Fast Modeling with a Single-Point Laser Range Finder

Jason Wither
Dept. of Computer Science
UC Santa Barbara
jwither@cs.ucsb.edu

Chris Coffin
Dept. of Computer Science
UC Santa Barbara
ccoffin@cs.ucsb.edu

Jonathan Ventura
Dept. of Computer Science
UC Santa Barbara
jventura@cs.ucsb.edu

Tobias Höllerer
Dept. of Computer Science
UC Santa Barbara
holl@cs.ucsb.edu

ABSTRACT

This paper presents methodology for integrating a small, single-point laser range finder into a wearable augmented reality system. We introduce a method using the laser range finder to incrementally build 3D panoramas from a fixed observer's location. To build a 3D panorama semi-automatically, we track the system's orientation and use the sparse range data acquired as the user looks around in conjunction with real-time image processing to construct geometry around the user's position. Using full 3D panoramic geometry, it is possible for new virtual objects to be placed in the scene with proper lighting and occlusion by real world objects, which increases the expressivity of the AR experience.

1. INTRODUCTION

Mobile Augmented Reality (AR) allows users with wearable or portable computers to view, and interact with, virtual content that is registered to the real world around them. In order to use the full interaction potential of this technology, AR research and application development increasingly focus on going beyond pre-created content, and instead focus on easy online creation of new virtual content. Ideally, a mobile AR system will build up knowledge of the physical environment on the fly so that real world objects can be referenced more easily and virtual annotations can be placed more accurately.

Facilitating online modeling is difficult with traditional outdoor AR equipment however. Existing online modeling techniques are not ideal, as they require walking [3][8] as well as extensive user interaction and/or external data sources. Having a model of the environment is very useful though for things like proper occlusion of virtual object by real-world objects. For proper occlusion it is necessary to have a model of the real-world environment, so that virtual content can disappear (become occluded) at the correct time. In many outdoor systems a model is built as part of a preparatory offline process; however this is very time consuming and doesn't scale well.

Ideally, we would like a system that can provide correct occlusion of annotations with very little effort. We would like this system to fit the framework of *Anywhere Augmentation* [6], requiring only negligible start up cost, no environment instrumentation, and only off-the-shelf hardware components.

In the spirit of this Anywhere Augmentation agenda, we decided to add a small, affordable, single-point laser range finder to our wearable system. With this new interactive sensing modality in place, we can more easily meet the described requirements.

The main contribution of this work is a novel AR technique for 3D world building from a static location. By enabling simpler and faster modeling we view a laser range finder as a promising tool for ubiquitous AR. As hardware continues to shrink a laser range finder could easily be integrated into either hand-held or head-worn devices, opening up new opportunities for user interaction and user-generated content in physical environments.

2. RELATED WORK

Several different approaches have been proposed for the problem of creating a panoramic environment map with depth information. The depth can be specified by the user in an interactive modeling system [9], however as with the user interfaces for single viewpoint images described above, the depth model produced by this type of system is only defined qualitatively. With multiple cameras [10], or a moving camera [7], panoramas with parallax can be automatically produced. More similar to our scenario is that of Bahmutov et al. [2]. They couple a 7x7 laser range finder array with a moving camera to produce highly detailed, textured indoor scene models.

Note the stark difference in focus between our static viewpoint technique and multi-view geometry approaches that recover or track sparse depth in moving user views [4], or recover depth information from landmark features in overlapping photographs [1].

3. HARDWARE AND CALIBRATION

Our testing platform can be seen in Figure 1. The laser range finder we have chosen to use is an Opti-Logic RS400 which gives calibrated range readings continuously at 10 Hz and has a factory-specified range of 400 yards with accuracy of $\pm$ 1 yard. It weighs less than 8 ounces. We are also using a Pt. Grey Firefly MV camera, and a Garmin GPS 18 receiver. For orientation tracking we are using DiVerdi et al.'s Envisor system [5]. Envisor provides completely vision based orientation tracking, using both sparse optical flow for frame to frame features and heavyweight landmark fea-

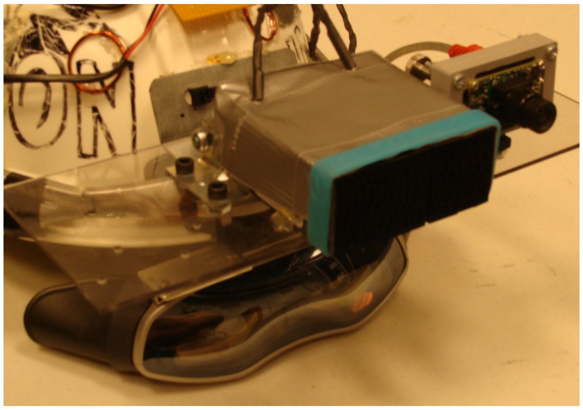

Figure 1: Our hardware on it's development platform. The laser range finder is rigidly mounted to a piece of Lexan, and the camera is carefully calibrated to point in the same direction. Here the development platform is mounted to our wearable system's helmet.

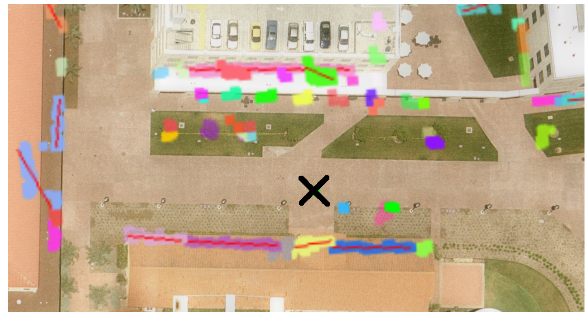

Figure 2: An aerial view of the group and planar information for the data set from Figure 3. Each range point is represented as a dot, and color represents groups. Extracted planes are represented as red lines. The user collected the data set from the black X location. Note that the aerial photograph is not orthorectified. Groups not on building surfaces are due to entryways, trees, or lamp poles.

tures. It builds an environment map on the fly, which we use for image processing.

We calibrate the laser range finder to point parallel to the user's viewing direction. For an outdoor scene this means that the laser will always hit objects near the center of the user's field of view. By measuring the baseline between the laser and camera we can further determine exactly what pixel in the image the laser is hitting depending on the distance returned by the range finder.

4. PANORAMIC DEPTH MAP CONSTRUCTION

As the user looks around the environment, our technique for building panoramic depth maps continually integrates color, depth, and temporal data to refine the estimated 3D model. In our experiments creating a number of panoramic depth maps we found that the process of looking around to completely fill the full 360 degree depth map takes between two and four minutes, providing between one and three thousand depth samples from the laser range finder. This time could likely be reduced with a more robust tracking system.

To begin creating a 3D panorama the user simply looks around. As the user pans the laser range finder, different objects are ranged. Because we only receive sparse depth samples compared to the resolution of the camera, we need to propagate point labels across the image. As a first step, we group the range points. Any time there is a large difference between one depth value and the next we conclude we have observed a group boundary.

This technique for dividing the range points into groups is robust because of the high update rate of the range finder. Two consecutive updates will always have a small angle in between them, so our algorithm is unlikely to change groups when moving the range finder along a single surface. The results of the grouping and plane creation can be seen in Figure 2 from an aerial perspective. Sets of points of a single color represent a single group.

One significant advantage to dividing the range points into groups is the semantic information gained about the spatial layout of the scene. We use this information to seed a diffusion based flood fill process that expands the groups across the image. For this process a confidence value is associated with each pixel, which is highest at known sample points, and initially zero everywhere else. At each iteration, each pixel looks at its 8-connected neighborhood and averages its group (foreground or background) and confidence with neighboring pixels of higher confidence. This diffusion process is weighted by edge information as well.

We automatically detect object boundaries by examining the intensity gradient (using the 3x3 Sobel operator). The measure of boundary $E_p = f(\frac{dI}{dp})$ at pixel p is a function of the magnitude of the gradient. We use $f(x) = x^4$ so that the function will drop off quickly as the gradient decreases. To use the edge information to regulate diffusion the edge value at a neighboring pixel is subtracted from the neighbor's confidence value before being considered by the diffusion algorithm. The result is that pixels on an edge have a very low chance of diffusing their value, effectively stopping the diffusion along boundaries. One particular advantage of this method is its speed; on the GPU, 120 iterations per second can be achieved.

In areas of relatively dense laser range finder samples, the technique gives excellent segmentation along natural image boundaries. In areas with less dense sample resolution, the technique works less well, sometimes stopping short of filling the correct semantic region, or leaking through boundaries into incorrect semantic regions. However, the user can easily improve the results by adding more samples as appropriate.

From the group expansion we produce a group map which labels each pixel with its group number. Now, the depth of each pixel must be determined. To do this we try to model the scene with vertical planar surfaces. This is a reasonable thing to do for urban environments where most objects around a user are buildings. If we find that all of the points in a group are co-planar it is reasonable to assume that they are all on the same planar surface, and use that surface for smooth extrapolation across the whole group. By assuming that all planes are vertical we greatly reduce our search space, allowing us to create accurately oriented planes with a small number of points.

We use a perpendicular regression to find the best fit line through

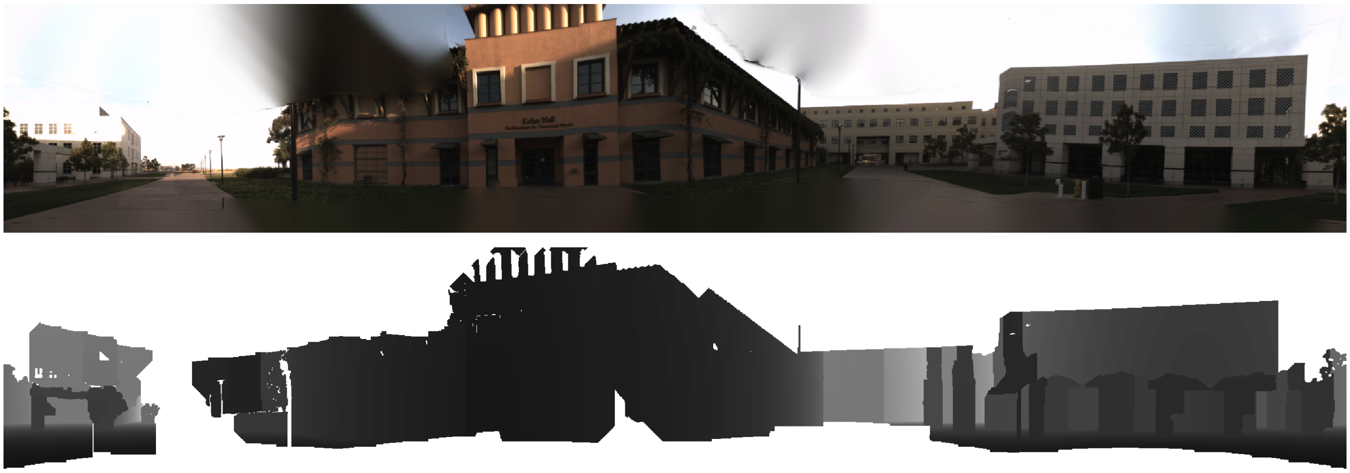

Figure 3: An example color panorama and semi-automatically generated depth map pair. Darker regions of the depth map are closer to the user. To generate both images the user simply has to look around the scene. Both images are composed of the four surrounding faces of a cube map, and are not warped to a cylindrical projection. This cube projection causes the strange peak on the roof line in the center of the images.

Figure 4: A virtual statue occluded by the real world.

the set of 2D points where all heights are ignored. This regression minimizes the sum of squares of the perpendicular distances of each point from the line. We also use RANSAC to throw out any outliers in the set. Outliers are often created by small foreground objects like light poles that partially occlude the the surface of interest and can be ignored.

For a group with no detected planar objects, we take the average depth of the samples in the group as the depth of the entire group. Finally, we use the average height of a user to determine the ground plane, and add that plane to the our complete depth map.

5. CONCLUSIONS AND FUTURE WORK

We have presented results that demonstrate how a single-point laser range finder can improve the AR experience by building a depth map around a static user. The depth maps we create are accurate enough to be useful for a number of AR applications, arguably the most useful of which is automatic occlusion of virtual objects by real world objects.

6. REFERENCES

[1] T. Aoki, T. Tanikawa, and M. Hirose. Virtual 3D world construction by inter-connecting photograph-based 3d models. *Virtual Reality Conference, 2008. VR '08. IEEE*, pages 243–244, 8-12 March 2008.

[2] G. Bahmutov, V. Popescu, and E. Sacks. Depth enhanced panoramas. *Visualization, 2004. IEEE*, pages 11p–11p, 10-15 Oct. 2004.

[3] Y. Baillot, D. Brown, and S. Julier. Authoring of physical models using mobile computers. In *Proceedings of the 5th IEEE International Symposium on Wearable Computers*, pages 39–46, 2001.

[4] A. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proceedings of the International Conference on Computer Vision*, Oct. 2003.

[5] S. DiVerdi, J. Wither, and T. Höllerer. Envisor: Online environment map construction for mixed reality. In *Proc. IEEE VR 2008 (10th IntŠl Conference on Virtual Reality)*, Reno, NV, Mar. 2008.

[6] T. Höllerer, J. Wither, and S. DiVerdi. "Anywhere Augmentation": Towards mobile augmented reality in unprepared environments. In G. Gartner, M. Peterson, and W. Cartwright, editors, *Location Based Services and TeleCartography, Series: Lecture Notes in Geoinformation and Cartography*, pages 393–416. Springer Verlag, Feb. 2007.

[7] S. Peleg and M. Ben-Ezra. Stereo panorama with a single camera. *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, 1:–401 Vol. 1, 1999.

[8] W. Piekarski and B. H. Thomas. Tinmith-metro: New outdoor techniques for creating city models with an augmented reality wearable computer. In *Proceedings of the 5th IEEE International Symposium on Wearable Computers*, pages 31–38, 2001.

[9] H.-Y. Shum, M. Han, and R. Szeliski. Interactive construction of 3d models from panoramic mosaics. *IEEE Computer Vision and Pattern Recognition*, pages 427–433, 23-25 Jun 1998.

[10] H.-Y. Shum and L.-W. He. Rendering with concentric mosaics. In *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 299–306, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co.