Persepolis: Recovering history with a handheld camera

M. Sainz, A. Susin (†), A. Cervantes and N. Bagherzadeh

Image Based Modeling and Rendering Lab, University of California, Irvine, USA
(†) Departament de Matem'atica Aplicada I, Universitat Polit'ecnica de Catalunya, Spain

Abstract

In this paper we present new improvements to our novel pipeline for image based modeling of objects using a camcorder. Our system takes an uncalibrated sequence of images recorded around a scene, it automatically recovers the underlying 3D structure and camera path and then a volumetric scene reconstruction is performed using a hardware accelerated voxel carving approach. Finally a triangular mesh is obtained and the available information from the images is combined to generate a full 3D photo-realistic reconstruction. As an application, we use this system to reconstruct parts of the archeological site of Persepolis (Iran).

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling, I.4.8 [Image Processing and Computer Vision]: Scene Analysis

1. Introduction

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.

On the other hand, research fields such as archeology are becoming aware of the great advantages of 3D visualization systems [2] and computerized databases of images. Since in their daily routine archeologists take a large amount of images and measurements of their sites for documentation purposes, a system to extract 3D reconstructions of the archeological sites can be very beneficial to enhance the quality of the work as well as its diffusion to the general public.

In this paper we present a method for extracting, meshing and rendering a 3D volumetric representation of an object from a set of unstructured images taken with a still camera or handheld camcorder. As an application, we present our ongoing work of reconstruction of the archeological site of Persepolis (Iran).

Figure 1: *Gate of Xerxes, at the North of the site.*

The magnificent palace complex at Persepolis (see Figure 1) was founded by Darius the Great around 518 B.C., although more than a century passed before it was finally completed. Conceived to be the seat of government for the Achaemenian kings and a center for receptions and ceremonial festivities, the wealth of the Persian empire was evident in all aspects of its construction. The splendor of Persepolis, however, was short-lived; the palaces were looted and burned by Alexander the Great in 331-330 B.C. The ruins were not excavated until the Oriental Institute of the University of Chicago sponsored an archaeological expedition to Persepolis and its environs under the supervision of Professor Ernst Herzfeld from 1931 to 1934, and Erich F. Schmidt from 1934 to 1939.

2. Image Based Reconstruction Method

The main goal of the presented project is to develop a tool, based on image based modeling techniques, that allows the automatic reconstruction of physical objects with all their properties (shape, color and texture) properly recovered. Figure 2 illustrates the block diagram of the suggested pipeline [3] for the 3D model reconstruction from images problem

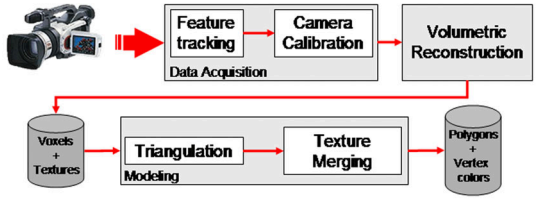

Figure 2: *Image Based Modeling pipeline.*

The input to the proposed system is a video sequence or set of images taken with an off-the-shelf digital camera or camcorder, by moving it around the physical object that is going to be reconstructed. Then, a calibration process reconstructs the 3D structure of the scene and the motion of the camera analyzing how a set of 2D tracked points move in the image sequence. A volumetric reconstruction step fits a volume of voxels to the object in the scene using the information of the calibration and the input images. The last stage of the pipeline is a modeling process that transforms the voxelized volume into a colored mesh suitable to be rendered in any 3D pipeline.

2.1. Camera Calibration

The first step of the pipeline consists of performing a camera calibration process to recover the 3D structure of the scene and the camera motion and internal parameters. More specifically, the goal is to recover the 3D geometry of a scene from the 2D projections obtained from the digital images of multiple reference views, taking into account the motion of the camera. Moreover, since to obtain a robust reconstruction, the image sequence must show the object from different perspectives, self-occlusions are constantly present increasing the difficulty of the problem.

The proposed novel calibration approach [5] 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 objects in the scene we have improved the approach presented in [4], 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.

The proposed implementation of the space carving methodology consists of an iterative algorithm that has an outer loop that performs a progressive plane sweep along one axis of the bounding box that contains the object to be reconstructed until the set of non-carved voxels remains constant. On each iteration of the sweep, only the voxels that belong to the surface of the volume slice are processed. Then, during the color consistency test, each voxel is tested against each of the views to determine the set of contributing reference images. Then the voxel footprint is calculated and the color consistency function decides, whether the voxel is consistent and then kept, or is inconsistent and removed. This color test is performed by correlating each pair of color histograms of the contributing views, and if more than a threshold of images are correlated (i.e. similar color information), the voxel is considered to be consistent.

2.3. Scene Modeling

The final step of the pipeline is to generate a triangular mesh suitable to be rendered in a standard graphics pipeline. We use a variation of the SurfaceNet algorithm [1] that creates a globally smooth surface model from the binary segmented volume data retaining the detail present in the original segmentation. The mesh is constructed by linking nodes on the surface of the binary-segmented volume and relaxing node positions to reduce energy in the surface net while constraining the nodes to lie within a surface cube defined by the original segmentation.

Finally a color per vertex is assigned by using a weighted average of colors from the reference images that see the vertex being colored. The weight for each view is calculated taking into account the relative orientation of the vertex and the view direction and also the proximity of the viewpoint to the vertex.

3. Results

Several validation and verification tests were performed with images from a home video recording system with auto-focus and auto-exposure enabled. Two datasets are presented here, consisting of video sequences that have been preprocessed in order to remove the interlacing and to enhanced the color saturation, since the original tapes present a high content of

overexposed areas. All the tests and timings have been performed in a PC P4 2.0Ghz with 1Gb of RAM and a NVIDIA GeForce4 Ti 4600 videocard with Detonator drivers 40.03.

The first dataset, the *two-head-horse* (see Fig. 3, left column), contains 702 frames of 720x480 pixels each. A tracking system based on the KLT tracker [6] was used to track a total of 300 measures per frame with replacement of lost features.

The automatic calibration system selects a total of 94 keyframes divided into 4 subsequences that are calibrated and merging with a final overall average 2D reprojection error of 0.6 pixels with a standard deviation of 0.6 pixels. The sequence of keyframes is decimated to a final set of 12 frames that will be used during the volumetric reconstruction. The carving process is set to use an initial resolution of 200x176x138 voxels and it performs 165 iterations in 70 minutes, reducing the occupancy of the initial volume to a 43%, generating a final solid model of 4857600 voxels. If the same reconstruction if performed using half the initial resolution, the computation time is less than 10 minutes. The mesh generation takes 12 seconds to produce a colored triangulated mesh of 203391 vertices and 409978 faces. The rendering time is 6.5 frames per second on the above mentioned platform without any level-of-detail simplification.

The second dataset presented here, the *column-base* (see Fig. 3, right column) consists of a video sequence of 260 frames taken around a column pillar base, and the camera performs a 90 degrees rotation parallel to the ground around the object.

The same tracking approach has been used to find 200 measurements per each of the 640x480 pixels frames. The calibration algorithm divides the complete sequence in two fragments and calibrates and merges them into a final keyframe sequence of 81 frames. The final mean reprojection error after the bundle adjustment is 0.5 pixels with a standard deviation of 0.7 pixels. The volumetric reconstruction starts with an initial volume of 150x98x132 voxels, and using three manually selected frames produces in 134 iterations and 11 min. a final reconstruction of 1117488 voxels (a 57% of the initial volume). The final triangulated model has 93201 vertices and 188162 faces. The computation time of the smoothed and colored mesh is 7.2 seconds.

4. Conclusions

In this paper we have presented a novel complete pipeline for image based modeling that can be used to reconstruct 3D objects and scenes of archeological sites.

In this first attempt to use footage from a field trip to the site of Persepolis we have successfully recovered automatically some elements using the images from a camcorder. Moreover, a final colored geometric model is obtained that can be rendered in a standard graphics pipeline.

Currently we are working towards getting more reconstructed datasets and also to improve speed and reliability of the reconstruction algorithms.

The sensor device used for capturing the images is an off-the-shelf digital camcorder, and we have found out that the automatic settings of this type of cameras are not well suited for the purposes of object reconstruction from images because the automatic exposure compensation plays against the voxel carving algorithm by changing significantly the surface color of the objects. We expect to find a feasible way to compensate this on the actual tapes, and for future acquisitions we will design the proper protocol and sensor adjustments to avoid such problems.

Acknowledgements

This research was partially supported by the National Science Foundation under contract CCR-0083080 and by the Comissio Interdepartamental de Recerca i Innovacio Tecnologica, Gaspar de Portola grant C02-03.

We would like to thank Dr. Farrokh Shadab for kindly providing his personal video tapes of his trip to Persepolis as data for our experiments.

References

1. S. Gibson, “Constrained Elastic SurfaceNets: Generating Smooth Surfaces from Binary Segmented Data”, in *Proc. Medical Image Computation and Computer Assisted Interventions*, pp. 888- 898, October 1998. 2
2. M. Pollefeys, L. Van Gool, M. Vergauwen, K. Cornelis, F. Verbiest, J. Tops, “3D recording for archaeological fieldwork", in *IEEE Computer Graphics and Applications*, **23**(3):20–27, May-June 2003. 1
3. M. Sainz. *3D Modeling from Images and Video Streams*. PhD. Thesis, University of California Irvine, July 2003. 2
4. M. Sainz, N. Bagherzadeh and A. Susin, “Hardware Accelerated Voxel Carving”, in *1st Ibero-American Symposium in Computer Graphics (SIACG 2002)*, Guimaraes, Portugal. pp 289-297, July 2002. 2
5. M. Sainz, A. Susin and N. Bagherzadeh. “Camera Calibration of Long Image Sequences with the Presence of Occlusions”, in *Proc. IEEE International Conference on Image Processing*, September 2003. 2
6. J. Shi and C. Tomasi, “Good Features to Track”, in *Proc IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593-600, 1994. 3

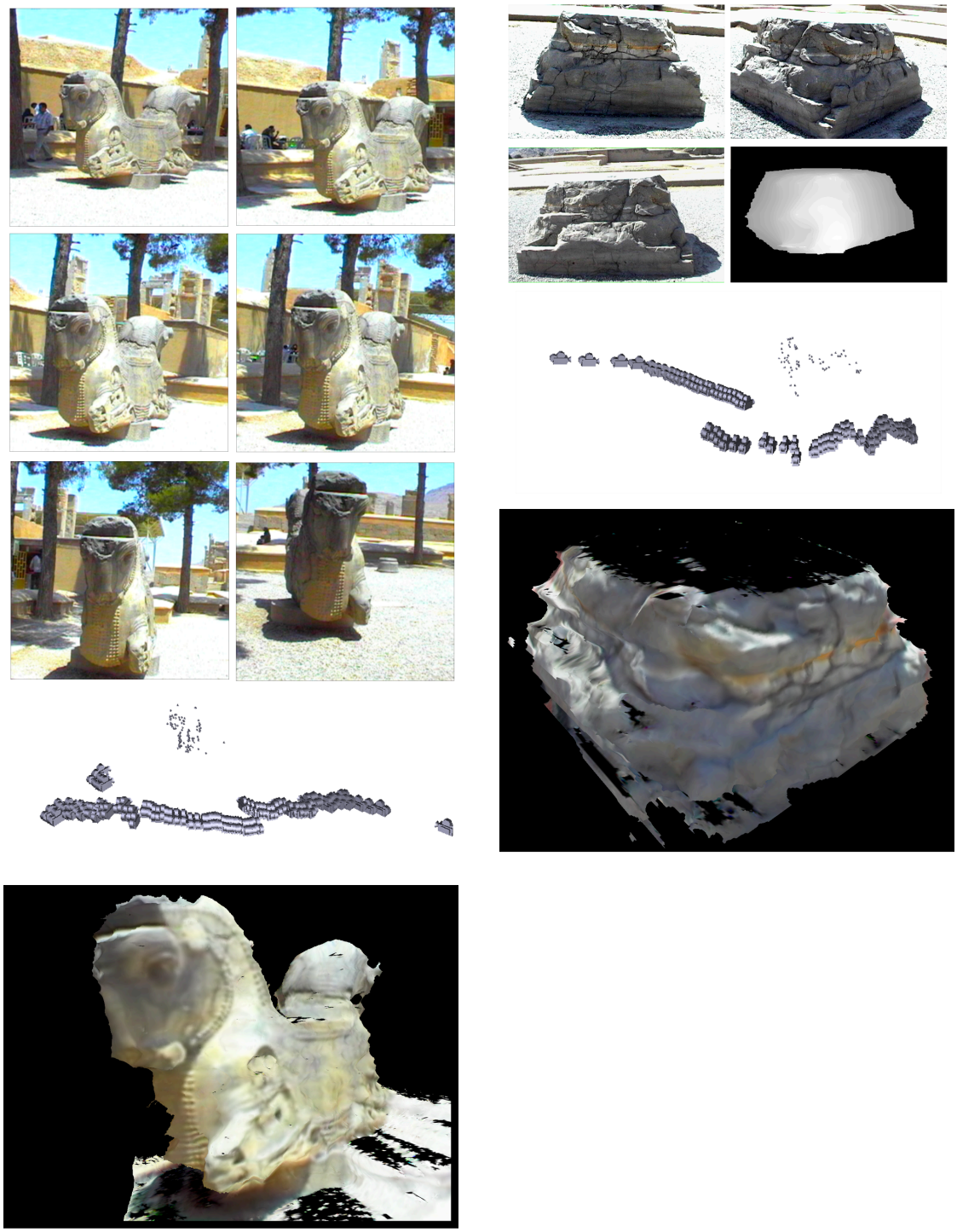

Figure 3: *Reconstruction results. The left column shows some of the* two-head-horse *dataset images, the reconstructed camera path and a novel view rendered using the reconstructed mesh. The right column shows some frames of the* column-base *dataset, a depthmap of the reconstruction, the camera path and a novel rendered view.*