Modeling and Rendering Architecture from Photographs

Paul Debevec

University of California at Berkeley
http://www.cs.berkeley.edu/~debevec
debevec@cs.berkeley.edu

Notes for SIGGRAPH 99 Course #28
3D Photography
Organized by Brian Curless and Steve Seitz
August 10, 1999

Contents

This section of the course notes is organized as follows:

1. Introductory material for this section. This includes a brief overview of related and complimentary material to photogrammetric modeling, such as structure from motion, stereo correspondence, shape from silhouettes, camera calibration, laser scanning, and image-based rendering.
2. A bibliography of related papers.
3. A reprint of:

   Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. *Modeling and Rendering Architecture from Photographs.* In SIGGRAPH 96, August 1996, pp. 11-20.
4. Notes on photogrammetric recovery of arches and surfaces of revolution written by George Borshukov.
5. Copies of the slides used for the presentation.

More information can be found in [10], [5], and [13], available at:
`http://www.cs.berkeley.edu/~debevec/Thesis`

1 Introduction

The creation of three-dimensional models of existing architectural scenes with the aid of the computer has been commonplace for some time, and the resulting models have been both entertaining virtual environments as well as valuable visualization tools. Large-scale efforts have pushed the campuses of Iowa State University, California State University – Chico, and swaths of downtown Los Angeles [23] through the graphics pipeline. Unfortunately, the modeling methods employed in such projects are very labor-intensive. They typically involve surveying the site, locating and digitizing architectural plans (if available), and converting existing CAD data (if available). Moreover, the renderings of such models are noticeably computer-generated; even those that employ large number of texture-maps generally fail to resemble real photographs.

Already, efforts to build computer models of architectural scenes have produced many interesting applications in computer graphics; a few such projects are shown in Fig. 1. Unfortunately, the traditional methods of constructing models (Fig. 2a) of existing architecture, in which a modeling program is used to manually position the elements of the scene, have several drawbacks. First, the process is extremely labor-intensive, typically involving surveying the site, locating and digitizing architectural plans (if available), or converting existing CAD data (again, if available). Second, it is difficult to verify whether the resulting model is accurate.

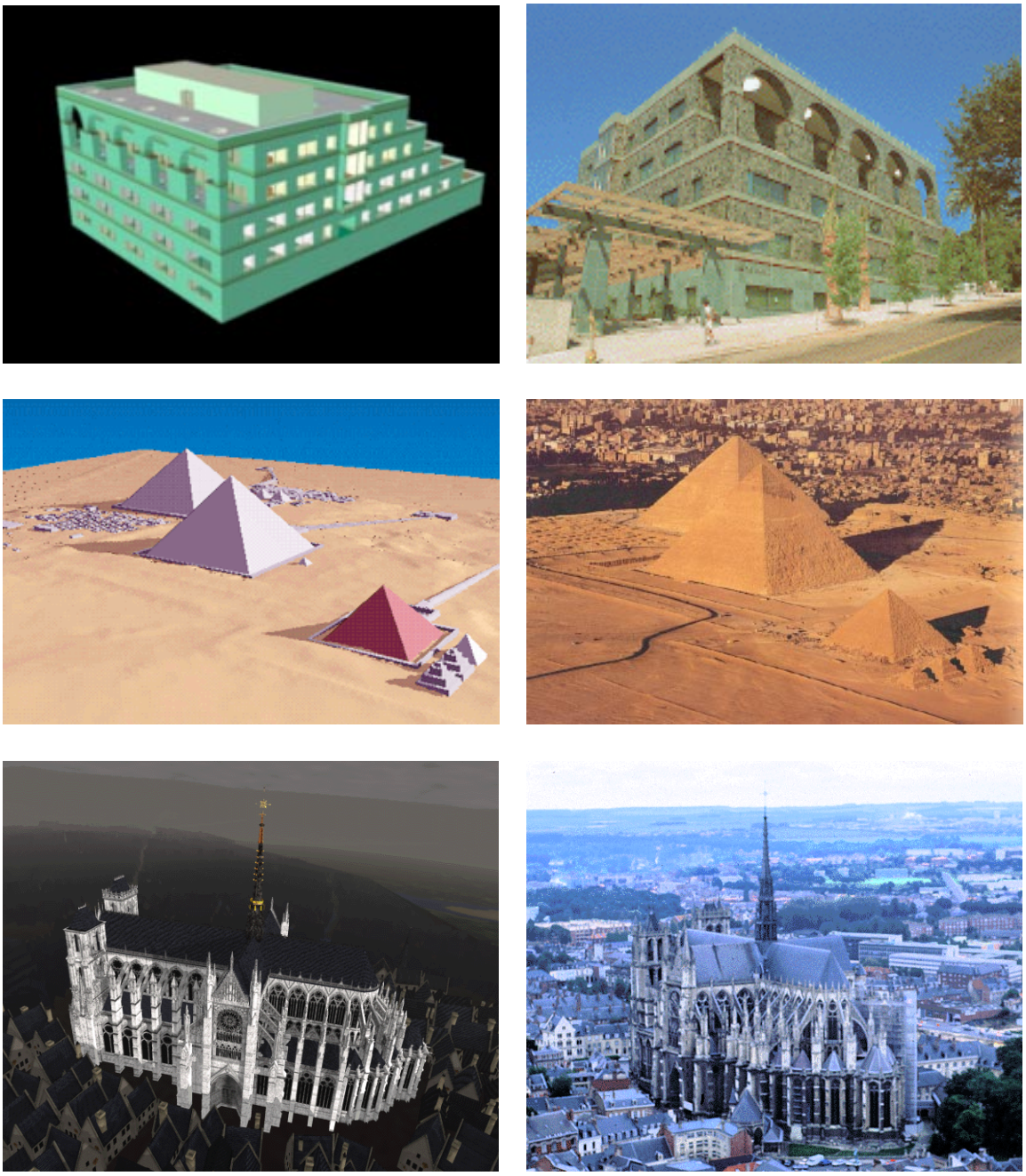

Figure 1: Three ambitious projects to model architecture with computers, each presented with a rendering of the computer model and a photograph of the actual architecture. Top: Soda Hall Walkthru Project [47, 19], University of California at Berkeley. Middle: Giza Plateau Modeling Project, University of Chicago. Bottom: Virtual Amiens Cathedral, Columbia University. Using traditional modeling techniques (Fig. 2a), each of these models required many person-months of effort to build, and although each project yielded enjoyable and useful renderings, the results are qualitatively different from actual photographs of the architecture.

Most disappointing, though, is that the renderings of the resulting models are noticeably computer-generated; even those that employ liberal texture-mapping generally fail to resemble real photographs. As a result, it is easy to distinguish the computer renderings from the real photographs in Fig. 1.

Recently, creating models directly from digital images has received increased interest in both computer vision and in computer graphics under the title of image-based modeling and rendering. Since real images are used as input, such an image-based system (Fig. 2c) has an advantage in producing photorealistic renderings as output. Some of these promising systems (e.g. [26, 32, 31, 44, 39], see also Figs. 3 and 4) employ the computer vision technique of computational stereopsis to automatically determine the structure of the scene from the multiple photographs available. As a consequence, however, these systems are only as strong as the underlying stereo algorithms. This has caused problems because state-of-the-art stereo algorithms have a number of significant weaknesses; in particular, the photographs need to have similar viewpoints for reliable results to be obtained. Because of this, current image-based techniques must use many closely spaced images, and in some cases employ significant amounts of user input for each image pair to supervise the stereo algorithm. In this framework, capturing the data for a realistically renderable model would require an impractical number of closely spaced photographs, and deriving the depth from the photographs could require an impractical amount of user input. These concessions to the weakness of stereo algorithms would seem to bode poorly for creating large-scale, freely navigable virtual environments from photographs.

The techniques presented in these notes aim to make the process of obtaining basic models of architectural scenes more convenient, more accurate, and more photorealistic than the methods currently available. The approach developed draws on the strengths of both geometry-based and image-based methods, as illustrated in Fig. 2b. The result is that our approach to modeling and rendering architecture requires only a sparse set of photographs and can produce realistic renderings from arbitrary viewpoints. In our approach, a basic geometric model of the architecture is recovered semi-automatically with an easy-to-use photogrammetric modeling system (explained in the following reprinted paper [12]), novel views are created using view-dependent texture mapping [12, 13], and additional geometric detail can be recovered through model-based stereo correspondence [12, 10]. The final images can be rendered with current image-based rendering techniques or with traditional texture-mapping hardware. Because only photographs are required, our approach to modeling architecture is neither invasive nor does it require architectural plans, CAD models, or specialized instrumentation such as surveying equipment, GPS sensors or laser range scanners.

2 Work Related to Photogrammetric Modeling

The process of recovering 3D structure from 2D images has been a central endeavor within computer vision, and the process of rendering such recovered structures is an emerging topic in computer graphics. Although no general technique exists to derive models from images, several areas of research have provided results that are applicable to the problem of modeling and rendering architectural scenes. The particularly relevant areas reviewed here are: Camera Calibration, Structure from Motion, Shape from Silhouette Contours, Stereo Correspondence, and Image-Based Rendering.

2.1 Camera calibration

Recovering 3D structure from images becomes a simpler problem when the images are taken with *calibrated* cameras. For our purposes, a camera is said to be *calibrated* if the mapping between image coordinates and directions relative to the camera center are known. However, the position of the camera in space (i.e. its translation and rotation with respect to world coordinates) is not necessarily known. An excellent presentation of the algebraic and matrix representations of perspective cameras may be found in [17].

Considerable work has been done in both photogrammetry and computer vision to calibrate cameras and lenses for both their perspective intrinsic parameters and their distortion patterns. Some successful methods include [49], [16], and [15]. While there has been recent progress in the use of uncalibrated views for 3D reconstruction [18], this method does not consider non-perspective camera distortion which prevents high-precision results for images taken with real cameras. We have found camera calibration to be a straightforward

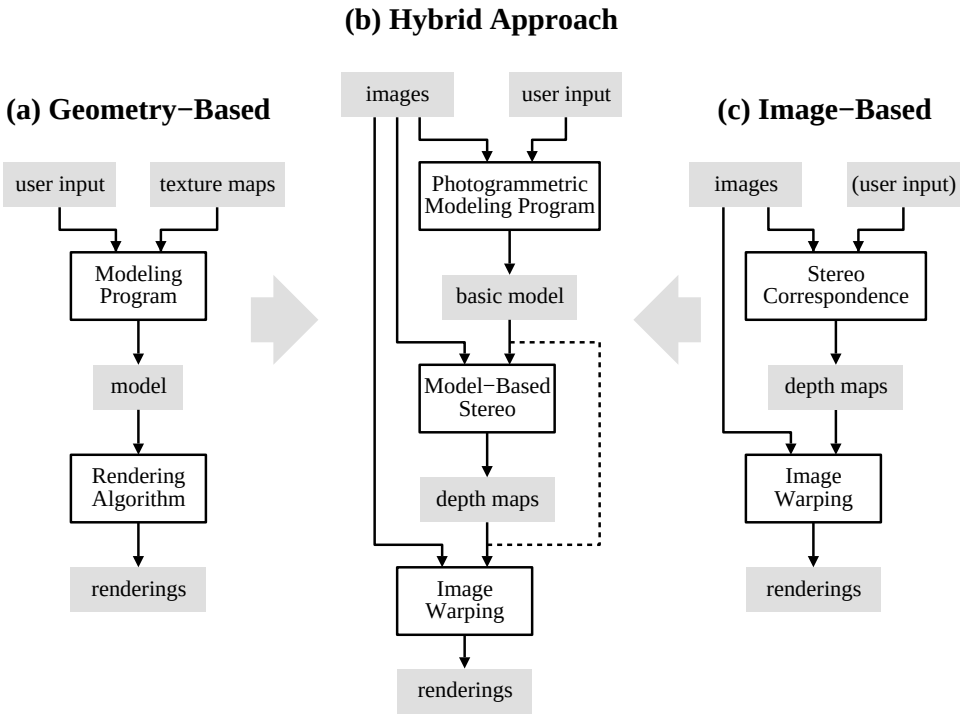

Figure 2: Schematic of how our hybrid approach combines geometry-based and image-based approaches to modeling and rendering architecture from photographs. The geometry-based approach illustrated places the majority of the modeling task on the user, whereas the image-based approach places the majority of the task on the computer. Our method divides the modeling task into two stages, one that is interactive, and one that is automated. The dividing point we have chosen capitalizes on the strengths of both the user and the computer to produce the best possible models and renderings using the fewest number of photographs. The dashed line in the geometry-based schematic indicates that images may optionally be used in a modeling program as texture-maps. The dashed line in the image-based schematic indicates that in some systems user input is used to initialize the stereo correspondence algorithm. The dashed line in the hybrid schematic indicates that view-dependent texture-mapping (discussed later in these notes and in [10, 13, 36]) can be used without performing stereo correspondence.

Figure 3: The Immersion ’94 [32] stereo image sequence capture rig, being operated by Michael Naimark of Interval Research Corporation. Immersion ’94 was one project that attempted to create navigable, photorealistic virtual environments from photographic data. The stroller supports two identical 16mm movie cameras, and has an encoder on one wheel to measure the forward motion of the rig. The cameras are motor-driven and can be programmed to take pictures in synchrony at any distance interval as the camera rolls forward. For much of the work done for the Immersion project, the forward motion distance between acquired stereo pairs was one meter.

process that considerably simplifies the problem of 3D reconstruction, although the methods presented here can also solve for focal lengths and other intrinsic parameters if necessary. [10], Chapter 4 provides a more detailed overview of the issues involved in camera calibration and discusses the camera calibration process used in this work.

2.2 Structure from motion

Given the 2D projection of a point in the world, its position in 3D space could be anywhere on a ray extending out in a particular direction from the camera’s optical center. However, when the projections of a sufficient number of points in the world are observed in multiple images from different positions, it is mathematically possible to deduce the 3D locations of the points as well as the positions of the original cameras, up to an unknown factor of scale.

This problem has been studied in the area of photogrammetry for the principal purpose of producing topographic maps. In 1913, Kruppa [25] proved the fundamental result that given two views of five distinct points, one could recover the rotation and translation between the two camera positions as well as the 3D locations of the points (up to a scale factor). Since then, the problem’s mathematical and algorithmic aspects have been explored starting from the fundamental work of Ullman [51] and Longuet-Higgins [29], in the early 1980s. Faugeras’s book [17] overviews the state of the art as of 1992. So far, a key realization has been that the recovery of structure is very sensitive to noise in image measurements when the translation between the available camera positions is small.

Attention has turned to using more than two views with image stream methods such as [48] or recursive approaches [1]. Tomasi and Kanade [48] (see Fig. 5) showed excellent results for the case of orthographic cameras, but direct solutions for the perspective case remain elusive. In general, linear algorithms for the problem fail to make use of all available information while nonlinear optimization methods are prone to difficulties arising from local minima in the parameter space. An alternative formulation of the problem by Taylor and Kriegman [46] (see Fig. 6) uses lines rather than points as image measurements, but the previously stated concerns were shown to remain largely valid. For purposes of computer graphics, there is yet another problem: the models recovered by these algorithms consist of sparse point fields or individual line segments, which are

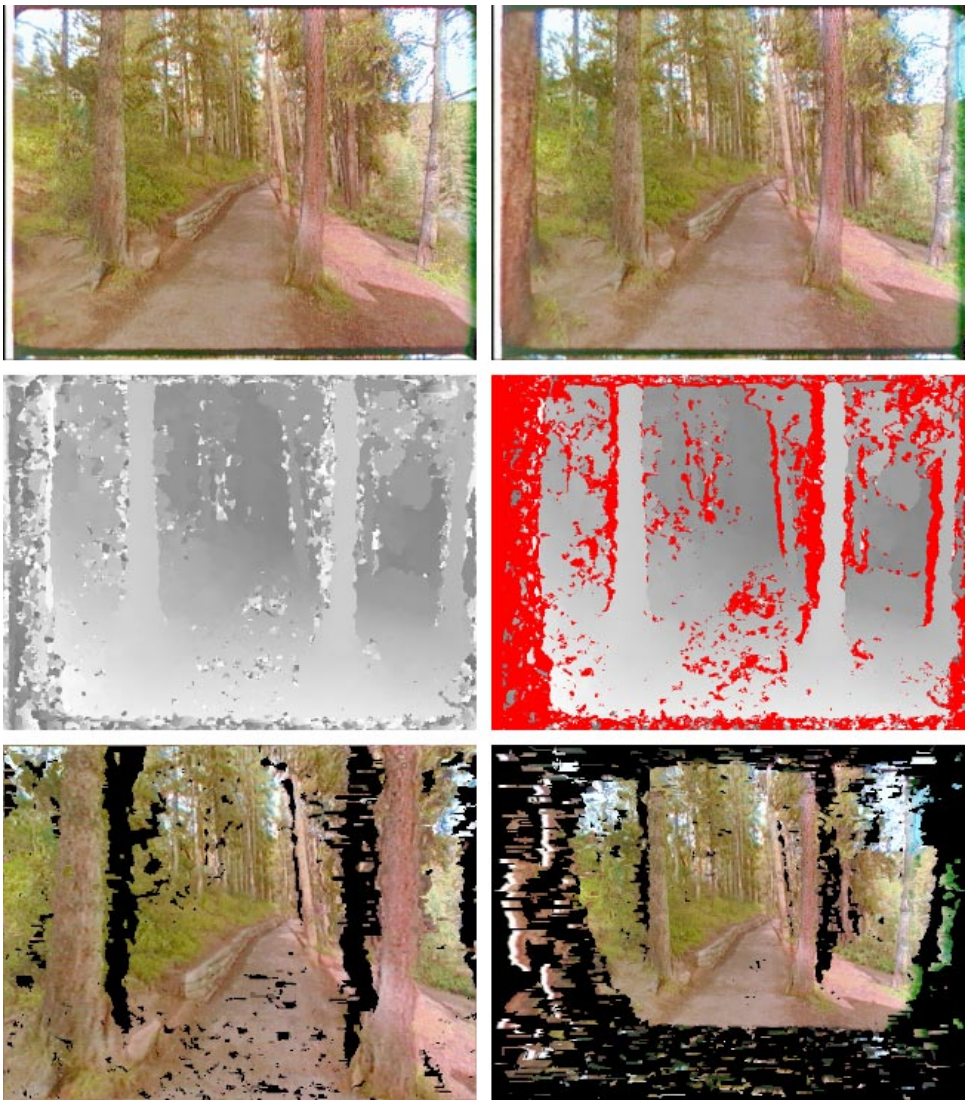

Figure 4: The Immersion ’94 [32] image-based modeling and rendering (see Fig. 2c) project. The top two photos are a stereo pair (reversed for cross-eyed stereo viewing) taken with the apparatus in Fig. 3 in Canada’s Banff National Forest. The film frame was overscanned to assist in image registration. The middle left photo is a stereo disparity map produced by a parallel implementation of the Zabih-Woodfill stereo algorithm [55]. 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 reimaging 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 de-occluded areas produced by these virtual camera moves; these areas were not seen in the original stereo pair. In the Immersion ’94 animations, these regions were automatically filled in from neighboring stereo pairs.

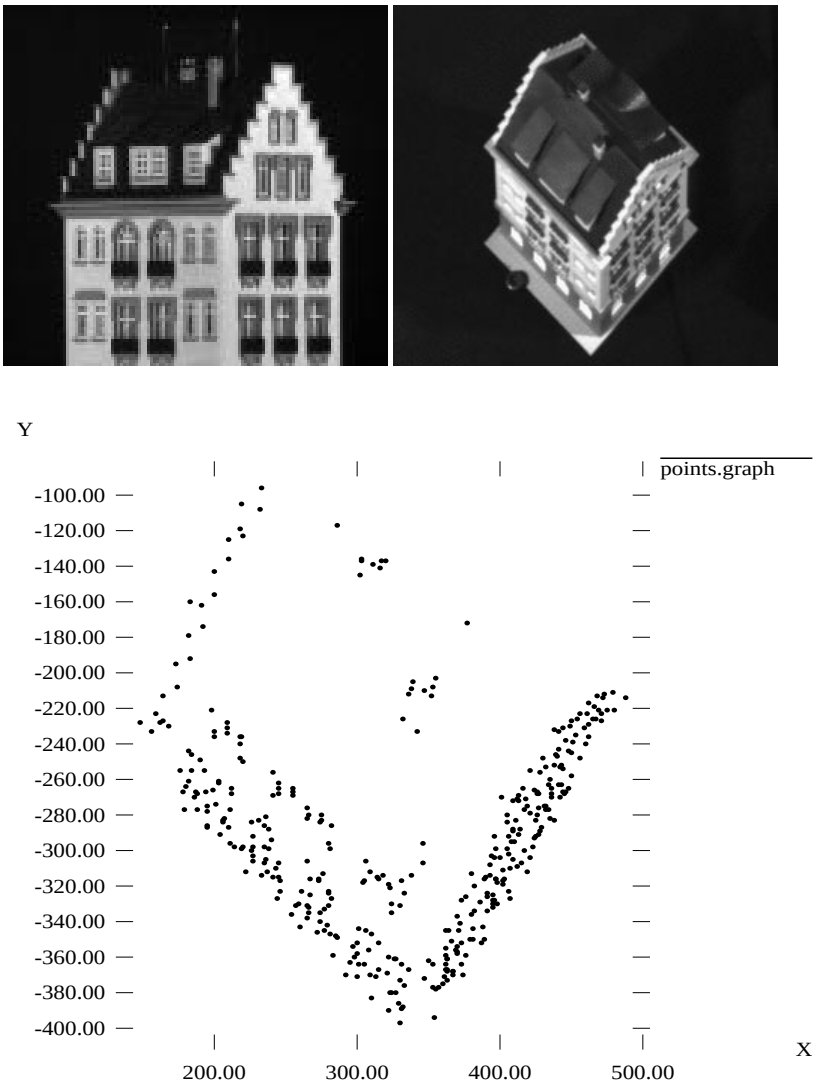

Figure 5: Images from the 1992 Tomasi-Kanade structure from motion paper [48]. In this paper, feature points were automatically tracked in an image sequence of a model house rotating. By assuming the camera was orthographic (which was approximated by using a telephoto lens), they were able to solve for the 3D structure of the points using a linear factorization method. The above left picture shows a picture from the original sequence, the above right picture shows a second image of the model from above (not in the original sequence), and the plot below shows the 3D recovered points from the same camera angle as the above right picture. Although an elegant and fundamental result, this approach is not directly applicable to real-world scenes because real camera lenses (especially those typically used for architecture) are too wide-angle to be approximated as orthographic.

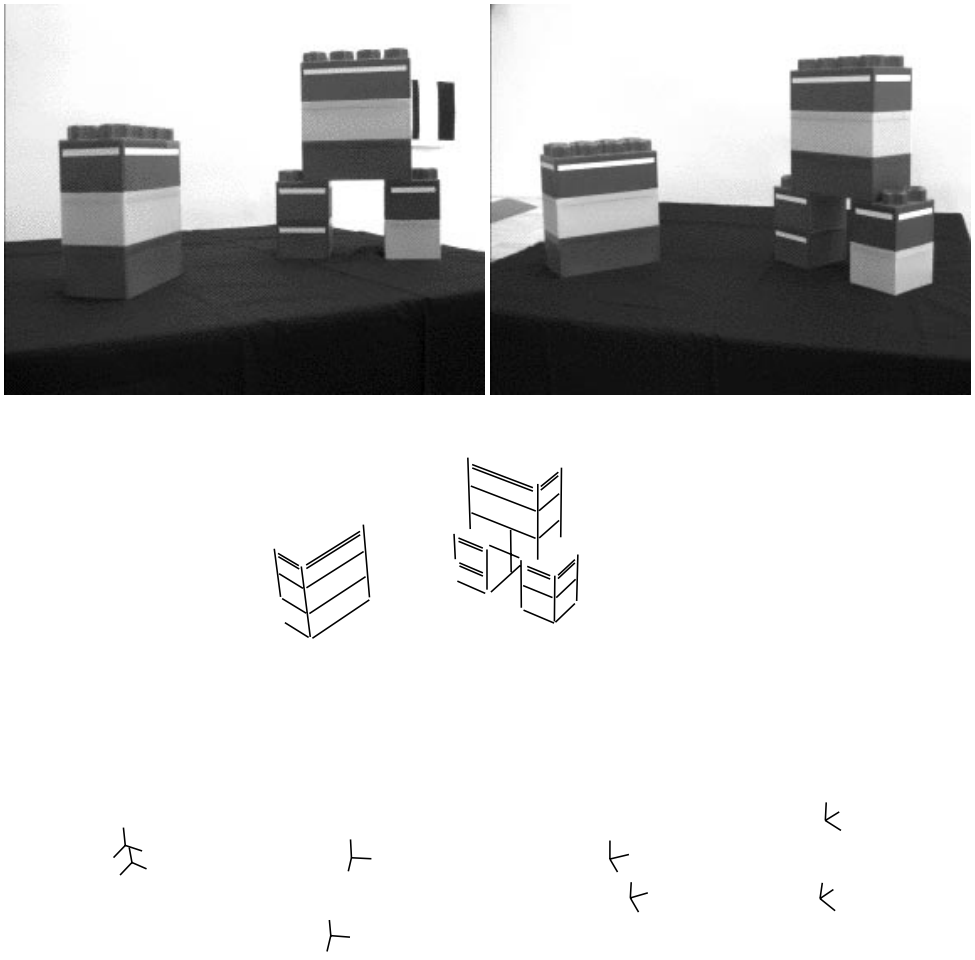

Figure 6: Images from the 1995 Taylor-Kriegman structure from motion paper [46]. In this work, structure from motion is recast in terms of line segments rather than points. A principal benefit of this is that line features are often more easily located in architectural scenes than point features. Above are two of eight images of a block scene; edge correspondences among the images were provided to the algorithm by the user. The algorithm then employed a nonlinear optimization technique to solve for the 3D positions of the line segments as well as the original camera positions, show below. This work used calibrated cameras, but allowed a full perspective model to be used in contrast to Tomasi and Kanade [48]. However, the optimization technique was prone to getting caught in local minima unless good initial estimates of the camera orientations were provided. This work was extended to become the basis of the photogrammetric modeling method presented in this section of these notes.

not directly renderable as solid 3D models.

In our approach, we exploit the fact that we are trying to recover geometric models of architectural scenes, not arbitrary three-dimensional point sets. This enables us to include additional constraints not typically available to structure from motion algorithms and to overcome the problems of numerical instability that plague such approaches. Our approach is demonstrated in an interactive system for building architectural models from photographs, described in the following paper.

2.3 Shape from silhouette contours

Some work has been done in both computer vision and computer graphics to recover the shape of objects from their silhouette contours in multiple images. If the camera geometry is known for each image, then each contour defines an infinite, cone-shaped region of space within which the object must lie. An estimate for the geometry of the object can thus be obtained by intersecting multiple such regions from different images. As a greater variety of views of the object are used, this technique can eventually recover the ray hull[1] of the object. A simple version of the basic technique was demonstrated in [8], shown in Fig. 7. In this project, three nearly orthographic photographs of a car were used to carve out its shape, and the images were mapped onto this geometry to produce renderings. Although just three views were used, the recovered shape is close to the actual shape because the views were chosen to align with the mostly boxy geometry of the object. A project in which a continuous stream of views was used to reconstruct object geometry is presented in [45, 44]; see also Fig. 8. A similar silhouette-based technique was used to provide an approximate estimate of object geometry to improve renderings in the Lumigraph image-based modeling and rendering system [20].

In modeling from silhouettes, qualitatively better results can be obtained for curved objects by assuming that the object surface normal is perpendicular to the viewing direction at every point of the contour. Using this constraint, [43] developed a surface fitting technique to recover curved models from images.

In general, silhouette contours can be used effectively to recover approximate geometry of individual objects, and the process can be automated if there is known camera geometry and the objects can be automatically segmented out of the images. Silhouette contours can also be used very effectively to recover the precise geometry of surfaces of revolution in images. However, for the general shape of an arbitrary building that has many sharp corners and concavities, silhouette contours alone can not provide adequately accurate model geometry.

Although not adequate for general building shapes, silhouette contours could be useful in recovering the approximate shapes of trees, bushes, and topiary in architectural scenes. Techniques such as those presented in [35] could then be used to synthesize detailed plant geometry to conform to the shape and type of the original flora. This technique would seem to hold considerably more promise for practically recovering plant structure than trying to reconstruct the position and coloration of each individual leaf and branch of every tree in the scene.

2.4 Stereo correspondence

The geometrical theory of structure from motion assumes that one is able to solve the *correspondence* problem, which is to identify the points in two or more images that are projections of the same point in the world. In humans, corresponding points in the two slightly differing images on the retinas are determined by the visual cortex in the process called binocular stereopsis. Two terms used in reference to stereo are *baseline* and *disparity*. The baseline of a stereo pair is the distance between the camera locations of the two images. Disparity refers to the difference in image location between corresponding features in the two images, which is projectively related to the depth of the feature in the scene.

Years of research (e.g. [2, 14, 21, 24, 30, 33, 34]) have shown that determining stereo correspondences by computer is difficult problem. In general, current methods are successful only when the images are similar in appearance, as in the case of human vision, which is usually obtained by using cameras that are closely spaced

[1] The ray hull of an object is the complement of the union of all rays in space which do not intersect the object. The ray hull can capture some forms of object concavities, but not, in general, complicated concave structure.

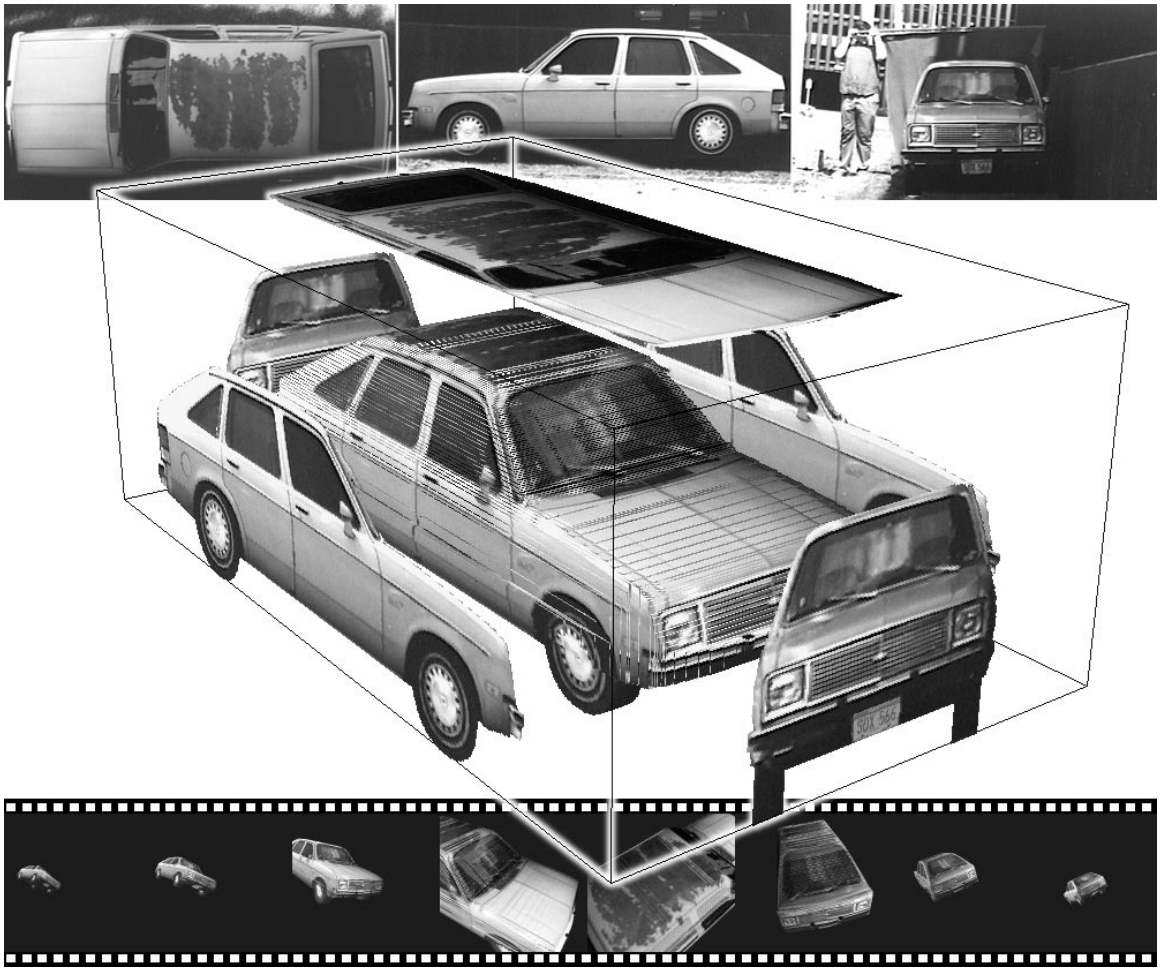

Figure 7: Images from the 1991 Chevette Modeling project [8]. The top three images show pictures of the 1980 Chevette photographed with a 210mm lens from the top, side, and front. The Chevette was semi-automatically segmented from each image, and these images were then registered with each other approximating the projection as orthographic. The registered photographs are shown placed in proper relation to each other on the faces of a rectangular box in the center of the figure. The shape of the car is then carved out from the box volume by perpendicularly sweeping each of the three silhouettes like a cookie-cutter through the box volume. The recovered volume (shown inside the box) is then textured-mapped by projecting the original photographs onto it. The bottom of the figure shows a sampling of frames from a synthetic animation of the car flying across the screen. Although (and perhaps because) the final model has flaws resulting from specularities, missing concavities, and imperfect image registration, it unequivocally evokes an uncanny sense of the actual vehicle.

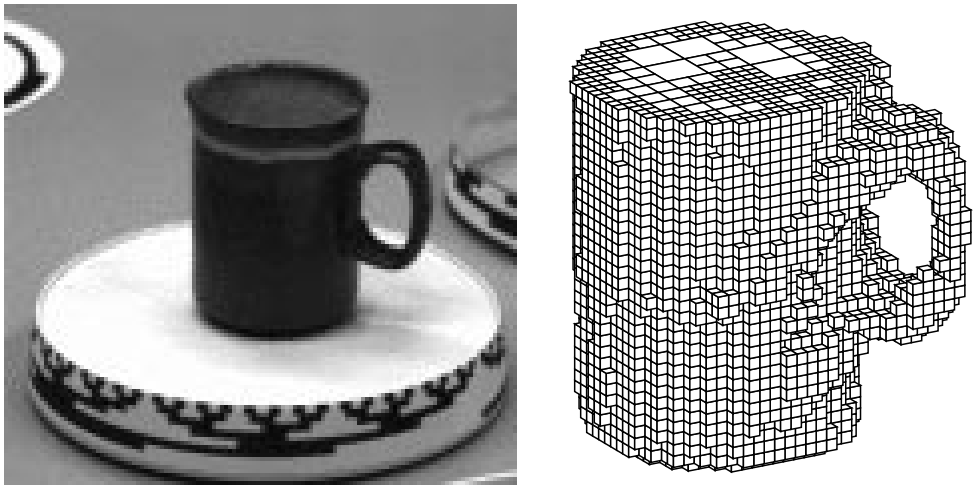

Figure 8: Images from a silhouette modeling project by Rick Szeliski [45, 44]. The cup was videotaped on a rotating platform (left), and the extracted contours from this image sequence were used to automatically recover the shape of the cup (right).

relative to the objects in the scene. As the distance between the cameras (often called the *baseline*) increases, surfaces in the images exhibit different degrees of foreshortening, different patterns of occlusion, and large disparities in their locations in the two images, all of which makes it much more difficult for the computer to determine correct stereo correspondences. To be more specific, the major sources of difficulty include:

1. **Foreshortening**. Surfaces in the scene viewed from different positions will be foreshortened differently in the images, causing the image neighborhoods of corresponding pixels to appear dissimilar. Such dissimilarity can confound stereo algorithms that use local similarity metrics to determine correspondences.

2. **Occlusions**. Depth discontinuities in the world can create half-occluded regions in an image pair, which also poses problems for local similarity metrics.

3. **Lack of Texture**. Where there is an absence of image intensity features it is difficult for a stereo algorithm to correctly find the correct match for a particular point, since many point neighborhoods will be similar in appearance.

Unfortunately, the alternative of improving stereo correspondence by using images taken from nearby locations has the disadvantage that computing depth becomes very sensitive to noise in image measurements. Since depth is computed by taking the inverse of disparity, image pairs with small disparities tend to give rise to noisy depth estimates. Geometrically, depth is computed by triangulating the position of a matched point from its imaged position in the two cameras. When the cameras are placed close together, this triangle becomes very narrow, and the distance to its apex becomes very sensitive to the angles at its base. Noisy depth estimates mean that novel views will become visually unconvincing very quickly as the virtual camera moves away from the original viewpoint[2].

Thus, computing scene structure from stereo leaves us with a conundrum: image pairs with narrow baselines (relative to the distance of objects in the scene) are similar in appearance and make it possible to auto-

[2] The error present in a synthetic view as a function of stereo correspondence accuracy can be described as the **re-rendering equation**. If the novel view is at the same position as the original view, then no amount of depth estimation error can effect the appearance of the re-rendered view; it will always be the same as the original view up to rotation. However, if the novel view is displaced up to one baseline away from the original view, then a stereo correspondence error of n pixels will cause up to n pixels of error in the reprojected position of the mis-corresponded pixel. For a displacement up to k baselines away from the original viewpoint, the reprojection error will be up to kn pixels in the reprojected view, with this bound realized if the camera motion is parallel to the baseline between the cameras. Thus, it is advisable to limit novel viewpoints to be within a few baselines of the original views, lest correspondence errors distort the images very noticeably.

matically compute stereo correspondences, but give noisy depth estimates. Image pairs with wide baselines can give very accurate depth localization for matched points, but the images usually exhibit large disparities, significant regions of occlusion, and different forms of foreshortening which makes it very difficult to automatically determine correspondences.

In these notes, we help address this problem by showing that having an approximate model of the photographed scene can be used to robustly determine stereo correspondences from images taken from widely varying viewpoints. Specifically, the model enables us to warp the images to eliminate unequal foreshortening and to predict major instances of occlusion *before* trying to find correspondences. This technique is a generalization of the plane-plus-parallax parameterization [38] which we call *model-based stereo*; it is presented in the following paper and in [10], Chapter 7.

2.5 Range scanning

Instead of the approach of using multiple images to reconstruct scene structure, an alternative technique is to use range imaging sensors (e.g. [4]) to directly measure depth to various points in the scene. Range imaging sensors determine depth either by triangulating the position of a projected laser stripe, or by measuring the time of flight of a directional laser pulse. While existing versions of these sensors are generally slow, cumbersome and expensive, active development of this technology is making it of practical use for more and more applications. Indeed, the improved practicality of these devices combined with their amazing resolution and range precision will advocate their use in more and more modeling projects. In particular, the Digital Michaelangelo project [27] being directed by Professor Marc Levoy of Stanford University will undoubtedly serve as a watershed event in the practical use of laser range devices and digital photography for creating realistic models of both objects and environments for computer graphics applications.

Algorithms for combining multiple range images from different viewpoints have been developed both in computer vision [53, 42, 41] and in computer graphics [22, 50, 6]; see also Fig. 9. In many ways, range image based techniques and photographic techniques are complementary and each have advantages and disadvantages. Some advantages of modeling from photographic images are that (a) still cameras are inexpensive and widely available, (b) for some architecture that no longer exists (historic buildings, disassembled film sets) all that is available are photographs, and (c) photogrammetry works at arbitrary distances, and is always eye-safe. Of course, geometry alone is insufficient for producing realistic renderings of a scene; photometric information from photographs is also necessary. In general, any image-based rendering technique that can work with geometry acquired from photogrammetry or stereo can work equally well or better with geometry acquired from range scanning.

2.6 Image-based modeling and rendering

In traditional image-based rendering systems, the model consists of a set of images of a scene and their corresponding depth maps. When the depth of every point in an image is known, the image can be re-rendered from any nearby point of view by projecting the pixels of the image to their proper 3D locations and reprojecting them onto a new image plane. Thus, a new image of the scene is created by warping the images according to their depth maps. A principal attraction of image-based rendering is that it offers a method of rendering arbitrarily complex scenes with a constant amount of computation required per pixel. Using this property, [52] demonstrated how regularly spaced synthetic images (with their computed depth maps) could be warped and composited in real time to produce a virtual environment.

In the Immersion '94 project [32], (Fig. 4) stereo photographs with a baseline of eight inches were taken every meter along a trail in a forest. Depth was extracted from each stereo pair using a census stereo algorithm [55]. Novel views were produced by supersampled z-buffered forward pixel splatting based on the stereo depth estimate of each pixel. ([26] describes a different rendering approach that implicitly triangulated the depth maps.) By manually determining relative camera pose between successive stereo pairs, it was possible to optically combine re-renderings from neighboring stereo pairs to fill in missing texture information. The project was able to produce very realistic synthetic views looking forward along the trail from any position within a meter of the original camera path, which was adequate for producing a realistic virtual experience of

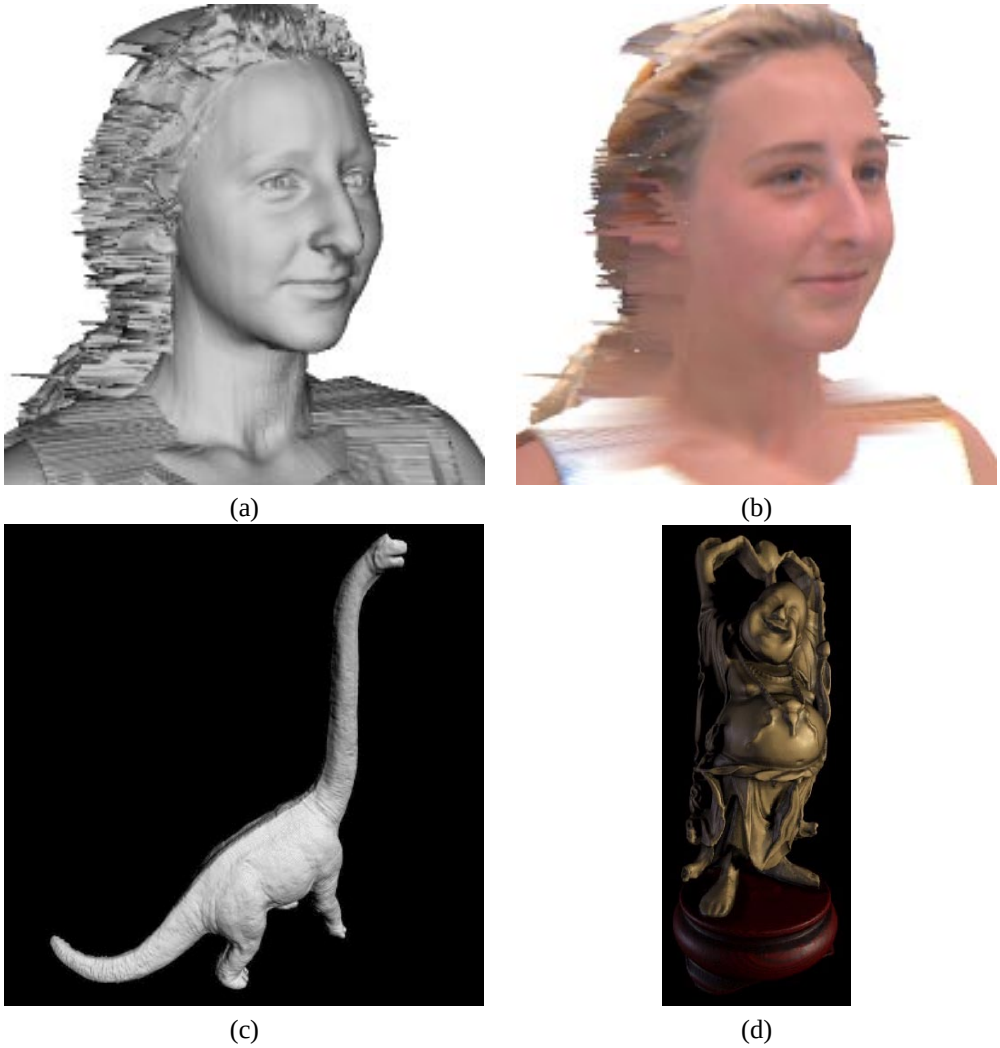

Figure 9: Several models constructed from triangulation-based laser range scanning techniques. **(a)** A model of a person's head scanned using a commercially available Cyberware laser range scanner, using a cylindrical scan. **(b)** A texture-mapped version of this model, using imagery acquired by the same video camera used to detect the laser stripe. **(c)** A more complex geometry assembled by zippering together several triangle meshes obtained from separate linear range scans of a small object from [50]. **(d)** An even more complex geometry acquired from approximately sixty range scans using the volumetric recovery method in [6].

walking down the trail. Thus, for mostly linear environments such as a forest trail, this method of capture and rendering seems promising.

[31] presented a real-time image-based rendering system that used panoramic photographs with depth computed, in part, from stereo correspondence. One observation of the paper was that extracting reliable depth estimates from stereo is very difficult. The method was nonetheless able to obtain acceptable results for nearby views using user input to aid the stereo depth recovery: the correspondence map for each image pair was seeded with 100 to 500 user-supplied point correspondences and also post-processed. Even with user assistance, the images used still had to be closely spaced; the largest baseline described in the paper was five feet.

The requirement that samples be close together is a serious limitation to generating a freely navigable virtual environment. Covering the size of just one city block could require thousands of panoramic images spaced five feet apart. Clearly, acquiring so many photographs is impractical. Moreover, even a dense lattice of ground-based photographs would only allow renderings to be generated from within a few feet of the original camera level, precluding any virtual fly-bys of the scene. Extending the dense lattice of photographs into three dimensions would clearly make the acquisition process even more difficult.

The modeling and rendering approach described in these notes takes advantage of the structure in architectural scenes so that only a sparse set of photographs can be used to recover both the geometry and the appearance of an architectural scene. As an example, the approach was able to create a virtual fly-around of the UC Berkeley bell tower and the surrounding campus from just twenty photographs (see the following slides and the web site http://www.cs.berkeley.edu/~debevec/Campanile).

Some research done concurrently with the work presented here [3] also shows that taking advantage of architectural constraints can simplify image-based scene modeling. This work specifically explored the constraints associated with the cases of parallel and coplanar edge segments.

An interesting aspect of image-based modeling and rendering is that the accuracy of the geometry can traded off with the number of images acquired and the freedom of movement attainable. [40] for example, uses no explicit geometry but rather a set of correspondences between two views of a scene to generate arbitrary views intermediate to the two original ones. And [20, 28] blend between a very large array images of an object or scene in a view-dependent manner to create the appearance of a 3D object, when the actual geometry being used can be as simple as a single plane passing through the object.

3 Conclusion

The philosophy of the work presented here is that geometry is a good thing to have, and that it should be acquired as accurately as possible. The particular techniques presented here make it possible to acquire the basic geometry for many sorts of architectural scenes, using just a set of still photographs and some effort by a trained user of the system. With the geometry recovered, the full realism of the scene can be rendered by projecting the original photographs onto the geometry, preferably combining them with a form of view-dependent texture mapping. Note that this technique can only render the scene in the original lighting conditions, and that it will not be able to convincingly render particularly shiny surfaces, which change in appearance too much with angle to be captured adequately in a sparse set of views. Addressing these problems requires obtaining estimates of the lighting conditions and material properties of the scene, which is the subject of work in image-based lighting [11, 9], BRDF recovery [7, 37], and Inverse Global Illumination [54].

More extensive information on Image-Based Modeling, Rendering, and Lighting and how it relates to 3D Photography may be found in the SIGGRAPH 99 Course notes for Course #39, “Image-Based Modeling, Rendering, and Lighting”.

Acknowledgments

Many thanks to C.J. Taylor, Jitendra Malik, George Borshukov, Yizhou Yu, and Golan Levin for their contributions to this work, and to Interval Research Corporation, the National Science Foundation, Silicon Graphics,

the California MICRO program, Rockwell International, the ONR MURI Program, and the JSEP program for their sponsorship.

References

[1] Ali Azarbayejani and Alex Pentland. Recursive estimation of motion, structure, and focal length. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(6):562–575, June 1995.

[2] H. H. Baker and T. O. Binford. Depth from edge and intensity based stereo. In *Proceedings of the Seventh IJCAI, Vancouver, BC*, pages 631–636, 1981.

[3] Shawn Becker and V. Michael Bove Jr. Semiautomatic 3-d model extraction from uncalibrated 2-d camera views. In *SPIE Symposium on Electronic Imaging: Science and Technology*, Feb 1995.

[4] P. Besl. Active optical imaging sensors. In J. Sanz, editor, *Advances in Machine Vision: Architectures and Applications*. Springer Verlag, 1989.

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

[6] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *SIGGRAPH 96*, pages 303–312, 1996.

[7] K. J. Dana, B. Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real-world surfaces. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 151–157, 1997.

[8] Paul Debevec. The Chevette Project. Summer 1991.

[9] Paul 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.

[10] Paul E. Debevec. *Modeling and Rendering Architecture from Photographs*. PhD thesis, University of California at Berkeley, Computer Science Division, Berkeley CA, 1996. http://www.cs.berkeley.edu/~debevec/Thesis.

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

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

[13] Paul E. Debevec, Yizhou Yu, and George D. Borshukov. Efficient view-dependent image-based rendering with projective texture-mapping. In George Drettakis and Nelson Max, editors, *9th Eurographics workshop on Rendering, Vienna, Austria*, pages 105–116, June 1998.

[14] D.J.Fleet, A.D.Jepson, and M.R.M. Jenkin. Phase-based disparity measurement. *CVGIP: Image Understanding*, 53(2):198–210, 1991.

[15] O.D. Faugeras, Q.-T. Luong, and S.J. Maybank. Camera self-calibration: theory and experiments. In *European Conference on Computer Vision*, pages 321–34, 1992.

[16] Oliver Faugeras and Giorgio Toscani. The calibration problem for stereo. In *Proceedings IEEE CVPR 86*, pages 15–20, 1986.

[17] Olivier Faugeras. *Three-Dimensional Computer Vision*. MIT Press, 1993.

[18] Olivier Faugeras, Stephane Laveau, Luc Robert, Gabriella Csurka, and Cyril Zeller. 3-d reconstruction of urban scenes from sequences of images. Technical Report 2572, INRIA, June 1995.

[19] Thomas Funkhauser and C. H. Sequin. Adaptive display algorithm for interactive frame rates during visualization of complex virtual environments. In *SIGGRAPH 93*, pages 247–254, 1993.

[20] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. The Lumigraph. In *SIGGRAPH 96*, pages 43–54, 1996.

[21] W. E. L. Grimson. *From Images to Surface*. MIT Press, 1981.

[22] H. Hoppe, T. DeRose, T. DUchamp, M. Halstead, H. Jin, J. McDonald, J. Schweitzer, and W. Stuetzle. Piecewise smooth surface reconstruction. In *ACM SIGGRAPH 94 Proc.*, pages 295–302, 1994.

[23] William Jepson, Robin Liggett, and Scott Friedman. An environment for real-time urban simulation. In *Proceedings of the Symposium on Interactive 3D Graphics*, pages 165–166, 1995.

[24] D. Jones and J. Malik. Computational framework for determining stereo correspondence from a set of linear spatial filters. *Image and Vision Computing*, 10(10):699–708, December 1992.

[25] E. Kruppa. Zur ermittlung eines objectes aus zwei perspektiven mit innerer orientierung. *Sitz.-Ber. Akad. Wiss., Wien, Math. Naturw. Kl., Abt. Ila.*, 122:1939–1948, 1913.

[26] Stephane Laveau and Olivier Faugeras. 3-D scene representation as a collection of images. In *Proceedings of 12th International Conference on Pattern Recognition*, volume 1, pages 689–691, 1994.

[27] Marc Levoy. The Digital Michaelangelo Project. http://www-graphics.stanford.edu/projects/mich/, 1999.

[28] Marc Levoy and Pat Hanrahan. Light field rendering. In *SIGGRAPH 96*, pages 31–42, 1996.

[29] H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, September 1981.

[30] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proceedings of the Royal Society of London*, 204:301–328, 1979.

[31] Leonard McMillan and Gary Bishop. Plenoptic Modeling: An image-based rendering system. In *SIGGRAPH 95*, 1995.

[32] Michael Naimark, John Woodfill, Paul Debevec, and Leo Villareal. Immersion '94. Interval Research Corporation image-based modeling and rendering project from Summer 1994, presented at SIGGRAPH 95 Panel "Museums Without Walls: New Media for New Museums".

[33] H. K. Nishihara. Practical real-time imaging stereo matcher. *Optical Engineering*, 23(5):536–545, 1984.

[34] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby. A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14:449–470, 1985.

[35] Przemyslaw Prusinkiewicz, Mark James, and Radomir Mech. Synthetic topiary. In *SIGGRAPH 94*, pages 351–358, July 1994.

[36] Kari Pulli, Michael Cohen, Tom Duchamp, Hugues Hoppe, Linda Shapiro, , and Werner Stuetzle. View-based rendering: Visualizing real objects from scanned range and color data. In *Proceedings of 8th Eurographics Workshop on Rendering, St. Etienne, France*, pages 23–34, June 1997.

[37] Yoichi Sato, Mark D. Wheeler, and Katsushi Ikeuchi. Object shape and reflectance modeling from observation. In *SIGGRAPH 97*, pages 379–387, 1997.

[38] Harpreet S. Sawhney. Simplifying motion and structure analysis using planar parallax and image warping. In *International Conference on Pattern Recognition*, 1994.

[39] D. Scharstein. Stereo vision for view synthesis. In *Computer Vision and Pattern Recognition*, June 1996.

[40] Steven M. Seitz and Charles R. Dyer. View morphing. In *SIGGRAPH 96*, pages 21–30, August 1996.

[41] H. Shum, M. Hebert, K. Ikeuchi, and R. Reddy. An integral approach to free-formed object modeling. *ICCV*, pages 870–875, 1995.

[42] M. Soucy and D. Lauendeau. Multi-resolution surface modeling from multiple range views. In *Proc. IEEE Computer Vision and Pattern Recognition*, pages 348–353, 1992.

[43] Steve Sullivan and Jean Ponce. Constructing 3d object models from photographs. Technical Sketch, Siggraph 1996, Unpublished.

[44] R. Szeliski. Image mosaicing for tele-reality applications. In *IEEE Computer Graphics and Applications*, 1996.

[45] Richard Szeliski and Rich Weiss. Robust shape recovery from occluding contours using a linear smoother. Technical Report 93/7, Digital Equipment Corporation, December 1993.

[46] Camillo J. Taylor and David J. Kriegman. Structure and motion from line segments in multiple images. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(11), November 1995.

[47] S. J. Teller and C. H. Sequin. Visibility preprocessing for interactive walkthroughs. In *SIGGRAPH 91*, pages 61–69, 1991.

[48] Carlo Tomasi and Takeo Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, November 1992.

[49] Roger Tsai. A versatile camera calibration technique for high accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, August 1987.

[50] Greg Turk and Marc Levoy. Zippered polygon meshes from range images. In *SIGGRAPH 94*, pages 311–318, 1994.

[51] S. Ullman. *The Interpretation of Visual Motion*. The MIT Press, Cambridge, MA, 1979.

[52] Lance Williams and Eric Chen. View interpolation for image synthesis. In *SIGGRAPH 93*, 1993.

[53] Y.Chen and G. Medioni. Object modeling from multiple range images. *Image and Vision Computing*, 10(3):145–155, April 1992.

[54] Yizhou Yu, Paul Debevec, Jitendra Malik, and Tim Hawkins. Inverse global illumination: Recovering reflectance models of real scenes from photographs. In *SIGGRAPH 99*, August 1999.

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