntable, video images are captured, the turntable angle is computed, and selected images (typically at 3° to 10° spacing) are retained for further processing. The relative position of the video camera and the turntable, and also the internal calibration parameters of the camera, must be known in order to recover 3D shape. We perform this calibration prior to model acquisition by placing a simple hexagonal pattern on the top of the