Fast Annotation and Modeling with a Single-Point Laser Range Finder

Jason Wither* Chris Coffin† Jonathan Ventura‡ Tobias Höllerer§

Department of Computer Science
University of California, Santa Barbara

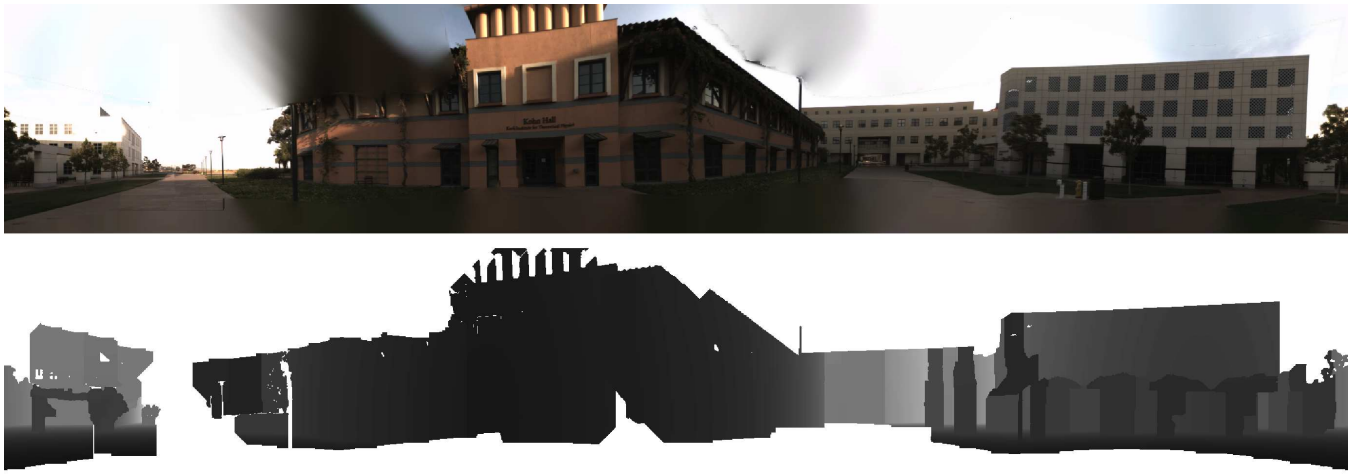

Figure 1: An example color panorama and semi-automatically generated depth map pair. Darker regions of the depth map are closer to the user. To generate both images the user simply has to look around the scene. Both images are composed of the four surrounding faces of a cube map, and are not warped to a cylindrical projection. This cube projection causes the strange peak on the roof line in the center of the images.

ABSTRACT

This paper presents methodology for integrating a small, single-point laser range finder into a wearable augmented reality system. We first present a way of creating object-aligned annotations with very little user effort. Second, we describe techniques to segment and pop-up foreground objects. Finally, 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.

Index Terms: I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Tracking; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality

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 annotation, and instead focus on easy creation of new annotations. 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 simple annotation and modeling is difficult with traditional outdoor AR equipment however. Existing techniques for creating new annotations are not ideal, as they require walking [3] [13] or user interaction and/or external data sources [18][17]. For proper occlusion by real-world objects it is necessary to have a model of the real-world environment. 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 both create new annotations with the press of a button, and provide correct occlusion of annotations with very little effort. We would like this system to fit the framework of *Anywhere Augmentation* [10], 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 contributions of this work are novel AR techniques for three important AR tasks: annotation placement and alignment (section 4), foreground object segmentation (section 5), and 3D world building from a static location (section 6). Each of these use our mobile AR platform with an integrated laser range finder.

By enabling simpler and faster techniques for these tasks we view a laser range finder as a promising tool for ubiquitous AR.

*e-mail: jwither@cs.ucsb.edu
†e-mail: ccoffi n@cs.ucsb.edu
‡e-mail: jventura@cs.ucsb.edu
§e-mail:holl@cs.ucsb.edu

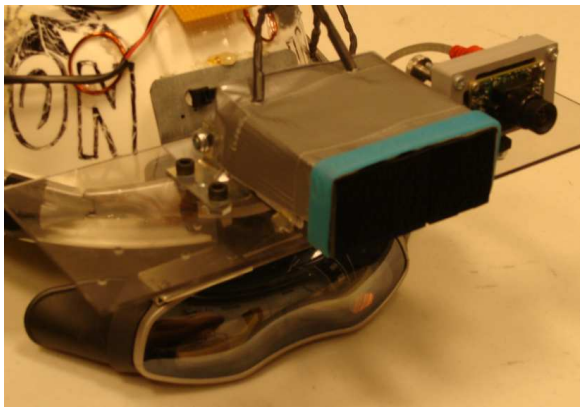

Figure 2: Our hardware on it's development platform. The laser range finder is rigidly mounted to a piece of Lexan, and the camera is carefully calibrated to point in the same direction. Here the development platform is mounted to our wearable system's helmet.

Figure 3: This figure shows two labels, "Edge Label" which is correctly oriented to the surface of the building it is annotating, and "Perpendicular Label" which is oriented perpendicular to the viewing angle from which it was created.

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

The first goal of our system is to aid in annotation. We distinguish four general ways to register a new annotation point from a distance in outdoor AR. First, we could rely on the user to estimate the distance to the object as in our previous work[18]. We found however that even with extra cues human depth estimation is not accurate enough for precise annotation placement. A second way of determining the depth to a point to be annotated is to use multiple views and triangulation. This can be done either by walking to several locations, as demonstrated by Piekarski and Thomas [13] and Baillot et al. [3], or by using an external source for a second viewing angle like aerial photographs as done in [17]. A third general approach is to rely on an a-priori model of the environment, in which case VR interaction techniques such as those developed by Bowman and many others will be applicable [4]. All of these approaches have drawbacks, whether it is inaccuracy, time requirements, or need for environmental knowledge. By using a laser range finder we employ a fast and accurate fourth way of creating an annotation at a distance, direct measurement.

A second goal of our system is a simple and effective way of segmenting out a foreground object and assigning it a constant depth (billboarding). The Tour into the Picture system presented by Horry et al. demonstrates the usefulness of billboarding for image-based modeling from a single viewpoint [11], although in their system object segmentation has to be produced manually. Hoiem et al. later developed an automatic object pop-up technique using machine learning [9]. Neither of these systems are able to build a to-scale depth map however, something we are able to do with a laser range finder. The remaining challenge in our system is to find the visible extent of an object in the image. For foreground object segmentation we use graph cut [5, 14].

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 [15], 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 [16], or a moving camera [12], 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 [6], or recover depth information from landmark features in overlapping photographs [1].

3 Hardware and Calibration

Our testing platform can be seen in Figure 2. 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 [7]. Envisor provides completely vision based orientation tracking, using both sparse optical flow for frame to frame features and heavyweight landmark features. 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 Annotation Creation

To create an annotation using the laser range finder users must use display motion (head motion for an HMD) to align the physical object they wish to annotate with a cursor on the screen that shows the pixel that is currently being ranged. Once the object and cursor are aligned a button press creates an annotation at the real world point by casting a ray from the user through the object and then placing the annotation at the reported distance along that ray.

The laser range finder also makes it easy for the user to orient an annotation with the surface they are placing it on. From the user's perspective all that needs to be done to correctly orient that annotation is to continue holding down a button, and then sweep the laser left or right across the surface of the object they are annotating. Because the range finder continuously updates this will give a number of points on the surface of that object. From this, we extract the orientation of the surface and correctly orient the annotation. An annotation that has been created with this process can be

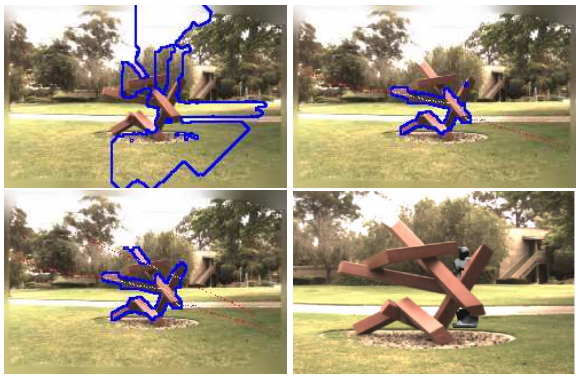

Figure 4: Result of diffusion and graph cut segmentation on a sculpture. Each range point is represented as a dot: red for background and white for foreground. The bottom right image shows the resulting segmentation being used for occlusion.

Figure 5: An aerial view of the group and planar information for the data set from Figure 1. 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 ortho-rectified. Groups not on building surfaces are due to entryways, trees, or lamp poles.

seen in Figure 3. One annotation is directly facing the user which looks correct from the user's current position, but would look incorrect from any other location. The second annotation is oriented to match the orientation of the wall that has been annotated.

To extract the orientation of a surface we first assume that the surface the user is annotating is vertical. This is a reasonable assumption for urban environments where the majority of annotations will be on buildings, and constrains the problem to a 2D search space when viewed from top down rather than 3D. We use a perpendicular regression to find the best fit line through 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 [8] 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.

5 Foreground Object Extraction

Another task that becomes much easier when using a laser range finder is popping out foreground objects to use for occlusion. The laser range finder gives us points on the object to start segmentation, and the correct depth to the object, so the entire object can be virtually billboarded. The user interaction for this task can be very simple. Pressing the input button while looking at an object starts the foreground object extraction mode. While holding down the button, the user looks at and around the object they wish to segment, generating depth samples $u_0, \ldots, u_n$ (until the button is released). When we first detect $T_u = |u_{i+1} - u_i| > \tau$ we assume that we have jumped between background and foreground, and set $u_F = \min(u_i, u_{i+1})$ as the foreground depth. All samples for which $|u_i - u_F| < T_u$ are labeled as foreground, and all other samples are labeled as background.

The next step is to propagate the depth sample labels to the rest of the image. For this, we use a form of diffusion similar to that described in DiVerdi et al.'s [7] work to expand information out to neighboring pixels. 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 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. We use edge information to regulate diffusion by subtracting the edge value at a neighboring pixel from the neighbor's confidence value before it is 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 an Nvidia GPU, 120 iterations per second can be achieved.

In some cases, the edge-stopping criterion of the diffusion process fails to separate background areas from foreground objects. Assuming that the foreground and background have distinct color distributions, we can use color-based segmentation to better separate them.

We apply a graph cut method for binary color segmentation [5, 14]. The output of the graph cut procedure is a binary labeling $A = (A_1, \ldots, A_n)$ of each pixel in the image as either belonging to the foreground F or the background B. Initially, we only know the correct label for the laser range finder samples around the object (based on depth). Using these pixels we build 3D color histograms H_F and H_B in RGB space, to capture the color distribution of each label, and then use this information to seed the graph cut procedure.

Figure 4 gives an example case where diffusion-based segmentation failed, but graph cut segmentation succeeded. After a single sweep across the object, the lower half of the object is segmented. By sweeping across the top of the object and running graph cut a second time, the entire object is segmented well, except for a very dark patch in the middle. In this way the user can refine the object segmentation by adding more depth samples where needed.

6 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

Figure 6: A virtual statue occluded by the real world.

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 5 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, as described in Section 5. 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 sweeping their view over the region again, 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 model the scene with vertical planar surfaces. This is a reasonable approach 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 we 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. To find planes we use the technique described in Section 4 to fit line segments (that represent the vertical planes from a top down view) to the points of each group. 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 our evolving depth map.

7 Conclusions and Future Work

We have presented annotation and modeling methodology using a single-point laser range finder in an outdoor AR setting. We have presented results that demonstrate how a laser range finder can improve the AR experience in several areas. First, we simplify the placement of new, possibly object aligned annotations. Building depth maps of foreground objects, or the entire scene surrounding the user also becomes possible with little work. The depth maps we create are accurate enough to be useful for a number of AR applications, arguably the most useful of which is enabling occlusion of virtual objects by real world objects, an application of which can be seen in Figure 6.

The primary focus of this paper has been on the semi-automatic creation of depth maps from a stationary location. In ongoing work, we are considering a number of application areas for our 3D panoramas. In addition to AR applications, we see great potential for using our 3D panorama technology in the incremental construction of large-scale environment models by multiple roaming users.

Acknowledgements

The authors wish to thank Stephen DiVerdi for his development of the Envisor system, that this work builds upon. This work was supported in part by NSF IGERT grant #DGE-0221713, a research contract with KIST through the Tangible Space Initiative Project, and NSF CAREER grant #IIS-0747520.

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] D. Bowman. *Interaction Techniques for Common Tasks in Immersive Virtual Environments*. PhD thesis, Georgia Tech, Atlanta, GA, 1999.

[5] Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 1:105–112 vol.1, 2001.

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

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

[8] M. Fischler and R. Bolles. Random sample consensus: a paradigm for model fi tting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[9] D. Hoiem, A. A. Efros, and M. Hebert. Automatic photo pop-up. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, pages 577–584, New York, NY, USA, 2005. ACM.

[10] 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.

[11] Y. Horry, K.-I. Anjyo, and K. Arai. Tour into the picture: using a spidery mesh interface to make animation from a single image. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 225–232, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[12] 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.

[13] 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.

[14] C. Rother, V. Kolmogorov, and A. Blake. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, 2004.

[15] 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.

[16] 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.

[17] J. Wither, S. DiVerdi, and T. Höllerer. Using aerial photographs for improved mobile ar annotation. In *Proceedings of the International Symposium on Mixed and Augmented Reality*, pages 159–162, 2006.

[18] J. Wither and T. Höllerer. Pictorial depth cues for outdoor augmented reality. In *Proceedings of the International Symposium on Wearable Computers*, pages 92–99, 2005.