IMAGE-BASED MODELING AND RENDERING OF ARCHITECTURE WITH INTERACTIVE PHOTOGRAMMETRY AND VIEW-DEPENDENT TEXTURE MAPPING

Paul E. Debevec, Camillo J. Taylor, Jitendra Malik, Golan Levin, George Borshukov, Yizhou Yu

Computer Science Division
University of California at Berkeley
Berkeley, CA 94720-1776
debevec@cs.berkeley.edu

ABSTRACT

In this paper we overview an approach to modeling and rendering architectural scenes from a sparse set of photographs. The approach is designed to require no special hardware and to produce real-time renderings on standard graphics hardware. The modeling method is an interactive photogrammetric modeling tool for recovering geometric models of architectural scenes from photographs. The tool is practical as an interactive method because it solves directly for architectural dimensions rather than for a multitude of vertex coordinates or depth measurements. The technique also determines the camera positions of the photographs, allowing the photographs to be taken from arbitrary unrecorded locations. The rendering method creates novel views of the scene based on the recovered model and the original photographs. A view-dependent texture mapping method allows all available image information to be used to produce the most photorealistic renderings. We focus on a number of results and applications of the method.

1. INTRODUCTION

Architecture, a nearly omnipresent visual element, is important to model and render photorealistically. Current techniques of converting architectural plans or survey data to CAD models are very labor intensive, and methods for rendering such models are generally not photorealistic. In this work we have developed methods for recovering the geometry and appearance of architectural scenes directly from a sparse set of photographs. By making use of user interaction and leveraging geometric properties common to architectural scenes, the modeling technique can recover accurate models from a relatively sparse set of photographs. Our image-based rendering technique, view-dependent texture mapping, makes use of all the original photographic data to produce the best rendering possible for any given viewing position. As a result, photorealistic renderings can be produced in real time on standard graphics hardware.

2. PHOTOGRAMMETRIC MODELING

Our photogrammetric modeling technique is a an interactive tool that allows the user to build a geometric model of an architectural scene based on a set of photographs. The technique is able to be convenient, efficient, and robust by exploiting the geometric structure common in architecture.

The prototype implementation of this modeling technique is called *Façade*. In Façade, the user builds a 3D model of the scene in the traditional way by assembling a collection of geometric primitives such as boxes, frustums, arches, and surfaces of revolution. However, the user does not need to specify the dimensions or the locations of these components. Instead, the user corresponds points, edges, and contours in the model with edges marked in the photographs, and the system computes the architectural dimensions and the camera positions so that the model agrees with the photographed geometry. (See Fig. 1)

The technique is powerful because it solves directly for the architectural dimensions of the scene: the lengths of walls, the widths of doors, the heights of roofs, rather than the multitude of vertex coordinates that a standard photogrammetric approach would try to recover. As a result, the reconstruction problem becomes simpler by orders of magnitude, both in computational complexity and, more importantly, in the number of image features that it is necessary for the user to mark. The technique also allows the user to fully exploit architectural symmetries by modeling repeated structures and computing redundant dimensions only once. This further simplifies the modeling process.

Figures 1-3 show the system being used to recover architecture composed of polyhedra, arches, and surfaces of revolution. More information about the various reconstruction algorithms can be found in [1, 2].

3. RENDERING WITH VIEW-DEPENDENT TEXTURE MAPPING

For rendering, we begin with the standard approach of forming texture maps by projecting the original photos onto the recovered geometry, taking into account visibility. However, we take the additional step of selectively blending between the original photographs depending on the user's viewpoint. For example, suppose that a particular surface of a building is seen in three of the photographs from the left, front, and right. If we are generating a novel view from the left, we will use the surface's appearance in the left view as the texture map. Similarly, for a view in front of the surface we will use the frontal view. For a view in between the left and front, we blend between the left and front texture maps accordingly; for a view between the front and right, we would blend between the front and right images. As a result, the view-dependent

This research was supported by a National Science Foundation Graduate Research Fellowship and support from Interval Research Corporation, the California MICRO program, Silicon Graphics, JSEP contract F49620-93-C-0014 and ONR/BMDO grant FDN00014-96-1-1200. Golan Levin is a member of the research staff at Interval Research Corporation. See also http://www.cs.berkeley.edu/~debevec/Research

0-7803-4455-3/98/$10.00 © 1998 IEEE

texture mapping approach allows the renderings to be considerably more realistic than static texture-mapping, since it better captures non-diffuse reflectance and can simulate the appearance of unmodeled geometry. Using OpenGL on mid-range Silicon Graphics workstations, we are able to render complex scenes in such a manner in real time. As PC-based graphics cards continue to increase in performance, we can expect realtime rendering on inexpensive computers in the near future. The rendering of the tower in Fig. 6 was created in real time using this technique.

4. MODEL-BASED STEREO

Further research enables the computer to automatically refine a basic recovered model to conform to more complicated architectural geometry. The technique, which we call model-based stereo, displaces the surfaces of the model to make them maximally consistent with their appearance across multiple photographs. Thus, a user can model a bumpy wall as a flat surface, and the computer will compute the relief. This technique was employed extensively in modeling the façade of a gothic cathedral for the interactive art installation "Rouen Revisited", described in the next section and in Fig. 4.

5. APPLICATIONS

This work has been used in a number of creative applications, including "Rouen Revisited", an interactive art installation by Golan Levin and Paul Debevec, and "The Campanile Movie", a short computer-animated film directed by Paul Debevec.

5.1. Rouen Revisited

Between 1892 and 1894, the French Impressionist Claude Monet produced nearly 30 oil paintings of the main façade of the Rouen Cathedral in Normandy. Interested in the play of light and atmosphere, Monet systematically painted the cathedral at different times of day, from different angles, and in varied weather conditions. The *Rouen Revisited* interactive kiosk allows users to explore the façade of the Rouen Cathedral, as Monet might have painted it, from any angle, time of day, and degree of atmospheric haze (see Figs. 4 and 5). Users can contrast these re-rendered paintings with similar views synthesized from century-old archival photographs, as well as from recent photographs that reveal the scars of a century of weathering and war.

To produce renderings of the cathedral's façade from arbitrary angles, a 3D model of its West façade was constructed from three time-series photographs taken in front of the cathedral in January 1996. During the same expedition reproductions of Monet's paintings as well as antique photographs of the cathedral were obtained.

With the 3D model constructed, the photographs and Monet paintings were registered with and projected onto the 3D model. Re-renderings of each of the projected paintings and photographs were then generated from hundreds of points of view; renderings of the cathedral in different atmospheric conditions and at arbitrary times of day were derived from our own time-lapse photographs of the cathedral and by interpolating between the textures of Monet's original paintings. The exhibit is currently on public display in the gallery of Interval Research Corporation in Palo Alto, CA[1].

[1] See http://www.interval.com/projects/rouen/

5.2. The Campanile Movie

The largest-scale project we have used the system for was for a short computer-animated film featuring the UC Berkeley campus. The film involved both live action cinematography as well as computer-generated footage of a flight around the campus bell tower. To produce such renderings, we used Façade to construct a detailed model of the bell tower as well as forty buildings of the surrounding campus. The twenty photographs used in the project were either taken from the ground (3), from the tip of the tower (10), from a kite looking down on the tower (3), or from a recent campus aerial mapping survey (4). The result was an image-based model of the tower and the surrounding environment which could be viewed in any direction from anywhere in the vicinity of the top of the tower, shown in Fig. 6. A hardware-assisted projective texture mapping technique made it possible to render the model at 60 frames per second on an SGI Onyx2 InfiniteReality, and at 20 frames per second with full view-dependent texture mapping[2].

[2] See http://www.cs.berkeley.edu/~debevec/Campanile

6. CONCLUSION AND FUTURE WORK

Besides increasing the degree of automation in the modeling process, it would also be desirable to render scenes in different lighting configurations than those observed in the original photographs. This will involve recovering the actual reflectance characteristics (i.e. bi-directional reflectance distribution functions, or BRDF's) of the surfaces in the scene. Such information will also make it possible to convincingly add objects to and otherwise manipulate the scene. Furthermore, as 3D scanning methods continue to advance, hybrid photogrammetric / active sensing approaches will be able to extract more detailed geometry. However, for scenes that exist only in photographs, pure photogrammetric approaches must suffice. Regardless of how the geometry is constructed, image-based techniques need to extrapolate the scene's appearance from the original photographs to form renderings from novel points of view. The view-dependent texture mapping method effectively performs linear interpolation between the available views. Better interpolation methods that use reflectance models to extrapolate the observed pixel values may produce better-conditioned results; however, it would be less straightforward to create renderings with existing graphics hardware. Lastly, digital images generally have limited dynamic range which makes it difficult to model the appearance of sunlit or interior scenes accurately. Combining the photogrammetric modeling technique with techniques for high dynamic range photography [3] may be able to solve this problem.

7. REFERENCES

[1] DEBEVEC, P. E., TAYLOR, C. J., AND MALIK, J. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *SIGGRAPH '96* (August 1996), pp. 11–20.

[2] BORSHUKOV, G. New algorithms for modeling and rendering architecture from photographs. Master's thesis, University of California at Berkeley, Computer Science Division, Berkeley CA, May 1997.

[3] DEBEVEC, P. E., AND MALIK, J. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH '97* (August 1997), pp. 369–378.

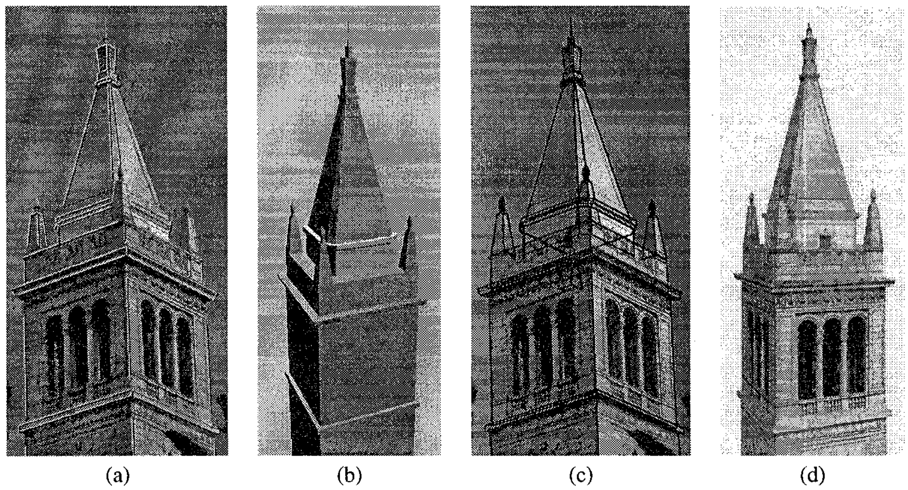


Figure 1: Reconstructing a simple model in Façade. **(a)** A photograph of the Campanile, Berkeley's bell tower, with marked edges shown in green. **(b)** The recovered model. Although only the left pinnacle was marked, the remaining three (including one not visible) were recovered from symmetrical constraints in the model. Our method allows any number of images to be used, but in this case constraints of symmetry made it possible to recover an accurate 3D model from a single photograph. **(c)** The accuracy of the model is verified by reprojecting it into the original photograph through the recovered camera position. **(d)** A synthetic view of the Campanile generated using three photographs and the view-dependent texture-mapping method.

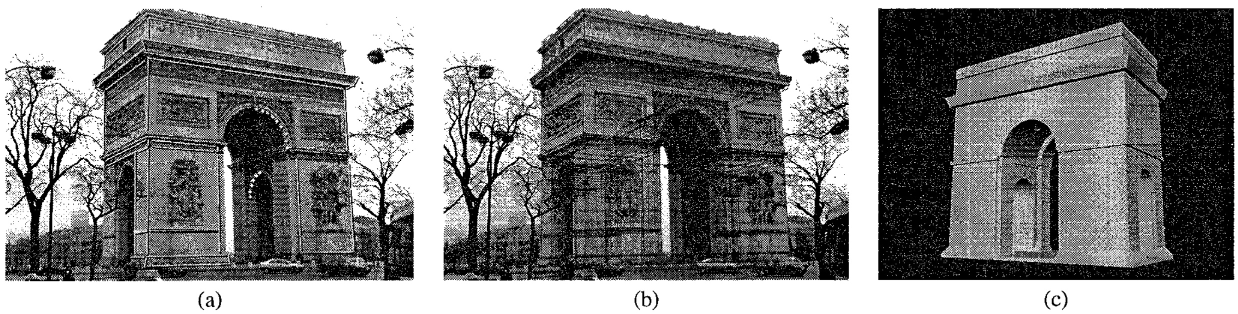


Figure 2: Reconstructing arches. **(a)** One of five photographs of the Arc de Triomphe, with edges used in the reconstruction shown. **(b)** The edges of the recovered model overlaid on the photograph. **(c)** A flat-shaded view of the model, featuring three round arches.

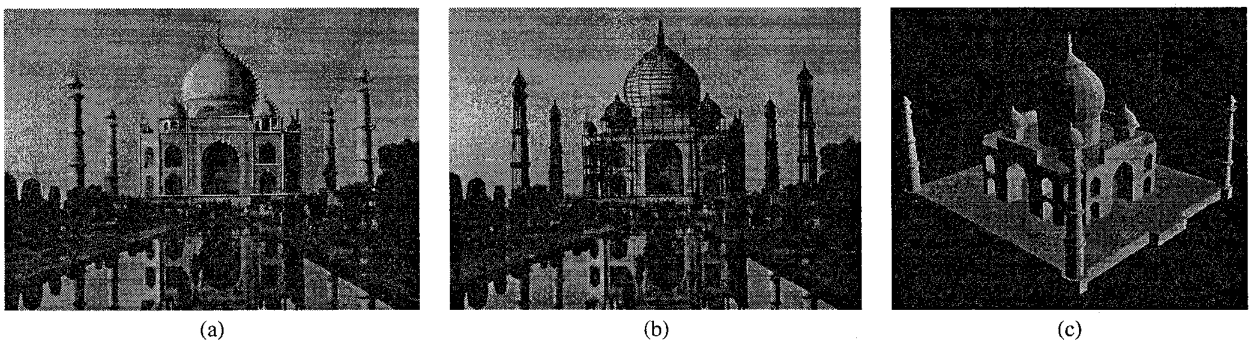
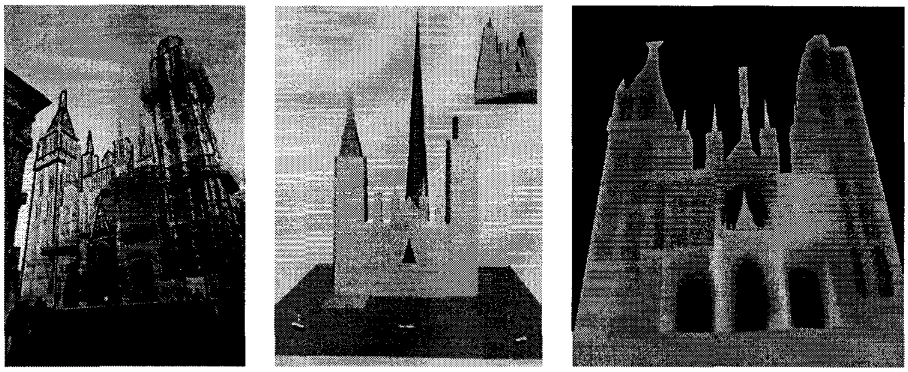
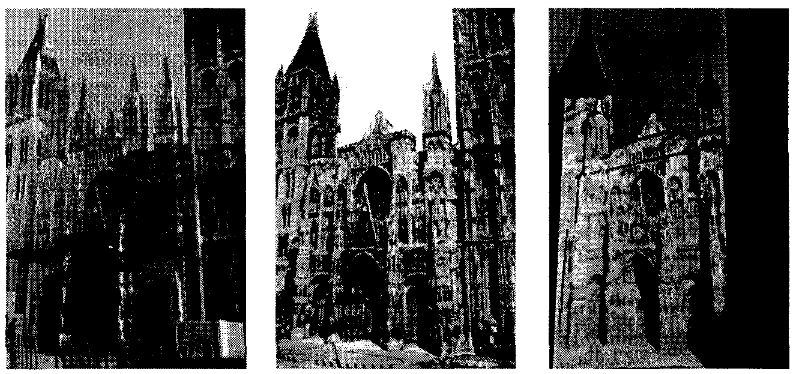


Figure 3: Reconstructing surfaces of revolution. **(a)** An image of the Taj Mahal, a 400×300 pixel JPEG image found on the internet. The twenty edges and five contours marked in this single photograph shown. **(b)** The edges of the recovered model, overlaid on the original photograph. **(c)** The recovered model, flat-shaded, with pointed Islamic arches. The system made it possible to leverage the symmetries in the Taj Mahal and to construct this complex model in about an hour from the single photograph.



(a) (b) (c)

Figure 4: Reconstruction of the West façade of the Rouen Cathedral for the *Rouen Revisited* installation in the SIGGRAPH'96 Art Show. **(a)** One of three original photographs, with marked edges indicated. **(b)** Basic model built from the three photographs. The three recovered camera positions are indicated on the ground in front of the cathedral. **(c)** Geometric detail recovered from the model-based stereo algorithm shown as a displacement map keyed to the middle photograph. Dark areas lie behind the modeled geometry; light areas are in front of it.



(a) Synthetic view, 1996 (b) Synthetic view, 1896 (c) Synthetic view, painting

Figure 5: Synthetic views of the modeled Rouen cathedral as it appeared in 1896, 1996, and with one of Monet's paintings reprojected onto the façade. All three pictures are rendered from the same viewpoint. See also http://www.interval.com/projects/rouen/

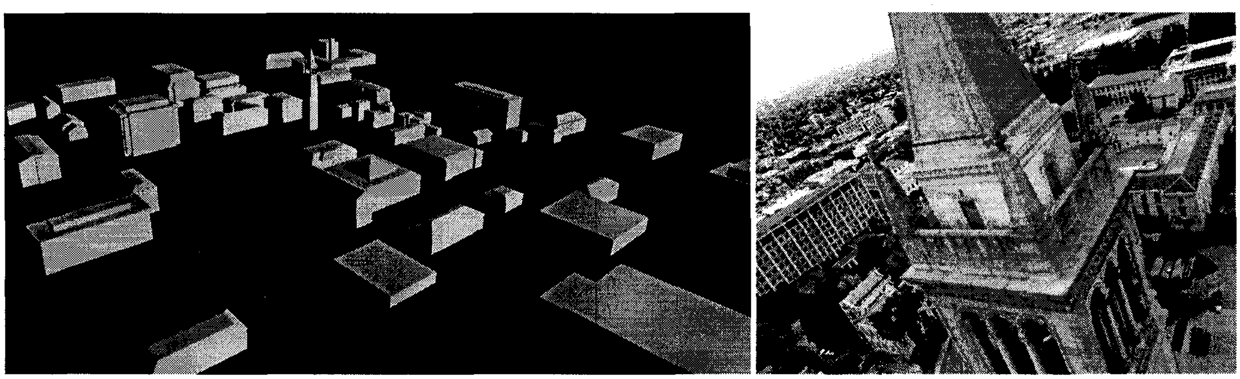


Figure 6: The Berkeley Campus model, constructed from 20 photographs. Constructed specifically for the Campanile movie, it is more detailed near the tower and fades out in complexity further away. Forty of the campus's buildings are represented. The film model also includes photogrammetrically recovered terrain geometry that extends out to the horizon. On the right is a rendering from the computer animation, showing the tower up close with several buildings and terrain in the background. See also http://www.cs.berkeley.edu/~debevec/Campanile