

Fast Omnidirectional 3D Scene Acquisition with an Array of Stereo Cameras

Jiajun Zhu, Greg Humphreys, David Koller
Dept. of Computer Science
Univ. of Virginia
{jz8p,humper,koller}@cs.virginia.edu

Skip Steuart
Steuart Systems
skip@steuart.com

Rui Wang
Dept. of Computer Science
Univ. of Massachusetts
ruiwang@cs.umass.edu

Abstract

We present an omnidirectional 3D acquisition system based on a mobile array of high-resolution consumer digital SLR cameras that automatically capture high dynamic range stereo pairs across a full 360-degree panorama. The stereo pairs are augmented with a time-varying lighting pattern created using standard photographic flashes, lenses, and patterned slides. Spacetime stereo techniques are used to generate 3D range images with corresponding color data from the HDR photographs. The multiple range images are aligned with egomotion estimation and ICP registration techniques, and volumetric merging and color texturing algorithms allow the rapid creation of complete 3D models. The resulting system compares favorably with other state of the art 3D acquisition technologies in the resolution and quality of its output, and can be faster and less expensive than 3D laser scanners for digitizing large 3D scenes such as building interiors.

1. Introduction

A wide variety of methods have been developed for 3D digitization of real-world scenes, including optical triangulation and time-of-flight laser scanning, active structured light techniques, photogrammetry, and stereo reconstruction. However, 3D modeling of large scenes with these devices is still expensive, cumbersome, and requires a high level of expertise. The cost and difficulty of 3D laser scanning, for example, limits its practice by potential user groups such as archaeologists who want to digitize cultural heritage sites, or law enforcement agencies who might want to capture crime scenes in 3D for forensic analysis.

Recent trends in digital camera technology point to a coming revolution in 3D shape acquisition. Just a few years ago, digital cameras and lenses of sufficient resolution and image quality to rival the 3D shape acquisition capability of laser scanners were prohibitively expensive. However, high-end consumer digital cameras are now sufficiently robust to produce comparable high quality 3D models, even

using a passive vision approach.

In this paper, we describe the design and implementation of a camera-based 3D scene acquisition system. The system uses a mobile array of off-the-shelf digital SLR cameras to capture high dynamic range (HDR) stereo pairs over a 360 degree horizontal field of view. By augmenting the stereo pairs with projected light patterns and applying spacetime stereo reconstruction techniques, we are able to generate high-resolution, accurate range images of the environment with perfectly corresponding HDR color texture.

As we will show, such a stereo camera-based acquisition system can be cheaper, more portable, and easier to use than traditional 3D scanning systems (such as time-of-flight laser technologies), without compromising on quality. In fact, our approach often performs better than laser scanning for areas with challenging reflection properties such as low albedo, gloss, or translucency. Furthermore, surface textures captured from stereo vision are perfectly registered with the resulting surface, resulting in no alignment error. Also, we achieve much faster acquisition times, as both 3D geometry and color information are acquired in a full panorama with each discrete exposure, as opposed to the relatively slow speed of laser systems that must mechanically sweep a laser line or beam across the surface for full coverage.

Other recent research in 3D acquisition has related goals. For example, robotics researchers have developed self-propelled robots with panoramic cameras to reconstruct 3D models of their environment [10, 18]. Hand-held devices based on video cameras have been used for interactive 3D scene modeling [26, 31]. Some research has focused on methods for real-time 3D model acquisition [28]; this is not an explicit goal of our system.

Other related work pertaining to specific aspects of our approach is referenced in the corresponding sections of this paper. In Section 2 below, we describe our 3D acquisition hardware design. In Section 3, we describe our software pipeline for the creation of complete, color 3D models from the digital color images. Section 4 includes experimental results from using our system to digitize 3D scenes, and includes comparisons with other state of the art 3D scanning

devices. In Section 5 of the paper, we discuss the ramifications of these results, and the potential advantages and disadvantages of our camera-based approach relative to other 3D acquisition methods.

2. Acquisition Hardware Design

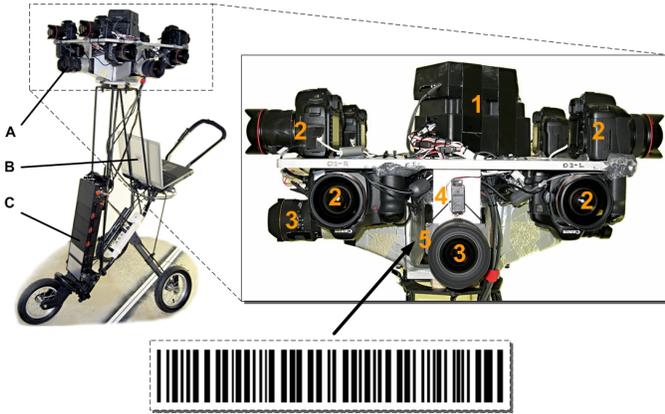


Figure 1. Our current prototype 3D acquisition device and the projected pattern: (A) System head; (B) Control laptop; (C) Strobe power pack for the flash. The zoomed image on the right shows the head with eight digital SLR cameras mounted on the aluminum frame: (1) USB controlled relays, USB controlled servo control, and USB hubs; (2) Camera and 14mm rectilinear lens; (3) Projection lens; (4) Pattern slide servo; (5) Pattern slide.

Our design prototype for a portable omnidirectional 3D acquisition device consists of an array of digital cameras mounted on a three-wheeled cart (Figure 1).

The primary components in the device are four pairs of Canon 5D 12-megapixel cameras with rectilinear Canon 14mm lenses. The resulting stereoscopic horizontal field of view is 104 degrees for each camera pair. The cameras are screwed onto a half-inch aluminum frame (camera motion is further minimized using epoxy), and are arranged so as to observe a full 360 degree horizontal panorama. Despite rough handling during field trials, the cameras have maintained calibration over periods of several weeks.

To make the system as compact as possible and permit passage through a 29 inch doorway, adjacent pairs of cameras are mounted on opposite sides of the frame; two camera pairs are mounted right-side up, and two upside down. This arrangement allows a stereo base of 14 inches. Although the camera shutters can be programmatically controlled through their USB connections, we ensure synchronized exposures by using an external relay board that is wired to each of the cameras’ external shutter triggers with Canon N3 connector. Another USB cable is attached to

each camera from the control laptop, through which the laptop controls the shutter speed, aperture, and ISO settings, and downloads the acquired image data.

To allow robust shape acquisition of textureless surfaces, we have included the capability to project a high-resolution, time-varying lighting pattern into the environment. A 1600 watt-second strobe projects the pattern across a full 360 degrees using four lenses, each with a 90 degree horizontal field of view. Compared to a standard projector, this provides a brighter light pattern and covers a larger field of view. The strobe can be triggered by the sync port in one camera, or by a USB-controlled strobe relay. The projected pattern is a non-repeating series of vertical lines printed on a 3cm x 7cm heat resistant transparency. In order to create a time-varying pattern to permit spacetime stereo analysis, a USB-controlled servo is used to move the slide laterally between exposures.

3. 3D Acquisition Pipeline

In this section we describe our 3D acquisition and processing steps; Figure 2 outlines the high level data flow of the pipeline.

3.1. Camera Pair Calibration

Because each camera pair uses wide-angle lenses to capture a wide field of view, there is significant nonlinear distortions in the raw acquired images. This lens distortion is first corrected using DXO Optics Pro [17], followed by a standard checkerboard-based stereo calibration procedure (we experimented with the toolbox by Bouguet [7] and also commercial software from Videre [33]; both have worked well). Once the four camera pairs have been calibrated individually, the inter-pair positions and orientations must also be calibrated to allow our scanner to operate as a single unit. This is achieved by scanning a room with calibration targets attached to the walls, and then manually aligning the four resulting point clouds together by establishing 3D point correspondences. Currently, this is the only portion of the processing pipeline that is not automated; we are currently investigating alternate inter-pair calibration procedures to achieve full automation; see Section 6.

3.2. Image Acquisition and Processing

For a single omnidirectional scan, each camera captures 18 exposures (144 images total) in 2.5 minutes. The first three exposures are captured at varying shutter speeds without flash; these images are combined into floating-point high dynamic range (HDR) images [15] and then tone-mapped using HDRSoft PhotoMatix [21] to be used for surface texture. This is followed by another 15 exposures with varying flash patterns. All images are acquired at the full

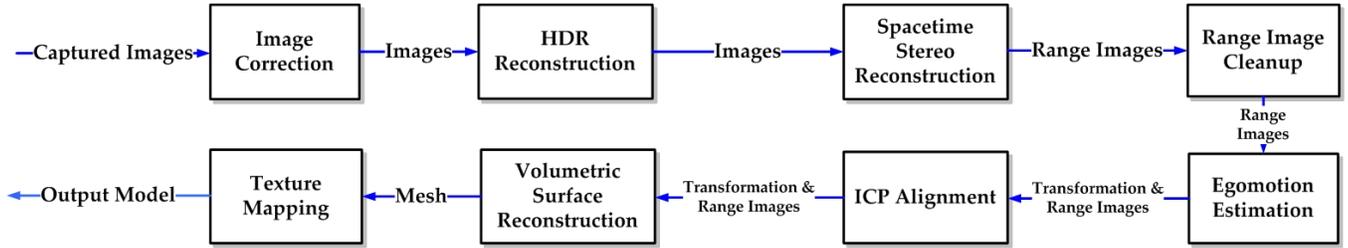


Figure 2. Our data processing pipeline. We first capture HDR images at different exposures and varying flashing patterns. After we correct the lens distortion, rectify the stereo pairs, and recover the HDR information, we use a spacetime stereo algorithm to reconstruct a range image which is then filtered. Multiple range images with their associated color images are used to estimate egomotion based on an image-based registration algorithm. We then use a volumetric surface reconstruction approach to generate dense meshes (with vertex color) from ICP-refined sets of range images. The final texture mapping step is optional.

camera resolution of 4328 x 2912 pixels (Figure 3). We empirically determined that 15 pictures give acceptable results, taking into account the tradeoff between surface quality and processing time that results from additional images.

Each omnidirectional scan thus consists of 18 high resolution images for each camera. Once a scan is completed, the automated control system downloads the 144 images from the eight cameras, removes geometric distortions using DXO Optics Pro [17], and performs image rectification to make all epipolar lines be horizontal. The scanner can then be repositioned and the procedure repeated if multiple 3D scans are desired.



Figure 3. Two example images from a scan of an office. Left: an image used to capture the surface color, without the pattern. Right: one of the 15 images acquired with the flashing pattern.

3.3. Spacetime Stereo

Our camera-based scanning system design is inspired in part by recent work on “spacetime stereo” techniques [13, 38], which generalize the traditional stereo problem to perform stereo matching in both the spatial and temporal domains simultaneously. For scenes with temporally-varying illumination, spacetime stereo can deliver significant accuracy improvements over traditional stereo algorithms. Additionally, it does not require that the lighting be calibrated, as is necessary in some triangulation-based techniques such as structured lighting [8].

We apply spacetime stereo to reconstruct 3D range im-

ages from the acquired set of 18 stereo images from each camera pair. In practice, the spacetime stereo algorithm requires much more memory than traditional stereo to store input images. In our implementation, we load only a few image scanlines at a time during processing to maximize the cache hit rate and incrementally compute stereo disparities.

In order to handle ambiguities that can arise due to occlusions, we estimate the confidence in each depth sample as the ratio between the image correlation at the best and second-best matches discovered by the stereo algorithm. If this ratio is too low (e.g., if the second-best match was still reasonable), then we discard the data point. This greatly reduces the number of small “floating” surface patches typical of stereo reconstructions. The confidence threshold for rejecting samples is unfortunately scene dependent, so the operator may have to adjust this until the results are acceptable.

To further improve the quality of the range images, we remove all 3D points that fall out of a user-specified working range, and finally apply an anisotropic filter [6, 34] to smooth surface noise.

3.4. Multi-Scan Registration

To eliminate occlusions and make a 3D model more complete, many scans must be merged. Multiple scans of a scene may be acquired by moving the scanner to multiple locations and capturing sets of range images. Because the device’s movement is not tracked, automatically registering these multiple scans into a single coordinate system consists of an initial image-based motion estimation, followed by an Iterated Closest Point (ICP) algorithm [4, 5].

3.4.1. Egomotion Estimation

Because our 3D range data are computed in the identical image space as their corresponding 2D input images, we can use image-based registration techniques [35, 27, 3, 22] to estimate the transformations between scans. Here

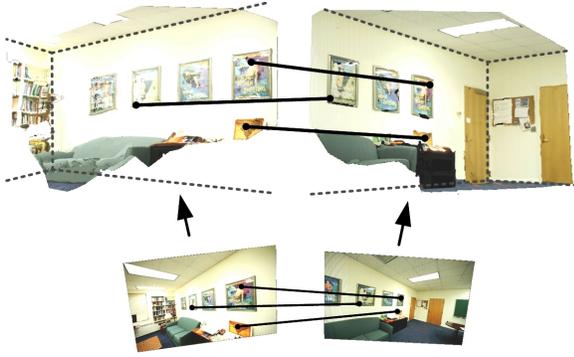


Figure 4. Egomotion estimation using image-based feature matching. 2D features are first detected and matched in the image space, and then back-projected onto the reconstructed 3D point clouds.

we describe our robust image-based egomotion estimation method.

Given two color images and their corresponding dense range images, we start by extracting scale-invariant features from both color images using the SIFT algorithm [23]. Then for every feature in each image, we search for the corresponding feature in the other image which minimizes the Euclidean distance between their feature representations and only keep those best pair-wise matches [9]. This 2D feature matching step can be efficiently implemented by using a *kd*-tree to find approximate nearest neighbors [2, 24]. Similar to Snavely et al. [32], we then estimate the fundamental matrix relating these two color images using a RANSAC-based eight-point algorithm [19]. We then refine the initial matches by rejecting the outliers to this fundamental matrix.

After we robustly match 2D image features between the two color images, we back-project each image feature pair onto their range images and link the corresponding 3D feature points (Figure 4). This gives us N 3D point matches $\{(P_i, Q_i) | P_i \in \text{Scan}_1 \text{ and } Q_i \in \text{Scan}_2, i = 1 \dots N\}$. To compute a rigid transformation to align the two range images, we compute a rotational matrix R and a translational vector T by minimizing the mean squared distance between the transformed match pairs, as described by Arun et al. [1]:

$$f(R, T) = \frac{1}{N} \sum_i^N \|(R \times Q_i + T) - P_i\|^2.$$

This approach is more robust than the previous work by Li [22] for two reasons. First, it extracts and matches SIFT feature representations instead of computing the cross correlation of corner regions; this greatly improves performance when the motion between images involves primarily forward motion as opposed to lateral translation. Also, we add a RANSAC refinement step to reject mismatches based on epipolar geometry.

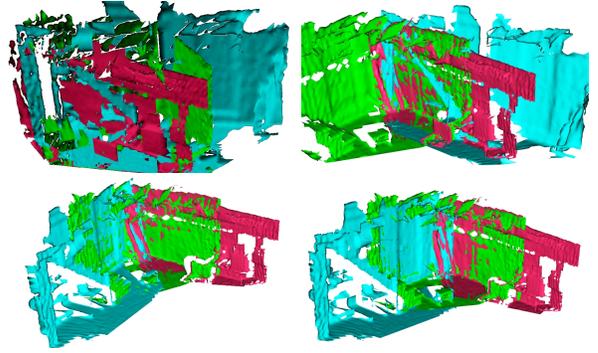


Figure 5. An example of improving 3D registration by egomotion estimation. Top-left: 3 range images in initial position; top-right: attempted ICP registrations; bottom-left: egomotion estimation; bottom-right: ICP-refined registration based on the egomotion estimation.

3.4.2. ICP Alignment

Following the egomotion estimation step, the registration of the individual range scans is further refined via application of the Iterative Closest Point (ICP) algorithm [4]. We use the Stanford *Scanalyze* software [30], which can be scripted to allow implementation in our automated scan processing pipeline.

Figure 5 shows an example of the use of egomotion estimation in scan registration. Just using ICP alone fails because of the bad initial positions, while egomotion estimation provides a well-aligned starting point. Egomotion estimation itself can fail if two scans have very small overlap where automatic image-based feature matching is either impossible or unreliable; this can be avoided by carefully selecting the location for each scan.

3.5. Surface Reconstruction

The registered range scans are typically integrated into a single 3D mesh surface model using a volumetric surface reconstruction method that merges the range data from each scan, as well as the color information [11, 34]. For some applications, conversion to a polygonal mesh may not be required, and the dense point cloud representation can be used directly for visualization or measurement. However, exporting to modeling applications like 3D Studio Max or AutoCAD typically requires a polygonal mesh. Meshing also permits data compression using polygonal simplification techniques.

3.6. Texture Mapping

Texture mapping is optional in our pipeline because a tessellated mesh with vertex colors is frequently dense enough to provide a convincing visualization. However, when a 3D simplified mesh model with texture maps is re-

quired, we simply project each triangle onto every aligned texture image and assign a texture ID and texture coordinates from the image that yields the largest projected area. A view-dependent solution such as the one proposed by Debevec et al. [14] would yield visually better results, but cannot be easily incorporated into existing commercial model viewers.

4. Results

In this section, we present visual and quantitative results that assess the accuracy and performance of our scanning methods. We have used our prototype camera-based device to digitize several scenes. Due to the characteristics of our flash and the melting point of the pattern transparencies, our scanner is optimized for indoor scanning.

Figure 7 shows the reconstructed mesh results from six single scans of various basement and office scenes. Each scan was acquired with an image resolution of 2400×1600 , which are downsampled from the original 4328×2912 images for performance reasons. Although our camera system captures range images from four camera pairs simultaneously to create a panoramic scan, each row in Figure 7 depicts only a single scan captured by one of the camera pairs for ease of visualization.

4.1. Accuracy

To quantitatively measure the scanning accuracy as a function of distance, we scanned a known planar surface from four different distances and at two different image resolutions. As shown in Table 1, the error grows linearly as the distance between the surface and the cameras increases. The mean absolute error is 3.59 mm at 2 meters, and 16.10(mm) at 10 meters, which is comparable to modern pulsed time-of-flight laser scanners. Table 1 also demonstrates that using higher image resolution significantly reduces the error at far distances, as more pixels give finer measurements in the stereo disparity space.

Figure 6 shows a side-by-side comparison of our stereo scanner and a commercial time-of-flight laser scanner, the 3rdTech DeltaSphere-3000 [16]. We acquired two indoor scenes with both scanners at the same locations and with

Distance to target	1 m	2 m	5 m	10 m
1200 x 800 Res. Scan				
Avg. Error (mm)	1.23	3.70	6.70	30.06
Std. Deviation (mm)	1.69	5.44	8.40	51.22
2400 x 1600 Res. Scan				
Avg. Error (mm)	1.11	3.59	6.61	16.10
Std. Deviation (mm)	1.51	4.50	8.19	19.75

Table 1. Scan accuracy (deviation from plane) vs. distance at two different image resolutions.

similar resolutions. It can be seen that the camera-based scanner gives near-laser scan quality. Note that our stereo scan is more complete and less sensitive to the effect of shiny objects (e.g., fire extinguisher, ball, door handles) or dark objects (e.g., chair, partition mullions) because it incorporates local spatial smoothness. However, this also creates slightly blurrier scans compared to the laser scanner. We also observe that our projected lighting patterns bias our sub-pixel estimation and create artifacts such as the repeating patterns on the walls of the stereo scans (this is particularly evident on the door in the bottom row in figure 6). This problem has been recently addressed using a symmetric refinement [25], and we are in the process of incorporating it into our scan processing pipeline, which should greatly reduce these artifacts.

4.2. Scanning Time

Our camera-based stereo scanner requires 2.5 minutes to acquire a full 360 degree scan of 18 exposures per camera (144 images). The two primary bottlenecks are the flash recharge time and the communication time between the cameras and the control laptop. Our flash recharge takes approximately 6 seconds, which places a lower bound of 90 seconds on the scanning time with our current power source.

In future designs, camera communications can be parallelized with a networked group of tiny control PCs. After downloading the raw image data, our system requires approximately one hour to produce a textured 3D mesh using a modern dual-core PC. In comparison, the DeltaSphere laser scanner can complete a panoramic scan at comparable resolution in 10 minutes (with no color), and requires a further 10 minutes to generate the finished polygonal 3D model.

5. Discussion

Here we consider several characteristics of our camera-based stereo scanner approach. Table 2 provides a summarized comparison of the prototype system with the DeltaSphere-3000 time-of-flight laser scanner.

Scanning speed. Our camera-based approach exploits the full pixel-wise parallelization in depth sampling, and thus has the potential to allow extremely fast acquisition times. The major performance bottlenecks for our current prototype are addressable engineering limitations. The flash recharge time can be reduced by using larger capacitors, and we can substantially reduce the communication delay between the cameras and control laptop by parallelizing the communications with multiple USB ports. With these improvements, our next prototype is expected to reduce the acquisition time to less than one minute. Our system will thus be appropriate for use in environments that are ephemeral or dangerous and require minimized digitization time.

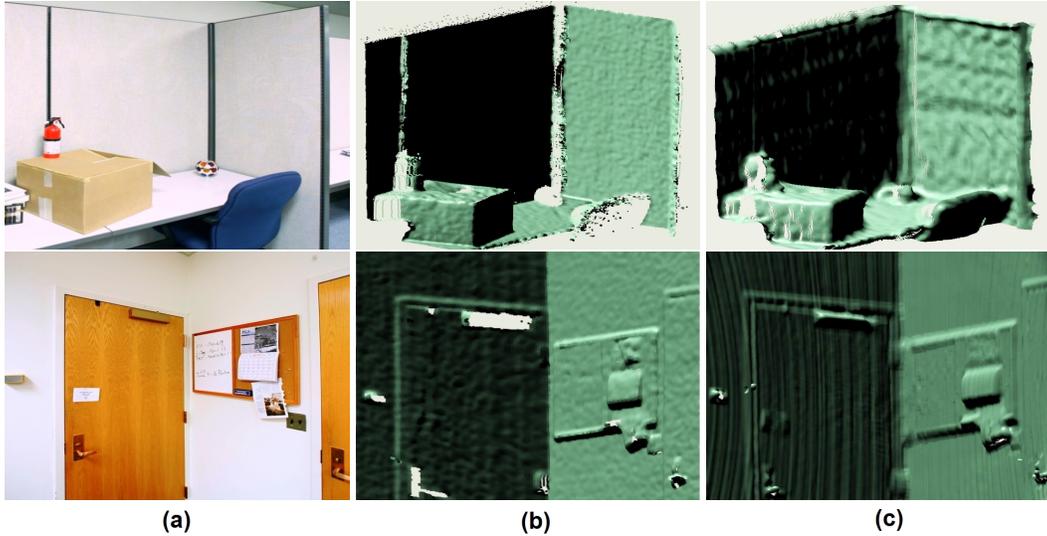


Figure 6. Two comparisons between laser scans and untextured camera-based scans. Columns from left to right: (a) photographs; (b) DeltaSphere laser scan; (c) our camera-based scan.

	DeltaSphere-3000 Laser Scanner	Camera-based Scanner Prototype
Avg. Scanning Time Per Scan	10 min. (without color)	2.5 min. (with HDR images)
Speed bottleneck	laser sampling rate	camera driver and recharge time
Data Processing Time	10 min. / scan	50 min. / scan
Range	0.3 m - 15 m	0.5 m - 12 m (with pattern) 12 m - 340 m (without pattern)
Range bottleneck	laser power, surface reflectance	flash energy, camera resolution, and baseline
Accuracy	7.5 mm at 12 m	16.1 mm at 10 m
Accuracy bottleneck	clock precision	camera and pattern resolution
Cost (Mid 2007, USD)	approx. \$40,000	approx. \$30,000
Color capture	requires a separate scan	captured simultaneously

Table 2. A comparison between our camera-based scanner prototype and DeltaSphere-3000 laser scanner.

Processing speed. After each scan, our current system can output a 3D “preview” model created from down-sampled 300x200 resolution images in less than 2 minutes. However, processing the full four sets of 18 2400x1600 stereo images takes about 50 minutes on a dual-core PC (including image rectification, stereo matching, and range image filtering). This slow processing speed is due to the huge amount of data I/O and the expensive spacetime stereo matching computation. One idea to improve this post-processing performance is to shift most computation onto GPUs [37]. Currently our algorithm is heavily compute limited and spends most of its time in stereo correspondence matching, so additional processing speed or more cores would yield immediate speedups.

Occlusions and model completeness. Almost every scanned scene will have holes due to occlusion. The only way to fill these occlusions is to scan the scene from multiple perspectives, a process made much more practical by the increased acquisition speed of cameras. For example,

to adequately scan all sides of a single rectangular object in a room would require at least three scans; forming a complete model of a complex scene might require dozens of scans. Automatically determining the best location for the next scan remains open research in computer vision. If the scene is not completely scanned, then holes must be filled either by manual editing or using automatic hole filling techniques [12].

Accuracy. There is still great potential to improve our camera-based scan accuracy, although many tradeoffs remain to be evaluated. Taking more pictures and increasing camera resolution will provide more accurate results at the expense of larger scanning and processing times. The design of an optimal pattern is an open problem in active stereo, although ours can certainly be improved.

Cost. Because our scanner is based on high-volume commodity digital camera components, different versions can be created to fit different budgets, a major advantage over the relatively low-volume laser-scanning indus-

try. Digital camera technology continues to decline sharply in price, so the cost of our scanning system should decline similarly. Also, the quality of lens manufacture and digital sensor design continues to rise, so at a fixed price point we expect the quality of our results to increase over time.

Color capture. Most long distance time-of-flight laser scanners require an independent acquisition process for surface color. This increases acquisition time and introduces error due to the misalignment of the range and color information. Our approach acquires perfectly aligned range and color simultaneously within a single scan. In addition to improving renderings, this implicit correspondence enables robust 3D scan registration via our egomotion estimation. While our current approach is limited to capturing color data under the lighting conditions present at acquisition time, we could use calibrated lighting to measure the surface reflectance [29].

Flexibility. Using off-the-shelf camera components in our system allows for flexibility and scalability in the configuration of the acquisition device. Varying the number and arrangement of cameras mounted on the mobile platform can easily match the economic resources available with the demands of capturing particular types of scenes. Users of the system can change the number of images acquired, the speed of pattern shifting, and the pattern resolutions to balance scanning speed and accuracy to meet application-specific requirements.

6. Conclusion and Future Work

We have described a camera-based stereo system for fast and flexible 3D scene acquisition. To produce high quality range data, we acquire a sequence of color images with different exposures and a time-varying lighting pattern, and then use spacetime stereo techniques for 3D reconstruction. Because the high resolution color images are perfectly aligned with their range images, we can employ a robust, image-based registration approach for initializing ICP registration. Our experiments show that our system generates 3D models of large scenes that are sufficient for many potential applications and that rival time-of-flight laser scanners.

We are continuing to develop our camera-based 3D acquisition system. Our current work focuses on reducing the communication cost between cameras and computers to reduce acquisition time, automating the inter-pair calibration procedure by using a linear calibration rod that can be seen by neighboring camera pairs, and improving sub-pixel estimation by using symmetric refinement [25].

Many experiments remain to further improve the quality of our eventual 3D reconstruction, such as the best projection pattern to use, amount of lateral shift between photographs, number of photographs, etc. On the hardware

side, we would like to add additional cameras to each of the four scanning directions to better acquire features that are parallel to our current baseline, as well as a scanning set pointing up so as to create a hemispherical panoramic scan. Finally, we are actively working to solve the outdoor scanning problem by automatically combining naturally-lit color photographs taken during the day and acquired, pattern-projected 3D geometry scanned at night.

Acknowledgements

This work was supported by NSF Grant #0535118 and sponsored research from Steuart Systems. The scanning hardware was provided by Steuart Systems and is based on their patent pending (# 20040246333) omnidirectional stereoscopic camera system. The authors would like to thank Szymon Rusinkiewicz, James Davis, Tim Weyrich, Jason Lawrence, and Diego Nehab for fruitful discussions regarding the performance of our scanning system.

References

- [1] K.S. Arun, T.S. Huang, and S.D. Blostein, "Least-Squares Fitting of Two 3D Point Sets," *IEEE Trans. PAMI*, vol. 9, no. 5, May 1987, pp. 698-700.
- [2] J. Beis and D. Lowe, "Shape Indexing using Approximate Nearest-Neighbor Search in High-Dimensional Spaces," *Proc. CVPR*, 1997, pp. 1000-1006.
- [3] F. Bernardini, I. Martin and H. Rushmeier, "High-quality Texture Reconstruction from Multiple Scans," *IEEE Trans. on Visualization and Computer Graphics*, 7(4):318-332, 2001.
- [4] P. J. Besl and N. D. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. PAMI*, 14(2), Feb. 1992, pp 239-256.
- [5] Y. Chen and G. Medioni. "Object Modeling by Registration of Multiple Range Images," *Proc. IEEE Conf. on Robotics and Automation*, 1991.
- [6] M. Black, G. Sapiro, D. Marimont, and D. Heeger, "Robust Anisotropic Diffusion," *IEEE Trans. on Image Processing*, 7(3):421-432, 1998.
- [7] Y. G. Bouguet, "Camera Calibration Toolbox for Matlab," http://www.vision.caltech.edu/bouguetj/calib_doc/.
- [8] K. L. Boyer and A. C. Kak, "Color-encoded Structured Light for Rapid Active Ranging," *IEEE PAMI*, vol. 9, no. 1, Jan. 1987.
- [9] M. Brown and D. G. Lowe, "Unsupervised 3D Object Recognition and Reconstruction in Unordered Datasets," *Proc. 3DIM*, Jun. 2005, pp. 56-63.
- [10] R. Bunschoten and B. Krose, "Robust Scene Reconstruction from an Omnidirectional Vision System," *IEEE Trans. on Robotics and Automation*, vol. 19, no. 2, pp. 351-357.
- [11] B. Curless and M. Levoy, "A Volumetric Method for Building Complex Models from Range Images," *Proc. of SIGGRAPH*, 1996, pp. 303-312.
- [12] J. Davis, S. Marschner, M. Garr, M. Levoy, "Filling Holes in Complex Surfaces Using Volumetric Diffusion," *Proc. 3DPVT*, 2002

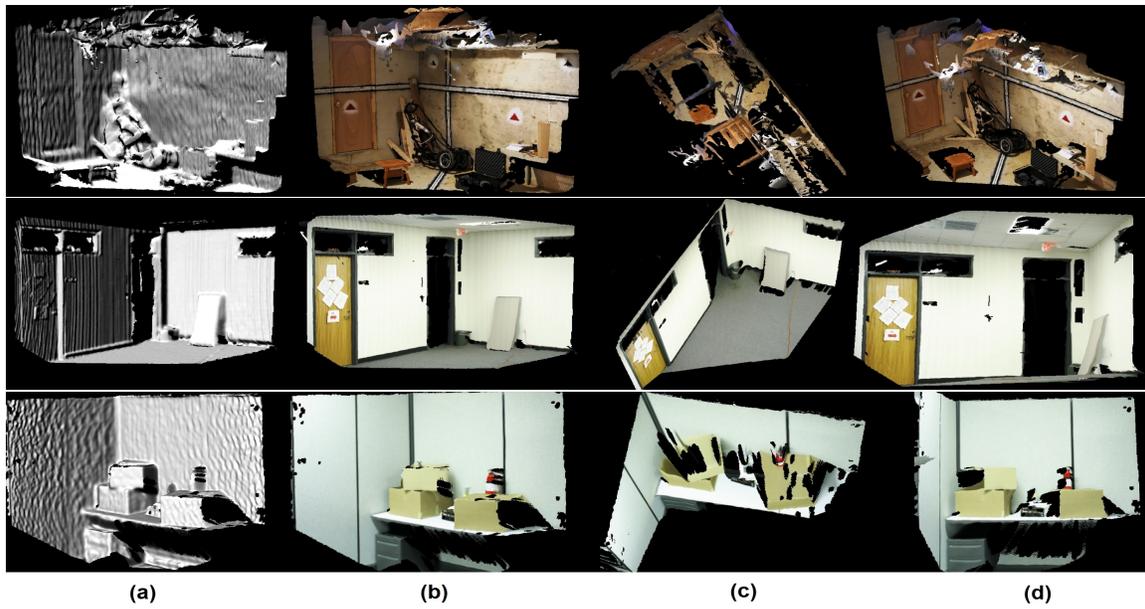


Figure 7. Scan results. Columns from left to right: (a) Reconstructed 3D mesh; (b),(c),(d) the same mesh with vertex colors, rendered from different viewpoints.

- [13] J. Davis, D. Nehab, R. Ramamoorthi, S. Rusinkiewicz, "Spacetime Stereo: A Unifying Framework for Depth from Triangulation," *IEEE Trans. PAMI*, vol. 27, no. 2, Feb. 2005.
- [14] P. E. Debevec, G. Borshukov, and Y. Yu, "Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping," *Proc. 9th Eurographics Rendering Workshop*, June 1998.
- [15] P. E. Debevec and J. Malik, "Recovering High Dynamic Range Radiance Maps from Photographs," *Proc. SIGGRAPH 97*, Aug. 1997.
- [16] DeltaSphere-3000 Scanner, <http://www.deltasphere.com/>.
- [17] DxO Optics Pro, <http://www.dxo.com/intl/photo>.
- [18] S. Fleck, F. Busch, P. Biber, W. Strasser, and H. Andreasson, "Omnidirectional 3D Modeling on a Mobile Robot Using Graph Cuts," *Proc. IEEE ICRA*, Apr. 2005, pp. 1748-1754.
- [19] R. I. Hartley and A. Zisserman, *Multiple View Geometry*, Cambridge University Press, 2004.
- [20] HDRShop, <http://www.hdrshop.com>.
- [21] HDRSoft PhotoMatix, <http://www.hdrsoft.com>.
- [22] T. Li, "A Robust Image-based Method for 3D Registration," *Proc. 3DIM*, 2005, pp. 270-276.
- [23] D. G. Lowe, "Distinctive Image Features from Scale-invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, 2004, pp. 91-110.
- [24] D. M. Mount and S. Arya, "ANN: A Library for Approximate Nearest Neighbor Searching," <http://www.cs.umd.edu/mount/ANN/>.
- [25] D. Nehab, S. Rusinkiewicz, and J. Davis, "Improved Sub-pixel Stereo Correspondences through Symmetric Refinement," *Proc. ICCV*, Oct. 2005, pp. 557-563.
- [26] V. Popescu, E. Sacks, and G. Bahmutov, "The ModelCamera: A Hand-Held Device for Interactive Modeling," *Proc. 3DIM*, 2003, pp. 285-292.
- [27] K. Pulli, "Surface Reconstruction and Display from Range and Color Data," doctoral dissertation, Dept. Computer Science and Engineering, Univ. of Washington, 1997.
- [28] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy, "Real-Time 3D Model Acquisition," *Proc. SIGGRAPH*, 2002, pp. 438-446.
- [29] Y. Sato, M.D. Wheeler, and K. Ikeuchi, "Object Shape and Reflectance Modeling from Observation," *Proc. SIGGRAPH*, 1997, pp. 379-387.
- [30] Scanalyze software package, <http://graphics.stanford.edu/software/scanalyze/>.
- [31] S. Se and P. Jasiobedzki, "Photo-realistic 3D Model Reconstruction," *Proc. IEEE ICRA*, 2006, pp. 3076-3082.
- [32] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo Tourism: Exploring Photo Collections in 3D," *ACM Trans. on Graphics*, vol. 25, no. 3, 2006, pp. 835-846.
- [33] SRI Stereo Engine, <http://www.videredesign.com>.
- [34] R. Wang and D. Luebke, "Efficient Reconstruction of Indoor Scenes with Color," *Proc. 3DIM*, 2003, pp. 402-409.
- [35] S. Weik, "Registration of 3D Partial Surface Models Using Luminance and Depth Information," *Proc. 3DIM*, May 1997, pp. 93-100.
- [36] N. Williams, K. Low, C. Hantak, M. Pollefeys, and A. Latta, "Automatic Image Alignment for 3D Environment Modeling," *Proc. SIBGRAPI*, 2004.
- [37] R. Yang and M. Pollefeys, "Multi-Resolution Real-Time Stereo on Commodity Graphics Hardware," *Proc. CVPR*, 2003, pp. 211-217.
- [38] L. Zhang, B. Curless, and S. M. Seitz, "Spacetime Stereo: Shape Recovery for Dynamic Scenes," *Proc. CVPR*, 2003.