Challenges and advances in interactive modeling

Voicu Popescu, Elisha Sacks, Aliasgar Ganiji

Computer Sciences Department, Purdue University,

250 N University Street, West Lafayette, IN 47907-2066

ABSTRACT

Many important computer graphics applications rely on 3D models of real-world scenes. The current approach for modeling such scenes is first to acquire depth and color data and then to build a model off line. The approach is slow, the acquisition devices are expensive, and the resulting models are incomplete and of uneven quality. We propose an interactive modeling strategy where an operator scans the scene with a handheld device and the model is built incrementally. The evolving model is displayed continually for immediate feedback. The advantages of the interactive modeling approach are short modeling times, good scene coverage, even sampling, and low cost. Implementing the interactive modeling pipeline poses the following challenges: devising an acquisition device that is lightweight, compact and robust, real-time depth extraction, real-time registration, and real-time incremental modeling. In this paper, we analyze these advantages and challenges in light of our experience in designing an interactive modeling system.

1. INTRODUCTION

Computer graphics has traditionally been shielded from the full complexity of the real world by operating on a model. Transforming the model into images is a laborious task, but it is well suited to the talent of computing machinery for fast, reliable quantitative analysis. Although the images produced originally only vaguely resembled the scene they were depicting, they were sufficient to support popular applications that in turn fueled rapid progress in computer graphics. The progress of applications that render real-world scenes eventually slowed down because manual scene modeling could not keep up with the increase in hardware rendering speed. A good way to increase the realism of the rendered images is to automate scene modeling. Automated scene-model acquisition is a difficult inverse problem and the scope of the state of the art is limited. Particularly difficult is the “inside-looking-out” case where the acquisition device operates from within the scene. Prior methods have limited applicability because of the following reasons.

Depth acquisition Image-based rendering (IBR) techniques have shown that scenes can be rendered without any form of depth ([1], [2]); however, in the case of complex scenes, the required ray databases become very large. Hybrid geometry-and-color scene representations are more compact and can be efficiently rendered on current graphics hardware. Acquiring geometry poses the difficult problem of depth acquisition; current computer vision and laser rangefinding solutions produce good depth maps, but require several minutes or even hours for each view.

Scene coverage The acquired scene model should capture all surfaces that can become visible to the application user. The problem of choosing a set of views that satisfies this condition, known as the *view planning problem* or *next-best view problem* ([3], [4], [5]), is challenging for complex scenes because of numerous occlusions. A large number of views is required but the lengthy per-view acquisition time of current systems limits the practical number of views, so a significant fraction of the scene is often missed.

Sampling rate control View planning should also consider sampling quality. A minimum sampling rate must be maintained for high-quality rendering. Sampling rate control is orthogonal to scene coverage. Since the artifacts produced by insufficient coverage are more visible than inadequate sampling, sampling rate control is usually ignored.

View registration The individual views have to be placed in a common coordinate system. Because only a small number of views are acquired, the views differ considerably, so user specified correspondences are needed to derive an initial guess that is refined automatically.

Depth/color registration Most acquisition devices that acquire color do so in a second pass using digital cameras. The color and depth data have to be co-registered.

Model construction The registered data needs to be merged into the final color-and-geometry model by eliminating overlap between views and by triangulating the depth samples.

High cost The need for trained operators, the long scanning times, and the expensive equipment make modeling costly.

To address these challenges, we propose a novel *interactive* form of automated modeling based on the following considerations. For the foreseeable future, the only solution to the scene coverage problem is the acquisition of a large number of views. If this condition is satisfied, the amount of data acquired at each view can be decreased without reducing the overall modeling capability of the system. This substantially reduces the demand on depth acquisition, which can be greatly accelerated to reach interactive rates. This in turn allows for the acquisition of a dense set of views. The coherence between consecutive views enables fast registration and incremental modeling. The resulting modeling pipeline runs at interactive rates. The operator sweeps the scene with a handheld acquisition device and the model is built and displayed in real time. This way, scene coverage and sampling rate control can be assigned to the operator, taking advantage of the human talent for high-level reasoning.

The remainder of this paper is organized as follows. In the next section, we review prior work in interactive modeling. Section 3 describes the challenges of interactive modeling. Sections 4 and 5 describe our handheld interactive modeling system, and section 6 discusses our results and gives directions for future work.

2. PRIOR WORK

Several hand-held devices have recently been described in the literature. One type of device consists of a fixed camera and a mobile light-pattern source. One variant [7] uses a hand-held laser point projector on which three green LED's are mounted. The position of the LED's in the camera frame is used to infer the position and orientation of the laser beam. The red laser blob is detected in the frame and then triangulated as the intersection between the pixel ray and the laser beam. Another variant [6] extracts depth from the shadow of a rod captured by a camera under calibrated lighting. The Autoscan system [8] uses two cameras mounted on a tripod and a hand-held laser point projector. The main problem with these systems is that they are limited to a single view by the fixed camera.

Hebert [9] proposes a system where the operator can freely change the view. The device consists of two cameras and a cross-hair laser light projector. Frame to frame registration is achieved using a set of fixed points projected with an additional, fixed laser system. The fixed points are easy to discern from the crosshair and act as fiducials. The system is not well suited for large scenes, since a large number of fiducials would be needed. It acquires depth only over a very narrow field of view at each frame, which implies long acquisition times in the case of complex scenes. It does not acquire color.

Rusinkiewicz et al. [10] present a structured light system where the object being scanned is hand-held in the fields of view of a fixed projector and fixed camera. The modeling pipeline is very fast. The frames are registered in real time using an iterative closest point (ICP) algorithm. The evolving model is constructed in real time and is rendered to provide immediate feedback to the operator. The system is limited to the outside-looking-in modeling case, and does not acquire color. A similar system is proposed by Koninckx [13] where moving or deformable objects are captured in real time. The system acquires depth using a pattern of equidistant black and white stripes and a few transversal color stripes for decoding. The disadvantages of their system are limited acquisition range due to the fixed camera and projector configuration and the requirement of strictly controlling the lighting in the scene. Despite their shortcomings, both systems demonstrate the advantages of interactive modeling.

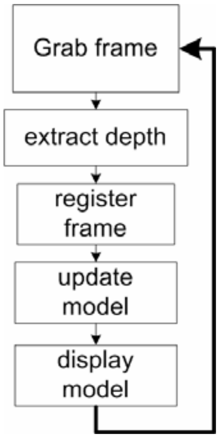

Figure 1 Interactive modeling pipeline.

3. CHALLENGES OF INTERACTIVE MODELING

Figure 1 shows an interactive modeling pipeline. Frame acquisition and model display are handled by consumer level technology. High-quality color can be obtained with a progressive-scan digital video camera. The frames can be transferred using standard interfaces (FireWire, USB 2.0). The acquired images have ample color and spatial resolution. Texture-mapped polygonal models support photorealistic visualization of complex scenes and can be displayed at interactive rates using PC graphics accelerators. Achieving interactivity in the middle three stages of the pipeline poses major research challenges.

Depth extraction Adequate depth sampling is needed to build a high-quality model. No current technology acquires high-resolution, precise depth at interactive rates. Interactive modeling has the advantage of a large number of frames. We hypothesize that sparse per frame sampling is sufficient for high-quality modeling.

Real-time frame registration We need to compute the motion of the acquisition device from frame to frame in order to register the incoming data in the model coordinate system. Registration means establishing the 6 degrees of freedom of the device, which is a non-linear optimization problem. This difficulty is alleviated by interactive modeling because of coherence between frames. However, the large number of frames raises the issue of registration drift, which must be quantified and controlled.

Real-time incremental modeling The challenge is to merge the geometry and color of the registered frames into an evolving scene model in real time. The redundant data has to be eliminated and the rest of the data has to be appended to the scene model. The definition of redundancy depends upon the desired level of detail. The reflective properties of the scene surfaces also come into play, for example a shiny surface has a different appearance when viewed from different angles. The model must support realistic interactive rendering. It should also support post processing, such as smoothing and compression.

4. MODELCAMERA INTERACTIVE MODELING SYSTEM

We have developed an interactive modeling system, called the *ModelCamera*, that addresses these challenges. The device consists of a hand-held digital video camera enhanced with a laser system that projects laser beams in the field of view of the camera (Figure 2). The laser beams produce blobs in the video frames that show where they hit scene surfaces. As the operator scans the scene, the video frames are read into a computer. The blobs are located in the frames and their 3D positions are inferred. Each incoming frame is registered using its color and depth data and then integrated into the scene model. The evolving model is rendered continually to provide immediate feedback to the operator. The ModelCamera processes five frames per second.

Depth extraction The ModelCamera acquires one depth sample per blob per frame by finding the blobs and triangulating their 3D positions in camera coordinates. We detect blobs quickly and robustly by exploiting epipolar geometry. Each blob is confined to the intersection line of the image plane with the plane through its laser ray and the center of projection of the camera (Figure 3). The blob detector searches the epipolar line for an intensity peak. We exploit coherent camera motion by starting the search at the peak from the previous frame. The peak is normally detected near the peak in the previous frame with minimal search. If this heuristic fails, the entire epipolar line is searched. Figure 4 shows a typical frame with its blobs and epipolar lines.

Real-time frame registration We have developed an algorithm for registering a frame with the previous frame using dense color and sparse depth. These results are composed to register a sequence of frames. Consider the frame (f_i) shown in Figure 4, where 16 blobs are on the armrest of a sofa, and consider a frame f_{i+1} obtained 200 ms later by translating the ModelCamera to the left. A surface fitted through the depth samples of f_i accurately approximates the geometry. Because of this, the surface on which the blobs of frame f_{i+1} move is known. The depth samples of frames f_i and f_{i+1} establish that the camera did not translate perpendicularly to the armrest.

Our depth-then-color algorithm (Figure 5) uses the depth samples to reduce the dimensionality of the registration search space from 6 to 3. This phase is very efficient because it involves a small number of depth samples. To continue the above example, the depth samples alone are not sufficient to

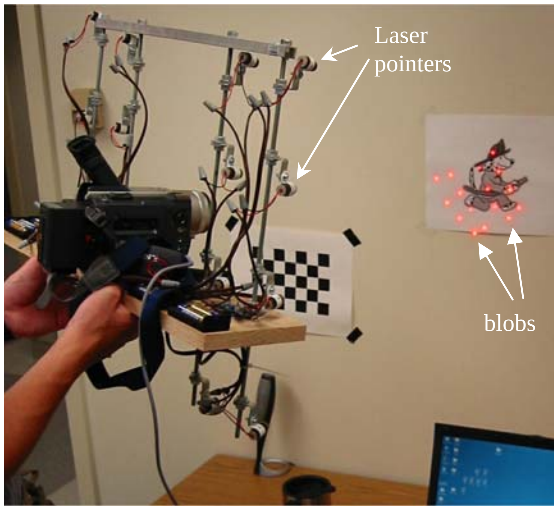

Figure 2 First ModelCamera prototype.

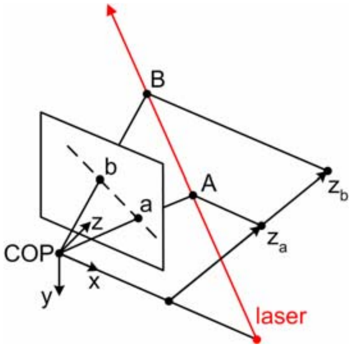

Figure 3 Epipolar line *ab* and triangulation.

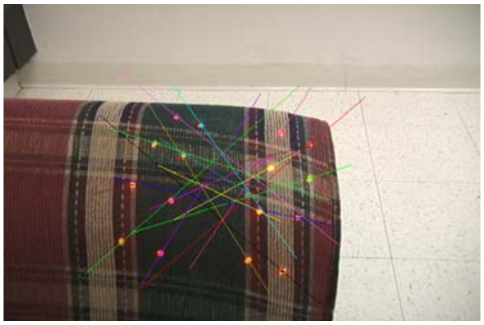

Figure 4 Blob detection on epipolar lines.

determine how much the ModelCamera translated to the left. Depth is invariant for that degree of freedom. The invariant depth degrees of freedom are established using the color between the blobs. The algorithm runs in real time because it matches the color data by searching in 3 dimensions. Thousands of color samples are used in this stage, so searching in 6 dimensions is impractical.

Figure 6 Depth image placement.

The partitioning of the camera's 6 degrees of freedom is done using the symmetries of the scene surfaces. The fundamental cases are plane, cylinder, cone, sphere, and helix. Even approximately symmetric surfaces provide too little depth variation for accurate determination of all degrees of freedom from depth. In Figure 4, the armrest is locally cylindrical, so the translation along and rotation about the armrest axis have to be found using the color data. A full description of the algorithm appears in a prior publication [12].

Figure 5 Consecutive frames before reg. (left), after depth reg. (middle), and after color reg. (right).

Real-time incremental modeling We model the scene as a collection of depth images that are created on demand (Figure 6). Depth images are regular color images enhanced with per-pixel depth [11], which can be merged and rendered efficiently. We merge the incoming frame by modeling the surface between the blobs with a depth image and inserting the non-redundant samples into the scene model [12]. Another advantage of depth images is convenient depth-and-color sampling-rate control.

5. SECOND MODELCAMERA PROTOTYPE

We have built a second ModelCamera prototype (Figure 8) that improves over the first prototype in three ways. The device is more compact for an easier hand-held operation. The laser modules are rigidly attached to the camera which prevents the laser beams to deviate from their original orientation. The diode modules are screwed in a plate manufactured using rapid prototyping. The plate is suspended to the right of the video camera using a rigid bracket. From our preliminary experimentation with the new prototype, the ModelCamera system maintains calibration over many days of use. The third improvement is that the epipolar lines do not intersect, which eliminates ambiguity in blob to laser assignment.

6. DISCUSSION

We have analyzed the challenges of interactive modeling and we have described two ModelCamera prototypes that address these challenges. The results are encouraging: we have scanned general curved surfaces at an average frame rate of 5 fps (Figure 8). Our system is the first real-time self-contained hand-held device. The blobs cast by the device itself are used to extract depth and to help register the camera frames.

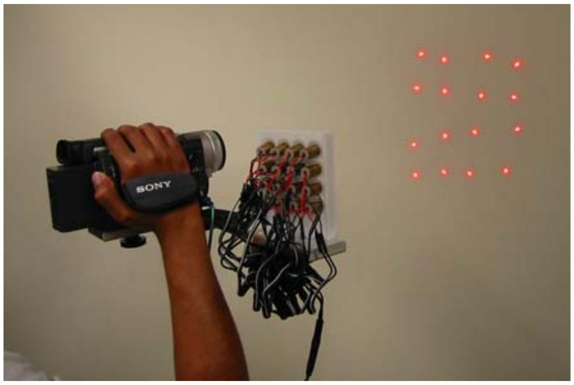

Figure 8 Second ModelCamera prototype.

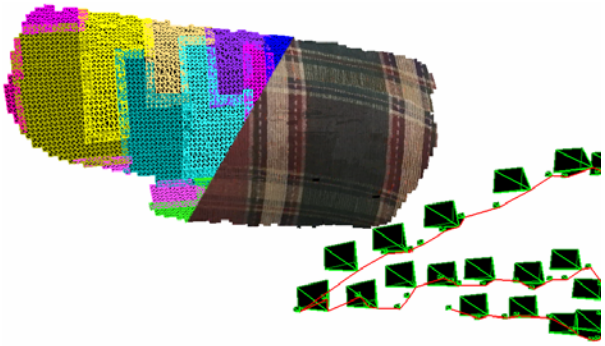

Figure 8 Curved surface model obtained from a sequence of frames; ModelCamera motion path also shown.

Considerable work remains to be done to realize the vision of a fast, accurate handheld device that the operator uses to scan scenes interactively. We will continue to reduce the size of the device. We plan to replace the multi-diode laser system with a single laser source and a diffraction grating head that splits the beam in a 7x7 dot matrix pattern. The new laser system is pen-sized and weighs 200 grams. The additional depth samples allow simultaneous scanning of two surfaces. The benefits are fewer symmetries, which accelerates registration, and the ability to scan from surface to surface.

The registration and modeling algorithms described so far assume that the blobs sample the surface densely enough in order to obtain a good approximation of the surface in between the blobs. In unstructured scenes, the blobs are distributed over many small surfaces. An example of such a scene is a messy bookshelf or a plant with dense foliage. We are investigating registration and modeling algorithms for these cases that restrict the camera motion to rotations around its center of projection. This restriction eliminates parallax and allows fast registration from color. Our registration method resembles stitching of color panoramas [14], but incorporates depth. Future extensions of the ModelCamera paradigm include outdoor modeling and dynamic scenes.

ACKNOWLEDGEMENTS

We would like to thank Gleb Bahmutov and Chun Jia for their help with the implementation of our system, and Chris Hoffmann, Pat Hanrahan and Leonard McMillan for useful discussions. This work was supported by NSF grants IIS-0082339 and CCR-0306214, by a Purdue OTC Trask grant, by the Purdue Center for Graphics, Geometry and Visualization, and by Intel and IBM Corporations.

REFERENCES

[1] M. Levoy, and P. Hanrahan. Light Field Rendering. Proc. of SIGGRAPH ’96, 31-42 (1996).

[2] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The Lumigraph. Proc. of SIGGRAPH ’96, 43-54 (1996).

[3] J. Maver and R. Bajcsy. Occlusions as a guide for planning the next view, IEEE Transactions on Pattern Analysis and Machine Intelligence 15(5), pp. 417-433, 1993.

[4] P. Allen, M. Reed, and I. Stamos, View Planning for Site Modeling Proc. DARPA Image Understanding Workshop, Monterey, November 21-23, 1998.

[5] W. Scott et al., View Planning with a Registration Constraint, In IEEE Third International Conference on 3D Digital Imaging and Modeling, Quebec City, Canada, May 28 – June 1, 2001.

[6] J.-Y. Bouguet and P. Perona, 3D Photography using Shadows in Dual-Space Geometry, International Journal of Computer Vision, Vol. 35, No. 2, Nov. 1999, pp. 129-149.

[7] M. Takatsuka et al., Low-cost Interactive Active Monocular Range Finder, in Proc. of the IEEE Conf. on Compter Vision and Pattern Rec-ognition, Fort Collins, CO, USA, Jun. 1999, pp. 444-449.

[8] N. A. Borghese et al., Autoscan: A Flexible and Portable 3D Scanner, IEEE Computer Graphics and Applications, Vol.18, No.3, May/ Jun. 1998, pp. 38-41.

[9] P. Hebert, A self-referenced hand-held range sensor, Proceedings of Third International Conference on 3-D Digital Imaging and Modeling, 5- 12, 2001.

[10] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Real-Time 3D Model Acquisition. In proc of SIGGRAPH 2002.

[11] L. McMillan L. and G. Bishop. Plenoptic Modeling: An Image-Based Rendering System. Proc. SIGGRAPH ’95, 39-46 (1995).

[12] V. Popescu, E. Sacks and G. Bahmutov. The ModelCamera: A Hand-Held Device for Interactive Modeling, Proceedings of Fourth International Conference on 3D Digital Imaging and Modeling 2003, 285-292, (2003).

[13] T. P. Koninckx, A. Griesser, and L. Van Gool, Real-Time Range Scanning of Deformable Surfaces by Adaptively Coded Structured Light. Proceedings of Fourth International Conference on 3D Digital Imaging and Modeling 2003, 293-301, (2003).

[14] S. Chen, Quicktime VR - An Image-Based Approach to Virtual Environment Navigation, Proc. SIGGRAPH ’95, 29-38 (1995).