FIRST EXPERIENCES WITH A MOBILE PLATFORM FOR FLEXIBLE 3D MODEL ACQUISITION IN INDOOR AND OUTDOOR ENVIRONMENTS – THE WÄGELE ¹ Peter Biber, Sven Fleck, Michael Wand, Dirk Staneker, Wolfgang Straßer

University of Tübingen Tübingen, Germany

{biber,fleck,wand,staneker,strasser}@gris.uni-tuebingen.de

KEY WORDS: 3D modeling, registration, omnidirectional vision, laser range scanner, 3DTV

ABSTRACT

Efficient and comfortable acquisition of large 3D scenes is an important topic for many current and future applications like cultural heritage, web applications and 3DTV and therefore it is a hot research topic. In this paper we present a new mobile 3D model acquisition platform. The platform uses 2D laser range scanners for both self localization by scan matching and geometry acquisition and a digital panorama camera. 3D models are acquired just by moving the platform around. Thereby, geometry is acquired continuously and color images are taken in regular intervals. After matching, the geometry is represented as unstructured point cloud which can then be rendered in several ways, for example using splatting with view dependent texturing. The work presented here is still "in progress", but we are able to present some first reconstruction results of indoor and outdoor scenes.

1 INTRODUCTION

Current research at University of Tübingen, WSI/GRIS is concerned with building a mobile platform for acquisition of 3D models. Both the platform and the processing software is work in progress, but there are already first promising results. In this paper we present our design principles and the parts of the mobile platform that are already implemented.

The mobile platform offers a very flexible sensor setup. Currently it is equipped with two 2D Sick laser range scanners (LMS 200) and a Canon EOS-20D digital camera with panoramic lens attachment (from 0-360.com). The laser range scanners record continuously while the platform is moved around. Panoramic images from the digital camera are taken in regular intervals. See Fig. 1 for an overview.

One of the laser scanners is mounted to record range values horizontally. This data is used to build a two dimensional map and to localize the mobile platform with respect to this map. Our techniques to tackle this problem are borrowed from robotics and in essence we have to solve the simultaneous localization and mapping (SLAM) problem. The other laser scanner is mounted perpendicularly, its data yields the geometric information for the 3D model. Thus, the environment is recorded in slices (opposed to 3D scanners) and the Wägele is operated in a fashion similar to former handheld image scanners. Section 3 describes the platform and the used sensors in more detail.

After a recording session the collected data is assembled to create a consistent 3D model in an offline processing step. First a 2D map of the scene is built and all scans of the localization scanner are matched to this map. After this step the position and orientation of the Wägele is known at each time step. This process is detailed in section 4.



Figure 1: One setup of our mobile platform. L1: Laser Scanner (scanning horizontally) used for 2D mapping and self-localization. L2: Laser Scanner (scanning vertically) for geometry acquisition. C1: Eight Megapixel panoramic camera for texture acquisition.

The panoramic camera has been calibrated, and as the relative positions of all sensors are known, geometry creation and texture mapping is easy with the known positions, see section 5. The result of this whole process is an unstructured point cloud with multiple color attributes per point. Flexibility of this modeling process is described in section 6. Section 7 discusses how the acquired data is used for rendering. Section 8 shows results for an indoor and an outdoor example scene. The paper concludes with a discussion of advantages and disadvantages of our approach and an outline of future work.

2 RELATED WORK

A brief summary of related work can be divided into purely geometric approaches, image-based approaches and hybrid

¹A Wägele – Swabian for a little cart

approaches for scene acquisition and representation.

2.1 Geometric approaches

Geometric representations of scenes include triangle meshes, curve representations or simply point clouds to model surfaces. Material properties, light sources and physical models provide the basis for rendering them. While it is possible to build mobile platforms that are able to acquire surface models of real world scenes by range scan techniques (Thrun et al., 2000, Hähnel et al., 2003, Surmann et al., 2003) even in real-time, estimation of material properties or light sources is a hard problem in general. So to render visual information convincingly without reconstructing or simulating physical properties it has been proposed to represent real scenes directly by images.

2.2 Image-based approaches

Image-based rendering is a now well established alternative to rendering methods based on geometric representations. The main promise is that it is able to generate photorealistic graphics and animations of scenes in realtime (McMillan and Bishop, 1995). Nowadays, panoramic views are the most well known variant of image-based rendering and can be discovered everywhere in the web. To allow all degrees of freedom, the so-called plenoptic function has to be sampled. Aliaga et al. (Aliaga et al., 2003) presented a system based on this approach that allows photorealistic walk-throughs in indoor environments. A panoramic camera mounted on a mobile platform captures a dense "sea of images", that is, the distance between two camera positions is only around 5 cm. Advanced compression and caching techniques allow walk-throughs at interactive speed. For the calculation of the camera positions, battery powered light bulbs were placed at approximately known positions. The largest area covered was $81m^2$, requiring around 10.000 images. The disadvantage of such a model is that despite its high memory requirements, only walk-throughs are possible: the user is not permitted to move too far away from a position where an image has been recorded. Especially the height of the viewpoint is fixed and cannot be chosen by the user.

It is now common to attempt to combine the best of both worlds in so-called *hybrid* approaches.

2.3 Hybrid approaches

Debevec et al. combined still photographs and geometric models in a hybrid approach (Debevec et al., 1996). In their work, the user had to interactively fit parameterized primitives such as boxes to the photographs to build a basic model. This model in turn was the basis of a modelbased stereo algorithm, which enriched the basic model with depth maps. Finally, *view-dependent texture mapping* was used to simulate geometric details not recovered by the model. This system allows generation of photo-realistic renderings from new viewpoints, as long as there exists a still photograph taken from a position close to that new viewpoint. El-Hakim et al. described a system that combines image-based modeling (also with human interaction) for surfaces like planes or cylinders, range-based modeling for fine details, aerial images and panoramas for far objects like mountains (El-Hakim et al., 2003). Their system was applied successfully to several heritage sites. Also Sequeira et al. presented a mesh based approach to automated construction of textured 3D models using a video camera and a laser scanner mounted on a pan-tilt unit (Sequeira et al., 2003). They employ an autonomous mobile platform using embedded software for triangulation, registration and integration.

Texture mapping per se, that is, mapping the color information of an image onto a plane, belongs to the oldest class of hybrid techniques, and is still the most commonly used method in computer graphics, so acquisition of textures from real world scenes is an important topic. A recent work in this field is due to Früh and Zakhor (Früh and Zakhor, 2003). They described a system that is able to generate 3D models of a city by combining textured facades with airborne views. Their model of downtown Berkeley, which is really worth a glance at, allows walk-throughs as well as bird's eye views.

3 DESCRIPTION OF PLATFORM

3.1 Platform itself

The Wägele platform consists of the following components: The main sensor acquisition platform is built using aluminium XC $44mm \times 44mm$ and $88mm \times 44mm$ structural beams. This combines great stiffness with limited weight and allows for a very flexible setup. So it can be easily adapted to various sensor and scene requirements. Together with a professional tripod it is mounted on the basic cart. The power supply consists of 24 NiMH cells with 1.2V/3000mAh each and some voltage regulators (5V, 6V, 24V).

3.2 Laser Scanner

Sick LMS-200 are eye-safe laser scanners widely used in robotics and industry. They feature a 180° field-of-view, max. 0.25° angular resolution and 75Hz sample rate. They transmit infrared laser beams and receive the distance to the reflection position via time-of-flight techniques offering an accuracy of 10mm, a statistical error of 5mm and a systematic error of 15-40mm. We use a RS422 connection via a multi-RS422 to USB bridge to the host, running synchronously at 500kbps each.

3.3 Digital camera

To achieve high quality images with a panoramic mirror a high spatial resolution is necessary. We utilize a Canon EOS-20D SLR camera with 8Mpixels. Its appealing properties include the very low noise CMOS sensor, the excellent optics and its great speed. It is accessible via USB2.0 by the Canon camera API. For our acquisitions, we use the manual mode, where both a constant aperture value and exposure time is set. The maximal possible aperture value of 32 is chosen for maximal depth of focus.

3.4 Panoramic mirror

On the camera a panoramic optics from "0-360.com" is mounted. It offers a large field of view: 115° , whereas 52.5° lie above the horizon and 62.5° lie below. It consists of an aluminium bar and a coated glass reflector. The camera with panoramic mirror is calibrated. Calibration is performed in Matlab using a calibration pattern.

4 BUILDING THE 2D MAP

An accurate 2D map is our basis to obtain accurate localization estimates. Our approach to build such a map belongs to a family of techniques where the environment is represented by a graph of spatial relations obtained by scan matching (Lu and Milios, 1997, Gutmann and Konolige, n.d., Frese and Duckett, 2003). The nodes of the graph represent the poses where the laser scans were recorded. The edges represent pairwise registrations of two scans. Such a registration is calculated by a scan matching algorithm. The scan matcher calculates a relative pose estimate where the scan match score is maximal, along with a quadratic function approximating this score around the optimal pose. The quadratic approximations are used to build an error function over the graph, which is optimized over all poses simultaneously (i.e., we have $3 \times nrScans$ free parameters). Details of our method can be found in (Biber and Straßer, 2003). Fig. 2 shows one map's graph along with the corresponding horizontal range scans, Fig. 3 shows a 2D map of our institute's floor obtained this way.



Figure 2: A 2D map of an indoor environment created by laser scan matching. White filled circles mark the position of nodes in the graph of spatial relations, whereas the white lines show edges in this graph (i.e. the corresponding scans have been registered successfully by scan matching.

4.1 Probabilistic Scan Matching

Matching two scans (i.e. finding the translation and the rotation between the positions where the scans were recorded) delivers the basic information for map building. Our approach to this problem is a probabilistic one. In summary, one scan is used to estimate a probability density function (in 2D space) that is thought of having generated the scan. This density function is used as a generative model and (using a maximum likelihood formulation) an energy function is derived from this generative model. This energy function describes how probable another scan is, dependent on



Figure 3: A 2D map of the whole floor of our institute.

translation and rotation. This energy function is optimized quickly using Newton's algorithm. Gradient and Hessian are thereby calculated analytically. More details can be found in (Biber and Straßer, 2003) and (Biber et al., 2004).

4.2 Global optimization of poses

The Hessian of the energy function after convergence of Newton's algorithm is used to build a spring like model of the link between the corresponding poses. Then an energy function is built over the whole graph of relations between scans using the energy function between links in a canonical way, which is then minimized. It would take too long to optimize over all poses simultaneously (the data set of Fig. 2 consist of around 30.000 scans and this corresponds to 90.000 parameters). Instead a new node is inserted into the graph only if the platform has moved a certain minimum distance or performed a minimum rotation. After global optimization the map has a correct topology and measurement errors have been distributed equally among the poses of the selected scans.

4.3 Localization

After the map has been built it is used to localize all scans. For this each scan is matched in turn to the map. The movement at each time step is predicted from the last time step and this prediction is used as a prior for the scan matching. Uncertainty in scan matching does not propagate without bounds, as position estimates are obtained by matching a current laser scan to the global map and not to the previous scan. So uncertainties in localization estimates are always with respect to the global map and thus to the global coordinate frame.

5 3D MODEL GENERATION

5.1 Point cloud generation

After localization the position of the platform is known for each laser range scan. The relative position and rotation between the horizontal and the vertical laser scanner are also known and so it is straightforward to convert a range reading from the vertical scanner into a 3D point in the global coordinate frame. This global coordinate frame is fixed at the starting position of the platform.

5.2 Color values generation

If the panoramic camera is calibrated and the relative position between camera and scanners is known, it is also straightforward to map the point cloud onto a single image. Unfortunately it turned out that it was very difficult to fix the focus manually so that the image of the mirror was sharp, as the combination of camera and attachment is a little bit shaky. So the camera was operated in auto focus mode which causes the image of the mirror to be on different positions in the camera image each time. In a postprocessing step we correct for this deviation by finding the black ring around the mirror (manually) in each image and then warping the image into a normalized coordinate system .

Subsequently, a z-buffer is constructed for each panoramic image: 3D-points that are occluded according to the z-buffer are not colored by this panoramic image. After this step each pixel receives color values from potentially multiple panoramic images. Final color values could be obtained by blending techniques. A more advanced method is explained in section 7. Alternatively, colored point clouds can be created from just a single panoramic image. Thereby missing depth values are interpolated from neighboring values. The point clouds of Fig. 7 were created using this method.

6 FLEXIBILITY OF THE PLATFORM

An important goal of our project was to allow 3D scanning in various scenarios. For that, the mobile platform is designed for maximal flexibility. The setup can be changed quickly. For transportation, only three screws have to be removed. Also more sensors can be mounted easily.

Up to now, two different setups have been evaluated, see Fig. 4 and 5. With setup #1 the second scanner scans to the left and to the right. But the floor is not scanned and therefore missing in the final model. This setup may however be useful in indoor environments, where the floor is known and colored uniformly so that it could be completed easily.

The second setup only scans to one side including to the ground. The other examples in this paper were acquired in this mode. The disadvantage of this mode is of course that the operator has to make his way twice in both directions if it is necessary to scan to both sides. But in many cases



Figure 5: Setup #2

this is not necessary. For example, if a facade of a building should be scanned it is sufficient to scan to one side while circling the building. Hence, various applications can be addressed by this platform. Both indoor and outdoor environments can be scanned. Orthogonal to this, two traditionally separately handled application classes can be solved with no special adjustment, namely the *inside out* view and the *outside in* view.

7 RENDERING

Rendering is performed using the point clouds obtained from the "vertical" laser scanner. For treating the radiance (texture) values, different strategies have been implemented: First, a fixed color value can be assigned to each sample point, effectively assuming ideal diffuse surfaces. In this case, we always chose the nearest camera to obtain color values of optimal resolution. However, the diffuse emission assumption breaks down for glossy surface, such as car paint. To deal with such cases, we have implemented a simple view-dependent texturing scheme in the spirit of (Debevec et al., 1996): Each sample point is assigned a set of color values from all panoramic images taken from camera positions that are not occluded from the point of view of the current point. During rendering, each color value is weighted by in how far the current viewing direction matches that of the camera (in our current implementation, we use the cosine of the angle between viewing and camera angle, taken to a fixed power). The color contributions from different cameras are added together and renormalized by dividing by the weights' sum. All calculations are performed in the vertex shader of the graphics board, leading to real-time performance for scenes of moderate complexity (say one million points and 10 camera positions). Fig. 6 shows results and a visualization of



Figure 6: View dependent rendering – the top row shows results from our view dependent rendering algorithm, the second row shows a color coding of the blending weights used for the different cameras.

the camera weighting functions.

7.1 Indoor Run

Various runs were performed in our department. The map of Fig. 3 was recorded there. A part of a point cloud created in the floor is shown in Fig. 8. Renderings of point clouds obtained from single panorama images with interpolated depth values are shown in Fig. 7.

8 RESULTS



Figure 8: Point cloud recorded in an indoor environment (floor of our department).

8.1 Outdoor Run

Although the flat floor assumption required by the 2D map approach may be violated in outdoor environments, good results can still be achieved. See Fig. 11 for a 2D map acquired outside our institute. In this environment we also recorded an outdoor scene with a car. Thus, it shall demonstrate both the inside out view and the outside in view around the car. One of the source images is shown in Fig. 9. and unwarped in Fig. 10.

Two screenshots of a textured outdoor scene are shown in Fig. 12.



Figure 9: A raw panorama image.

9 CONCLUSION AND FUTURE WORK

We presented a mobile and flexible platform for acquisition of 3D models. Both indoor and outdoor scenes have been recorded and rendered. Our first results are quite promising and we identified some advantages of our approach to 3D scanners:

- No extra registration step of 3D scans is necessary as the scan matching is an integral part of the acquisition process.
- By manually moving the platform the operator can easily control scene sampling.
- We have a more uniform scene coverage as we have thousands of viewpoints compared to only a few as it is typical when building models using a 3D scanner.



Figure 7: Some screenshots. Renderings of point clouds obtained from single panorama images. Some point clouds are partly post-processed to remove the ceiling and outliers.

Figure 10: An unwarped panorama image.

Figure 12: Two screenshots of a recorded outdoor model.

Figure 11: Map generated for an outdoor run.

• Additionally, this approach comes much more inexpensive compared to a 3D laser scanner.

There are however also inherent problems in comparison: if the cart wobbles while driving around the flat ground assumption is violated causing a decreased point cloud quality. We are, however, confident to cope with this problem and other issues by adding more sensors as detailed below.

More Sensors

• **GPS:** For large scale outdoor scenes. Loop closing is possible using only scan matching (in case of need supported also by a manual mode), but it is much eas-

ier (and will also be faster to calculate) if global position estimates are available.

- Additional video cameras: A video camera will be mounted on both laser range scanners. Using optical flow techniques we will try to estimate the platform's rotational deviation from upright position.
- **Inclinometer:** We will also employ a 2D inclinometer with microcontroller based temperature drift compensation and digital output for the same purpose.
- **Third laser scanner:** A third laser scanner enables viewing in both directions at the same time. Due to the modular approach, this can be easily mounted.

Other directions for future work include:

High Dynamic Range Imaging

Outdoor scenes often present a wide dynamic range of illumination intensities, from shadows to direct view into the sun. The same is true for the inside of buildings if they have windows. One can imagine that – in such situations – there are often underexposed and overexposed parts of the panoramic image. We will acquire automatically images with varying aperture and build HDR images

Incorporation of stereo algorithms

Stereo vision will lead to even higher quality results due to the higher spatial resolution of the panoramic images. Graph cut methods based on omnidirectional images lead to quite convincing results as described in (Fleck et al., 2005).

Splatting

We plan to implement two options for display: First, the points can be visualized directly by screen aligned rectangular splats of fixed size. However, this leads to visible holes in the reconstructed surface, depending on viewing distance and sampling density. To alleviate this problem, we calculate tangential quads (Pfister et al., 2000): For each point, the local point distribution (the k-nearest neighbors, with k = 10..20) is approximated by a Gaussian distribution, i.e. a covariance matrix is estimated. A principal component analysis of the covariance matrix yields a least-squares-optimal estimate of the tangent plane. The eigenvector of the smallest eigenvalue corresponds to the normal, the other two principal axis of the Gaussian serve as estimate of the local splat size. The resulting quadrilaterals are rendered as GL_QUADS. To determine the knearest neighbors of all points efficiently, a spatial hierarchy (an octree) is used to speed up the corresponding range query. Currently, our implementation of this technique still suffers from a large memory overhead so that tangential quads can only be computed for smaller (downsampled point clouds). Thus, the high resolution scans in this paper are still displayed using fixed sized screen aligned splats.

10 ACKNOWLEDGMENTS

We would like to thank Gerd Wolf for his implementation work on sensor communication. Thanks go also to Alexander Berner, Martin Bokeloh, Mark Hoffmann and Benjamin Maier for additional rendering implementations. This work is supported by EC within FP6 under Grant 511568 with the acronym 3DTV.

REFERENCES

Aliaga, D., Yanovsky, D. and Carlbom, I., 2003. Sea of images: A dense sampling approach for rendering large indoor environments. Computer Graphics & Applications, Special Issue on 3D Reconstruction and Visualization pp. 22–30.

Biber, P. and Straßer, W., 2003. The normal distributions transform: A new approach to laser scan matching. In: International Conference on Intelligent Robots and Systems (IROS).

Biber, P., Fleck, S. and Straßer, W., 2004. A probabilistic framework for robust and accurate matching of point clouds. In: 26th Pattern Recognition Symposium (DAGM 04).

Debevec, P. E., Taylor, C. J. and Malik, J., 1996. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. SIGGRAPH 96. El-Hakim, S. F., Beraldin, J.-A., Picard, M. and Vettore, A., 2003. Effective 3d modeling of heritage sites. In: 4th International Conference on 3D Imaging and Modeling (3DIM'03), Banff, Alberta, Canada, pp. 302–309.

Fleck, S., Busch, F., Biber, P., Andreasson, H. and Straßer, W., 2005. Omnidirectional 3d modeling on a mobile robot using graph cuts. In: IEEE International Conference on Robotics and Automation (ICRA 2005).

Frese, U. and Duckett, T., 2003. A multigrid approach for accelerating relaxation-based SLAM. In: Proc. IJ-CAI Workshop on Reasoning with Uncertainty in Robotics (RUR 2003).

Früh, C. and Zakhor, A., 2003. Constructing 3d city models by merging ground-based and airborne views. Computer Graphics and Applications.

Gutmann, J.-S. and Konolige, K., n.d. Incremental mapping of large cyclic environments. In: Computational Intelligence in Robotics and Automation, 1999.

Hähnel, D., Burgard, W. and Thrun, S., 2003. Learning compact 3d models of indoor and outdoor environments with a mobile robot. Robotics and Autonomous Systems.

Lu, F. and Milios, E., 1997. Globally consistent range scan alignment for environment mapping. Autonomous Robots 4, pp. 333–349.

McMillan, L. and Bishop, G., 1995. Plenoptic modeling: An image-based rendering system. SIGGRAPH.

Pfister, H., Zwicker, M., van Baar, J. and Gross, M., 2000. Surfels: Surface elements as rendering primitives. In: K. Akeley (ed.), Siggraph 2000, Computer Graphics Proceedings, ACM Press / ACM SIGGRAPH / Addison Wesley Longman, pp. 335–342.

Sequeira, V., K. Ng, E. W., Goncalves, J. and Hogg, D., 2003. Automated reconstruction of 3d models from real environments. ISPRS Journal for Photogrammetry and Remote Sensing 54, pp. 1–22.

Surmann, H., Nüchter, A. and Hertzberg, J., 2003. An autonomous mobile robot with a 3d laser range finder for 3d exploration and digitalization of indoor environments. Robotics and Autonomous Systems.

Thrun, S., Burgard, W. and D.Fox, 2000. A real-time algorithm for mobile robot mapping with applications to multirobot and 3d mapping. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA).