Image-Based Techniques for Digitizing Environments and Artifacts

Paul Debevec

University of Southern California Institute for Creative Technologies
Graphics Laboratory

13274 Fiji Way, 5th floor, Marina del Rey, California, USA, 90292
debevec@ict.usc.edu

Abstract

This paper presents an overview of techniques for generating photoreal computer graphics models of real-world places and objects. Our group's early efforts in modeling scenes involved the development of Façade, an interactive photogrammetric modeling system that uses geometric primitives to model the scene, and projective texture mapping to produce the scene appearance properties. Subsequent work has produced techniques to model the incident illumination within scenes, which we have shown to be useful for realistically adding computer-generated objects to image-based models. More recently, our work has focussed on recovering lighting-independent models of scenes and objects, capturing how each point on an object reflects light. Our latest work combines three-dimensional range scans, digital photographs, and incident illumination measurements to produce lighting-independent models of complex objects and environments.

1. Introduction

A basic goal of three-dimensional modeling and imaging has been to digitize photorealistic models of real-world places and objects. The potential benefits of these techniques to cultural heritage are significant: objects and places can be archived in digital forms less susceptible to decay and plunder; scholars can test and communicate theories of reconstruction without disturbing original artifacts; and the world's artistic and architectural heritage can be vividly and widely shared among all those with access to a computer. It will likely be a long time before an exact computer representation of an artifact – a record, one could suppose, of the object's atomic structure – will be able to be nondestructively recorded, stored, and rendered. Instead, most digitizing techniques have focussed on recording an artifact's surface geometry and its photometric properties - the information necessary to render it from different viewing angles and seen under various forms of illumination. Despite much insightful work in this area, no general fully technique exists that for faithfully digitizing objects having arbitrarily complex geometry and spatially-varying reflectance properties.

The computer graphics research my group has undertaken in recent years has involved a variety of object and scene modeling methods, ranging from techniques based principally on recovering geometry to those based more significantly on acquiring images. In this paper I will discuss several of these projects with a particular regard to the nature of the models acquired, why the models were appropriate for the particular applications being explored, and which aspects of the artifacts have been left unrecorded. While none of these projects on its own presents a general solution to the scene and object digitizing problem, evaluating their strengths and weaknesses together suggests new avenues for combining geometric and photometric measurements of objects for digitizing a more general set of complex objects.

2. Immersion '94: Modeling an Environment from Stereo

In 1994 I worked in a group led by media artist Michael Naimark working to analyze a sequence of forward-looking stereo images he had shot at one-meter intervals along several miles of the trails in Banff National Forest. Our goal for the summer was to turn these sets of stereo image pairs into a photorealistic virtual environment. The technique we used was to determine stereo correspondences, and thus depth, between left-right pairs of images, and then to project the corresponded pixels forward into the 3D world. For this we used a stereo algorithm developed by John Woodill and Ramin Zabih [18]. To create virtual renderings, we projected a supersampled version of the points onto a virtual

Proceedings of the Fourth International Conference on 3-D Digital Imaging and Modeling (3DIM'03)
0-7695-1991-1/03 $ 17.00 © 2003 IEEE

image plane displaced from the original point of view, using a Z-buffer to resolve the occlusions. With just one stereo pair, we could realistically re-render the scene from anywhere up to a meter away from the original camera positions, except for artifacts resulting from areas that were unseen in the original images, such as the areas originally hidden behind tree trunks. To fill in these occluded areas for novel views, our system would pick the two closest stereo pairs to the desired virtual point of view, and render both to the desired novel point of view. These images were then optically composited so that wherever one lacked information, the other would fill it in. In areas where both images had information, the data was linearly blended according to which original view the novel view was closer to. The result was the ability to realistically move through the forest, as long as one kept the virtual viewpoint within about a meter of the original path. Naimark presented this work at the SIGGRAPH 95 panel "Museums without Walls: New Media for New Museums" [1], and the animations may be seen at the Immersion project website [15].

The results that were achieved were exciting at the time; a scene that at the beginning seemed impossible to model had been "digitized" in the sense that nearly photoreal rendered images could be created of flying down the forest trail. Nonetheless, the model of the scene obtained by the technique did not allow the virtual camera to move more than a meter or so from the original path, and the scene could only be seen in the original lighting in which it was photographed. The next project I will discuss aimed to free the camera to fly around the scene.

3 Photogrammetric Modeling and Rendering with Façade

My thesis work [6] at Berkeley done in collaboration with C.J. Taylor presented a system for modeling and rendering architectural scenes from photographs. Architectural scenes are an interesting case of the general modeling problem since their geometry is typically very structured, and they are also one of the most common types of environment to model. The goal of the research was to model architecture in a way that is convenient, requires relatively few photographs, and produces freely navigable and photorealistic models.

The product of this research was Façade [8], an interactive computer program that enables a user to build photorealistic architectural models from a small set of photographs. In Façade, the user builds a 3D model of the scene by specifying a collection of geometric primitives such as boxes, arches, and surfaces of revolution. However, unlike in a traditional modeling program, the user does not need to specify the dimensions or the locations of these pieces. Instead, the user corresponds edges in the model to edges marked

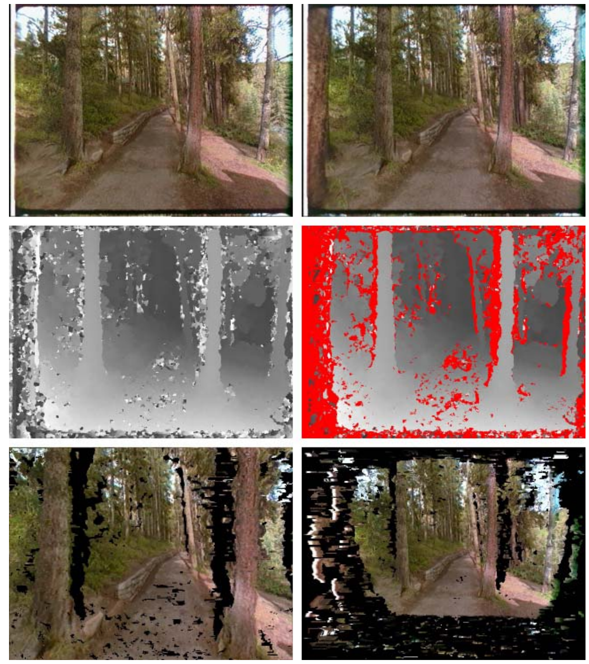


Figure 1. The Immersion '94 image-based modeling and rendering project. The top images are a stereo pair (reversed for cross-eyed stereo viewing) taken in Banff National Forest. The middle left photo is a stereo disparity map produced by John Woodfill's parallel implementation of the Zabih-Woodfill stereo algorithm [18]. To its right the map has been processed using a left-right consistency check to invalidate regions where running stereo based on the left image and stereo based on the right image did not produce consistent results. Below are two virtual views generated by casting each pixel out into space based on its computed depth estimate, and reprojecting the pixels into novel camera positions. On the left is the result of virtually moving one meter forward, on the right is the result of virtually moving one meter backward. Note the dark disoccluded areas produced by these virtual camera moves; these areas were not seen in the original stereo pair. In the Immersion '94 animations (available at http://www.debevec.org/Immersion , these regions were automatically filled in from neighboring stereo pairs.

in the photographs, and the computer uses an optimization process to determine the shapes and positions of the primitives that make the model agree with the photographed geometry (Fig. 2).

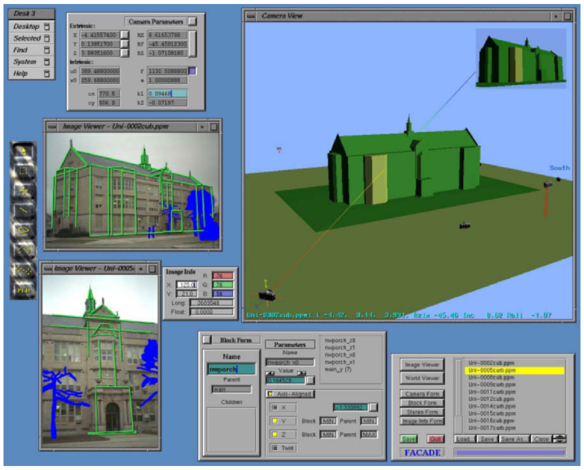


Figure 2. A screen snapshot from Façade. The windows include the image viewers at the left, where the user marks architectural edge features, and model viewers, where the user instantiates geometric primitives (blocks) and corresponds model edges to image features. Façade's reconstruction feature then determines the camera parameters and position and dimensions of all the blocks that make the model conform to the photographs. The other windows include the toolbar, the camera parameter dialog, the block parameter/constraint dialog, and the main image list window. See also http://www.debevec.org/Thesis/.

Façade solves directly for the architectural dimensions of the scene: the lengths of walls, the widths of doors, and 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 in the number of image features that it is necessary for the user to mark. The technique also allows the user to exploit architectural symmetries – modeling repeated structures and computing redundant dimensions only once.

Like any structure-from-multiple-views algorithm, Façade solves for where the original cameras were in the scene. With the camera positions known, any one of the photographs can be projected back onto the reconstructed geometry using projective texture mapping, which allows Façade to generate photorealistic views of the scene by using all of the available photographs.

Some additional research done in the context of the Façade system enables the computer to automatically refine a basic recovered model to conform to more complicated architectural geometry. The technique, called 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 in modeling the West façade of the gothic Rouen cathedral for the interactive art installation *Rouen Revisited* shown at the SIGGRAPH 96 art show. Most of the area between the two main towers seen in Fig. 3 was originally modeled as a single polygon. The Rouen project also motivated the addition of new features to Façade to solve for unknown focal lengths and centers of projection in order to make use of historic photographs of the cathedral.

Rendering: 1996 Rendering: 1896 Rendering: painting

Figure 3. Rouen Revisited. Synthetic views of the Rouen cathedral from the *Rouen Revisited* art installation. Left: a synthetic view created from photographs taken in January, 1996. Middle: a synthetic view created from historic postcards showing the cathedral at the time Monet executed his series of paintings (1892-1894). Right: a synthetic view of one of Monet's twenty-eight paintings of the cathedral projected onto its historic geometry, rendering it from a novel viewpoint.

4 The Campanile Movie: Rendering in Real Time

The Campanile Movie project aimed to use Façade to reconstruct a model of the UC Berkeley bell tower and its entire surrounding campus - to increase the photorealism of the renderings by showing both a building and its context. For data capture, I took photographs from the ground, from the tower, and (thanks to Berkeley professor of architecture Cris Benton) from above the tower using a kite. The final model we built in Façade contained forty of the campus buildings; the buildings further away appeared only as textures projected onto the ground. There were a few thousand polygons in the model, and the sixteen images (Fig. 4) used in rendering the scene fit precisely into the available texture memory of the Silicon Graphics RealityEngine we used for rendering. Using OpenGL and a hardware-accelerated view-dependent texture-mapping technique – selectively blending between the original photographs depending on the user's viewpoint [9] – made it possible to render the scene convincingly in real time.

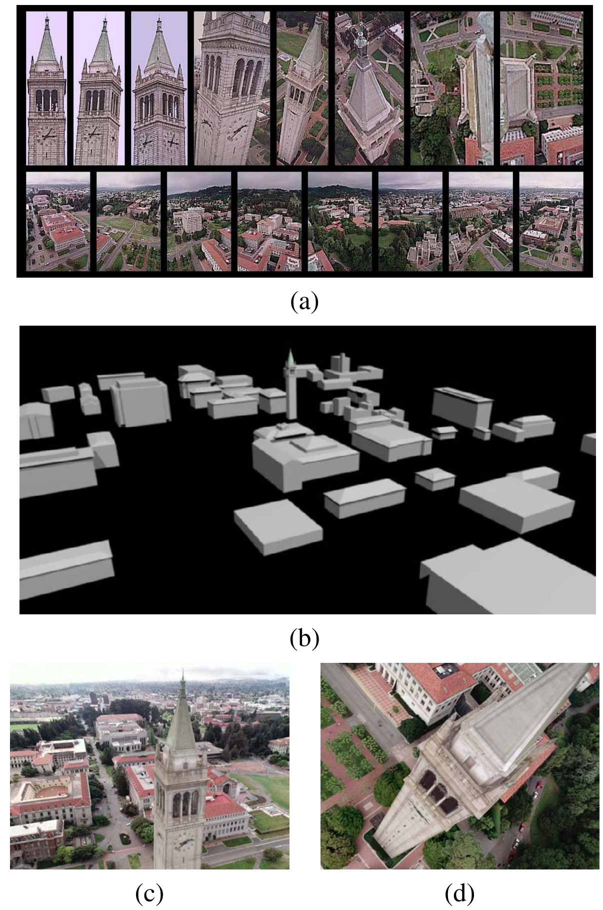


Figure 4. The Campanile Movie. At top are the original sixteen photographs used for rendering; four additional aerial photographs were used in modeling the campus geometry. In the middle is a rendering of the campus buildings reconstructed from the photographs using Façade; the final model also included photogrammetrically recovered terrain extending out to the horizon. At bottom are two computer renderings of the Berkeley campus model obtained through view-dependent texture mapping from the SIGGRAPH 97 animation. The film can be seen at http://www.debevec.org/Campanile/.

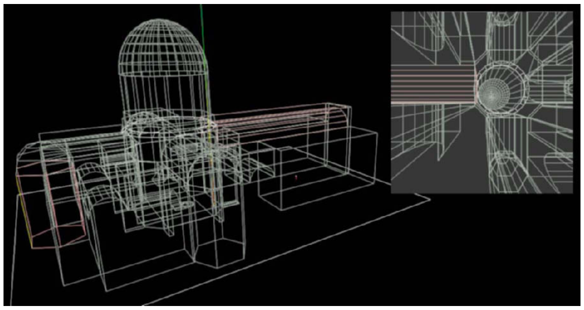
5 Fiat Lux: Capturing Lighting and Adding Objects

We most recently used Façade to model and render the interior of St. Peter's Basilica for the animation *Fiat Lux* (Fig. 5), which was shown at the SIGGRAPH 99 Electronic Theater. In *Fiat Lux*, our goal was to not only create virtual renderings of moving through St. Peter's, but to augment the space with animated computer-generated objects in the service of an abstract interpretation of the conflict between Galileo and the church.

The key to making the computer-generated objects appear to be truly present in the scene was to illuminate the CG objects with the actual illumination from the Basilica. To record the illumination we used a high dynamic photography method [7] we had developed in which a series of pictures taken with differing exposures are combined into a *radiance image* – without the technique, cameras do not have nearly the range of brightness values to accurately record the full range of illumination in the real world. We then used an image-based lighting [4] technique to illuminate the CG objects with the images of real light using a global illumination rendering system. In addition, we used an inverse global illumination [17] technique to derive lighting-independent reflectance properties of the floor of St. Peter's, allowing the objects to cast shadows on and appear in reflections in the floor. Having the full range of illumination was additionally useful in producing a variety of realistic effects of cinematography, such as soft focus, glare, vignetting, and lens flare.

6 The Light Stage: A Photometric Approach to Digitizing Cultural Artifacts

The best techniques presented to date for modeling real-world artifacts involve two stages. First, the geometry of the artifact is measured using a range scanning device and some sort of mesh merging algorithm [13, 16, 12]. Second,



(a)

(b)

(c)

Figure 5. ***Fiat Lux*****. The animation** *Fiat Lux* **shown at the SIGGRAPH 99 Electronic Theater used Façade [8] to model and render the interior of St. Peter's Basilica from single panorama assembled from a set of ten images. Each image was acquired using high dynamic range photography [7], in which each image is taken with a range of different exposure settings and then assembled into a single image that represents the full range of illumination in the scene. The recovered lighting was used to illuminate the synthetic objects added to the scene, giving them the correct shading, shadows, reflections, and highlights. The film is can be seen at http://www.debevec.org/FiatLux/.**

photographs taken under controlled lighting conditions are analyzed to determine the color of each point on the object's surface to produce texture maps for the object. A notable exception to this process is the trichromatic laser scanner device presented in [3, 2] which is able to measure object surface color in the same process as obtaining the geometric laser scan. All of these techniques have produced 3D models with accurate models of surface geometry as well as lighting-independent diffuse texture maps.

These techniques represent significant advances in the field of artifact digitization, and have shown themselves to work particularly well on objects with close to Lambertian reflectance properties such as aged marble sculptures. However, these existing techniques for model acquisition are difficult to apply to a large class of artifacts exhibiting spatially varying specular reflection properties, complex BRDFs, translucency, and subsurface scattering. As a result, artifacts featuring polished silver, gold, glass, fur, cloth, jewels, jade, leaves, or feathers remain very challenging to accurately digitize and to convincingly render. Since a large class of cultural artifacts are specifically designed to have intricate geometry and to reflect light in interesting ways, developing new techniques for such objects is an important avenue for research.

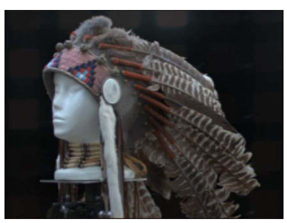
Our recent work involving the Light Stage [5, 11] (Figure 6) has explored a technique for creating re-lightable computer models of artifacts and people without explicitly modeling their geometric or reflectance properties. Instead, the artifact is illuminated from a dense array of incident illumination directions and a set of digital images are taken to record the artifact's *reflectance field*, a function that records how the object would appear seen from any angle as illuminated by any lighting environment. The data within a reflectance field can be quickly combined together in order to produce images of an artifact under any form of illumination, including lighting environments captured from the real world.

In [11], our group used an improved light stage device to digitize several Native American cultural artifacts exhibiting many of the geometric and reflectance properties which have traditionally been difficult to model and render. We furthermore described an interactive lighting tool that allows artifacts to be re-illuminated by a user in real time, and proposed image-based rendering techniques that allow an artifact to be manipulated in 3D as well as being arbitrarily illuminated. An advantage of this technique for capturing and rendering objects is that the object need not have well-defined surfaces or easy to model reflectance properties; the object can have arbitrary translucency, self-shadowing, interreflection, subsurface scattering, and fine surface detail. This is helpful for modeling and rendering subjects such as human faces, jewelry, and clothing which exhibit all of these properties.

(a)

(b)

(c)



(d)

Figure 6. Digitizing objects in the Light Stage. The Light Stage [5, 11] is designed to illuminate an artifact from every direction light can come from in a short period of time. In (a), a Native American headdress is being digitized. This allows a digital video camera to directly capture the subject's *reflectance field* – how it transforms incident illumination into radiant illumination – as an array of images (b). As a result, we can then synthetically illuminate the subject under any form of complex illumination directly from this captured data. The Native American headdress is shown as if illuminated by light from a cathedral (c) and a forest (d).

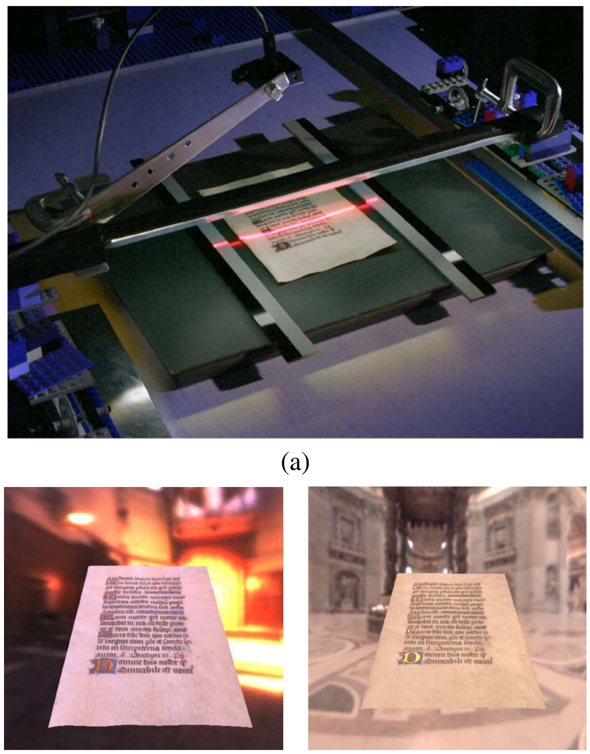
The significant drawback to this approach is the large amount of data that needs to be captured and stored for the object: on the order of at least a thousand lighting directions for each object viewing direction. As a result, we have yet to capture a model with both a dense array of viewing and lighting directions, so the changes in viewpoint we have been able to obtain thus far have been limited.

A recent project at MIT and MERL [14] used a device similar to our light stage to capture several objects from moderately dense arrays of viewing directions and approximately 64 lighting directions, and used a silhouette-based volumetric intersection technique to obtain rough geometry for the digitized objects. As we saw earlier with our Façade work, using basic scene geometry as a structure for projecting images allows sparsely sampled viewing directions to be effectively extrapolated to new viewing directions, at least for relatively diffuse objects where the light reflecting from a surface remains nearly constant under changing viewing directions. Something that still has not been leveraged in the digitizing process is that the way a particular point on an object surface reflects light (the point's four-dimensional BRDF) can often be well-approximated by a low-parameter reflectance model encompassing, for example, diffuse and specular parameters. The next project I will describe exploits this characteristic.

7. Linear Light Source Reflectometry

An important aspect of many cultural artifacts is that they have spatially-varying reflectance properties: they are made of different materials each of which reflect light in a different way. An example is an illuminated manuscript, in which different colors of inks and different types of precious metals are applied to the surface of parchment to produce textual designs and illustrations. We have found the case of digitizing an illuminated manuscript to be particularly good for developing algorithms to digitize spatially-varying reflectance properties since the geometry of a manuscript is relatively flat, making the geometry recovery and reflectometry processes simpler.

Digitizing the manuscript within the light stage would be possible, but since the manuscript contains both diffuse and sharply specular regions it would need to be illuminated by a large number of lighting and viewing directions to be faithfully captured. To avoid this problem, we developed a new digitizing technique described in [10] based on using a translating linear light source as the illuminant (see Fig. 7). We found that in passing the light source over the artifact and viewing it from an oblique angle, we could independently observe the diffuse and specular reflection of the light at each surface point. To turn these observations into digitized surface reflectance properties, we fit the observed diffuse and specular reflections to a standard physically-motivated reflectance model with diffuse, specular, and specular roughness coefficients. Furthermore, we found that by passing the light source across the object twice at different angles, we can estimate surface normals at every point on the object as well. Simple extensions to the technique – adding a laser stripe and a back light – make it possible to capture per-pixel surface height as well as per-pixel surface translucency in addition to the other reflectance parameters.



(a) (b) (c)

Figure 7. ***Linear Light Source Reflectometry*. The spatially-varying surface reflectance parameters of an illuminated manuscript are recorded using a simple apparatus based on a translating linear light source. The per-pixel estimates of diffuse and specular parameters, as well as surface normals, can be used to create photoreal real-time renderings (below) of the artifact from any viewing angle and in any lighting environment.**

8. Discussion

Looking at these projects together, we can see that each one of them can be characterized as acquiring or estimating a portion of the reflectance field of an environment or artifact: a dataset of how a environment or artifact would look from any angle under any form of incident illumination. The problems of digitizing environments and objects are related but have differences. Environments are larger, and often more geometrically complex, making them harder to photograph from a dense sampling of all possible angles. Environments are also harder to illuminate with controlled lighting - outdoor scenes are most easily illuminated by natural lighting rather than controlled sources, and indoor scenes tend to exhibit a great deal of mutual illumination between surfaces. As a result, many computer models which have been made of real-world environments are lighting-dependent, showing the scene under just one form of illumination, such as in the Banff forest, Berkeley campus, and St. Peter's Basilica models. Nonetheless, these models can be very effective digitizations, perhaps because for environments we are able to consider the illumination conditions to be part of the environment.

For digitizing artifacts, being able to capture their lighting-independent reflectance properties is much more important. Part of the reason is that when we view a virtual artifact from different angles, we most naturally assume that the lighting stays in our own frame of reference rather than rotating with the artifact. Also, in order to place a virtual artifact into a virtual museum environment, it is necessary to illuminate the artifact with the light present in the appropriate part of the museum. Fortunately, artifacts are in many respects easier to digitize than environments. They are smaller, and thus can be brought into object scanning facilities to be photographed from many angles and illuminated under controlled forms of illumination. As a result, most artifact modeling projects, such as the Native American Artifacts and the illuminated manuscript, have been successful at recovering lighting-independent models of the artifacts. However, the native artifacts could only be viewed from a limited set of viewing angles, and the manuscript, being a flat object, could be digitized without concern for self-occlusion and mutual illumination effects.

9. Conclusion

The projects surveyed in this paper each present a different approach to digitizing artifacts and environments. While each aims to produce photoreal results, the methods significantly differ in the nature of the measurements taken, the amount of user input required, the amount of geometric detail recovered, and the degree to which the models can be seen under novel lighting and from new viewpoints. Methods for environment and object digitizing in the future will likely draw upon several of these techniques to improve their results, though techniques for digitizing objects as compared to environments will likely continue to be significantly different.

Acknowledgements

The projects described in this paper are joint work with Michael Naimark, John Woodfill, Leo Villareal, Golan Levin, C.J. Taylor, Jitendra Malik, George Borshukov, Yizhou Yu, Tim Hawkins, Jonathan Cohen, Andrew Gardner, and Chris Tchou. Abjiheet Ghosh and Chris Tchou prepared the illuminated manuscript renderings in Fig. 7, and Maya Martinez arranged for the use of the cultural artifacts used in this work. This work has been funded by Interval Research Corporation, the National Science Foundation, the California MICRO program, Interactive Pictures Corporation, JSEP contract F49620-93-C-0014, U.S. Army contract DAAD19-99-D-0046, and the University of Southern California. The content of this information does not necessarily reflect the position or policy of the sponsors and no official endorsement should be inferred.

References

[1] A. C. Addison, D. MacLeod, G. Margolis, M. Naimark, and H.-P. Schwartz. Museums without walls: New media for new museums. In R. Cook, editor, *Computer Graphics annual Conference Series (SIGGRAPH 95)*, pages 480–481, August 1995.

[2] R. Baribeau, L. Cournoyer, G. Godin, and M. Rioux. Colour three-dimensional modelling of museum objects. *Imaging the Past, Electronic Imaging and Computer Graphics in Museum and Archaeology*, pages 199–209, 1996.

[3] R. Baribeau, M. Rioux, and G. Godin. Color reflectance modeling using a polychromatic laser sensor. *IEEE Trans. Pattern Anal. Machine Intell.*, 14(2):263–269, 1992.

[4] P. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH 98*, July 1998.

[5] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. *Proceedings of SIGGRAPH 2000*, pages 145–156, July 2000.

[6] P. E. Debevec. *Modeling and Rendering Architecture from Photographs*. PhD thesis, University of California at Berkeley, Computer Science Division, Berkeley CA, 1996. http://www.debevec.org/Thesis.

[7] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH 97*, pages 369–378, August 1997.

[8] P. E. Debevec, C. J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *SIGGRAPH 96*, pages 11–20, August 1996.

[9] P. E. Debevec, Y. Yu, and G. D. Borshukov. Efficient view-dependent image-based rendering with projective texture-mapping. In *9th Eurographics workshop on Rendering*, pages 105–116, June 1998.

[10] A. Gardner, C. Tchou, T. Hawkins, and P. Debevec. Linear light source reflectometry. In *Proceedings of SIGGRAPH 2003*, Computer Graphics Proceedings, Annual Conference Series, July 2003.

[11] T. Hawkins, J. Cohen, and P. Debevec. A photometric approach to digitizing cultural artifacts. In *Proc. 2nd International Symposium on Virtual Reality, Archaeology, and Cultural Heritage (VAST 2001)*, pages 333–342, December 2001.

[12] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital michelangelo project: 3d scanning of large statues. *Proceedings of SIGGRAPH 2000*, pages 131–144, July 2000.

[13] S. Marschner. *Inverse Rendering for Computer Graphics*. PhD thesis, Cornell University, August 1998.

[14] W. Matusik, H. Pfister, A. Ngan, P. Beardsley, R. Ziegler, and L. McMillan. Image-based 3d photography using opacity hulls. *ACM Transactions on Graphics*, 21(3):427–437, July 2002.

[15] M. Naimark, J. Woodfill, P. Debevec, and L. Villareal. Immersion '94. http://www.debevec.org/Immersion/, 1994.

[16] H. Rushmeier, F. Bernardini, J. Mittleman, and G. Taubin. Acquiring input for rendering at appropriate levels of detail: Digitizing a pietà. *Eurographics Rendering Workshop 1998*, pages 81–92, June 1998.

[17] Y. Yu, P. Debevec, J. Malik, and T. Hawkins. Inverse global illumination: Recovering reflectance models of real scenes from photographs. In *SIGGRAPH 99*, August 1999.

[18] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *European Conference on Computer Vision*, pages 151–158, May 1994.