# Photorealistic Image Based Objects from Uncalibrated Images

Miguel Sainz*
Computer Graphics Lab
Information and Computer Science
University of California, Irvine

Renato Pajarola†
Computer Graphics Lab
Information and Computer Science
University of California, Irvine

Antonio Susin‡
Dynamic Simulation Lab
Applied Mathematics Dept.
Polytechnical University of Catalonia

## Abstract

In this paper we present a complete pipeline for image based modeling of objects using a camcorder or digital camera. Our system takes an uncalibrated sequence of images recorded around a scene, it automatically selects a subset of keyframes and then it recovers the underlying 3D structure and camera path. The following step is a volumetric scene reconstruction performed using a hardware accelerated voxel carving approach. From the recovered voxelized volume we obtain the depth images for each of the reference views and then we triangulate them following a Restricted Quadtree meshing scheme. During rendering, we use a highly optimized approach to combine the available information from multiple overlapped reference images generating a generate a full 3D photo-realistic reconstruction. The final reconstructed models can be rendered in real time very efficiently, making them very suitable to be used to enrich the content of large virtual environments.

**CR Categories:** I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—;

**Keywords:** Volumetric reconstruction, Voxel carving, Hardware acceleration, Overlapping textures, image-based rendering, multiresolution modeling, level-of-detail, hardware accelerated blending.

## 1 Introduction

In this paper we present a complete pipeline for extracting a 3D volumetric representation of an object from a set of calibrated images taken with a digital camera or handheld camcorder. This reconstruction is then processed using a projective texture-mapped depth mesh model description and we provide an efficient rendering algorithm obtaining high quality images in realtime.

Since the very beginning of computer technology, the possibility of reproducing the real world for simulation purposes has been a primary goal of researchers. The growth of computer graphics technology has generated an important demand for more complex and realistic 3D content. However, even though the supporting tools for complex 3D model creation are more powerful (but also more expensive and difficult to use), obtaining realistic models is still difficult and time consuming.

In recent years Image Based Modeling and Rendering techniques have demonstrated the advantage of using real image data to greatly improve rendering quality. New rendering algorithms have been presented that reach photo-realistic quality at interactive speeds when rendering 3D models based on digital images of physical objects and some geometric information (i.e. a geometric proxy). While these methods have emphasized the rendering speed and quality, they generally require an extensive preprocessing in order to obtain well calibrated images and geometric approximations of the target objects. Moreover, most of these algorithms heavily rely on user interaction for the camera calibration and image registration part or require expensive equipment such as calibrated gantries and 3D scanners.

*e-mail: msainz@ics.uci.edu
†e-mail: pajarola@acm.org
‡e-mail: toni.susin@upc.es

## 2 Pipeline Description

Our goal is to extract 3D geometry and appearance information of the target objects in the scene, based on given camera locations and their respective images. Different approaches such as photogrammetry, stereo vision, contour and/or shadow analysis techniques work with similar assumptions. Figure 1 illustrates the block diagram of the proposed pipeline for the image based 3D model reconstruction from images problem

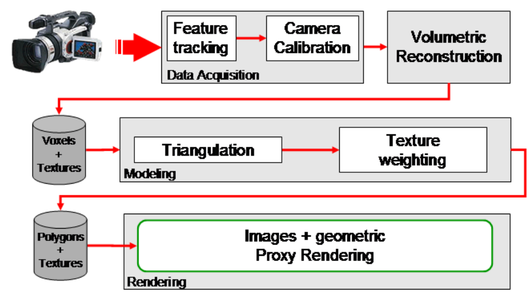


Figure 1: Image Based Modeling pipeline.

The complete pipeline starts with an initial calibration process of the images themselves, in order to recover the camera internal and external parameters. The next step in the pipeline is a scene reconstruction to obtain a complete model representation that can be used to render novel views of the object. Depending on the chosen representation for the model, solutions ranging from point based approaches to complete 3D textured meshes are available in the literature ([Pollefeys 1999], [Sainz 2003]). We propose a novel model representation that consists of set of textured depth meshes obtained from a voxelized reconstruction and that uses the images as overlapping texture maps. During rendering, our approach efficiently combines all the images as view-dependent projected texture maps obtaining photorealistic renders at interactive speeds.

### 2.1 Camera Calibration

The first step of the pipeline consists of recovering the 3D geometry of a scene from the 2D projections of measurements obtained from the digital images of multiple reference views, taking into account the motion of the camera. The proposed calibration approach [Sainz et al. 2003] is based on a divide and conquer strategy that automatically fragments the original sequence into subsequences and, in each of them, a set of key-frames is selected and calibrated up to a scale factor, recovering both camera parameters and structure of the scene. When the different subsequences have been successfully calibrated a merging process groups them into a single set of cameras and reconstructed features of the scene. A final non-linear optimization is performed in order to reduce the overall 2D re-projection error.

### 2.2 Volumetric Scene Reconstruction

In order to reconstruct the volume occupied by the object in the scene we have improved the approach presented in [Sainz et al. 2002], that is based on carving a bounding volume using a color similarity criterion. The algorithm is designed to use hardware accelerated features from the videocard. Moreover, the data structures have been highly optimized in order to minimize run-time memory usage. Additional techniques such as hardware projective texture mapping and shadow maps are used to avoid redundant calculations.

### 2.3 Object Modeling

The final representation of the reconstructed object is based on an efficient depth-image representation and warping technique called *DMesh* ([Pajarola et al. 2003]) which is based on a piece-wise linear approximation of the reference depth-images as a textured and simplified triangle meshes. During rendering the algorithm selects the closest reference views to the novel viewpoint, and it renders them and combines the result using a per-pixel weighted sum of the respective contribution, obtaining the final colored image. This weighted sum and the corresponding final normalization are achieved in real-time using the programmability of the actual GPU's.

### 2.4 Results

We present the *monster* dataset(see Fig. 2) that consists of a set of 16 still images of 1024x1024 pixels each taken from an object on a turntable. A manual tracking process of the fiducials on the surface of the object is performed to obtain the proper calibration of the images using our approach.

The volumetric reconstruction starts with an initial volume of 250 x 225 x 170 (9562500 voxels), and using five manually selected frames from the initial set, it produces in 43 iterations and 3.5 min. a final reconstruction of 1349548 voxels (a 14% of the initial volume). The DMesh final model is presents an average of 36000 faces per reference view and they are rendered and combined at 400 frames per second.

## References


PAJAROLA, R., SAINZ, M., AND MENG, Y. 2003. Dmesh: Fast depth-image meshing and warping. *International Journal of Image and Graphics (IJIG)*, (to appear).

POLLEFEYS, M. 1999. *Self-calibration and metric 3D reconstruction from uncalibrated image sequences*. PhD thesis, K.U. Leuven.

SAINZ, M., BAGHERZADEH, N., AND SUSIN, A. 2002. Hardware accelerated voxel carving. In *Proceedings of 1st Ibero-American Symposium in Computer Graphics*, 289–297.

SAINZ, M., BAGHERZADEH, N., AND SUSIN, A. 2003. Camera calibration of long image sequences with the presence of occlusions. In *Proceedings of IEEE International Conference on Image Processing*.

SAINZ, M. 2003. *3D Modeling from Images and Video Streams*. PhD thesis, University of California Irvine.


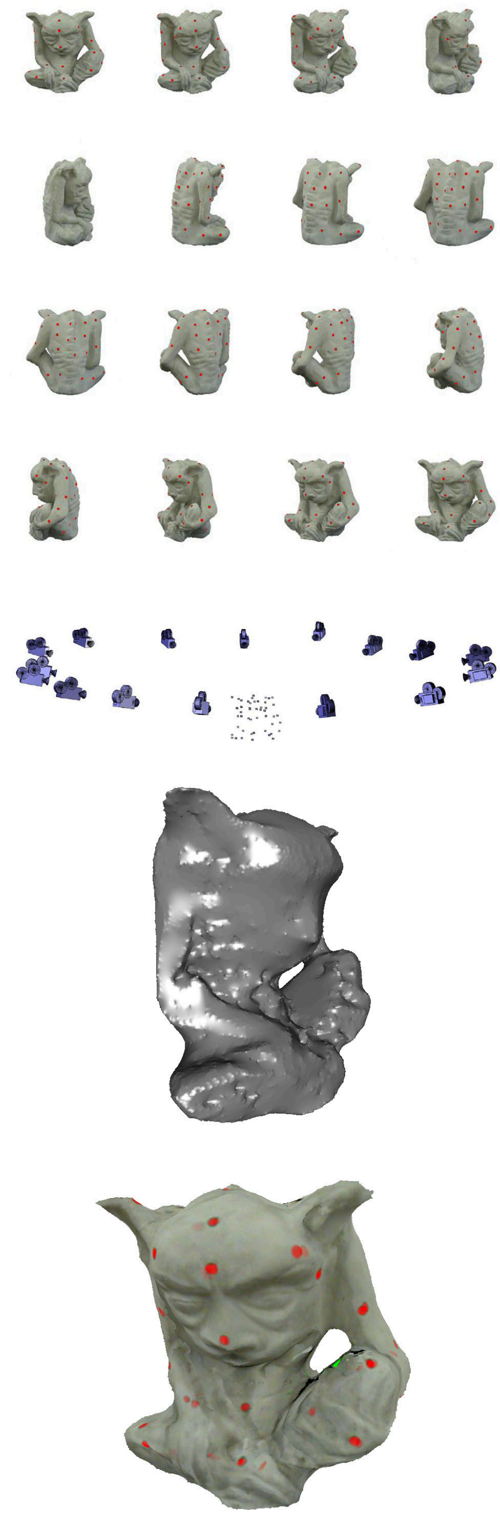

Figure 2: From top to bottom: the original 16 frames of the *monster* dataset; the calibrated camera path; the reconstructed volume without coloring and a novel rendered view of the object using the DMesh approach.