Joint Calibration of a Range and Visual Sensor for the Acquisition of RGBZ Concentric Mosaics

Ingo Bauermann and Eckehard Steinbach

Media Technology Group, Institute of Communication Networks, Technische Universität München,
Arcisstr. 21, 80290 Munich, Germany
Email: {ingo.bauermann,eckehard.steinbach}@tum.de

Abstract

In this paper we present an algorithm for joint extrinsic calibration of a moving sensor device consisting of a low cost industrial 2D laser range finder and a single high quality consumer DV video camera mounted on a camera crane. The calibration is based on the minimization of the Euclidean projection error of scene points in many frames captured at different viewpoints. The calibration procedure is designed to support any arbitrary motion trajectory of the acquisition device. In this work the capture geometry for concentric mosaics, a well known image-based rendering technique, is used to evaluate the performance of the proposed algorithm. The projection error after calibration is less than one pixel on average.

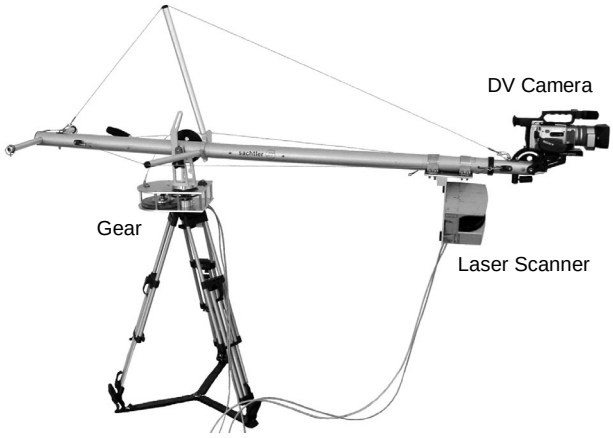

Figure 1: Device for joint image and depth data acquisition. A camera crane with gear moves the Sony DX2000 consumer camera and the SICK LMS290 laser range finder.

1 Introduction

Industrial laser measurement systems like those from the SICK-LMS family [7] have manifold applications in computer vision and geometric modeling and have become very popular amongst researchers in many fields recently. Three main fields of research combine range data from such devices and intensity information from video cameras and therefore require intrinsic and extrinsic calibration of both.

For mobile robot navigation tasks cheap depth sensors provide reliable information in real time for motion planning and obstacle avoidance. Laser scanners used in these applications provide approximately one depth sample per degree and 5cm of range resolution while scanning one line at a time. Additionally, visual sensors are often used to gain further information about the environment using computer vision techniques [10]. Due to the kind of task these systems have to solve, depth values are sampled sparsely and visual sensors often only provide low quality images.

Beside mobile robot vision which generally focuses on task specific real time geometric modeling, a second class of applications for laser range finders in computer graphics and vision is pure geometric modeling with texture mapping for visualization purposes (e.g. [2], [8], [20]). Expensive laser measurement systems used in these applications like the Riegl LMS Z360 [9] or Cyrax 2500 [19] usually provide dense range maps while simultaneously providing high quality texture data. The spatial resolution of these laser scanners can be as fine as 0.01 degree and 1mm in depth.

Along with quality assurance in industrial applications there is another class of applications for laser range finders. In image-based rendering a high quality reconstruction of the plenoptic function [13] is required. In this scenario geometric scene models are used for rendering and compression purposes and have proved to give a significant

gain (e.g. [14],[17]). However, though multiple viewpoint imaging is the backbone of image-based scene representations, depth information and image data registration is mostly performed at sparsely spaced viewpoints leading to high quality but at most view dependent textured geometric models.

In this work a densely sampled plenoptic function along with dense geometry information for rendering and compression purposes is considered. Here, the registration of many high quality images along with sparse 2D laser range data both captured from different viewpoints is discussed. While intrinsic calibration of visual and range sensor devices is well understood [15], work on fusion of range and intensity data is still ongoing and requires accurate intrinsic and extrinsic calibration of the used sensors. Making use of standard sensors and equipment we propose a joint calibration algorithm and evaluate its properties for image-based rendering purposes using concentric mosaics.

The remainder of this paper is structured as follows. Chapter 2 gives some notes on related work. In Chapter 3 we briefly describe the system setup and the resulting capture geometry. In Section 4 the notations we use are given. Chapter 5 and 6 discuss the intrinsic and extrinsic calibration procedure followed by a presentation of the results in Chapter 7. Some notes on the learned lessons are given in Chapter 8. Chapter 9 concludes the paper and gives notes on further work.

2 Related work

Significant work has been done on multiple viewpoint reconstruction and extrinsic calibration of single or multi camera systems (see e.g. [15], [16]). Registering multiple range images has also been studied extensively (e.g. [18]). Some work has been done on the joint calibration of heterogeneous systems assembled of a 2D laser scanner and a video camera [5], [11], [12]. In [4] joint extrinsic calibration of a depth sensor and a video camera is addressed. While we have the same goal, our approach differs in one main aspect. In [4] the acquired geometry is only visible in a very small portion of only one of the captured images. In comparison, our algorithm takes multiple viewpoint geometry into account.

3 System setup and capture geometry

A low cost acquisition device assembled of a SICK LMS290 2D laser range finder and a consumer DV camera SONY VX2000 is mounted on a camera crane as shown in Figure 1. A gear provides constant velocity for a circular acquisition path with a radius of about 1.5 meters. The rotation speed of the crane is 1/12 rotation per minute. Both sensors are outward-looking. The scanner is mounted nearer to the center of rotation and therefore its scan line is slightly slanted in order not to let the camera block the laser beam. The video camera provides 4 high quality DV images (720x576) per second with a field of view of approximately 40 degrees. The scanner provides 50 range samples per scan along one scanline and approximately performs 80 scans per second covering 50 degrees of vertical field of view.

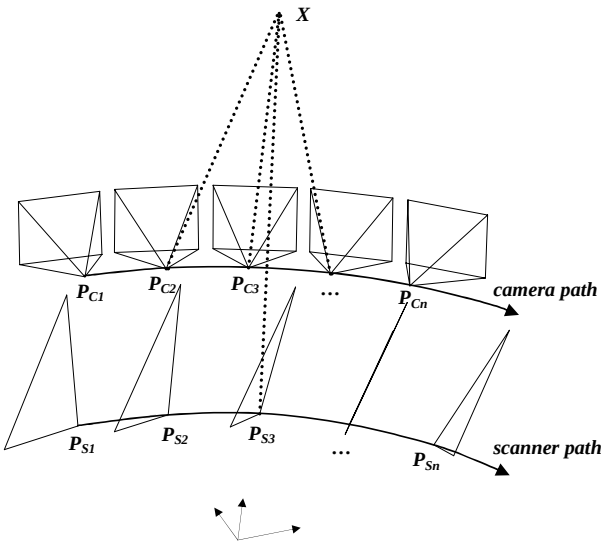

Figure 2: Capture geometry for joint acquisition of intensity and range data (seen from a position above and behind the center of rotation). Example viewing frustums of the video camera on the moving trajectory along with some 2D scan lines of the laser scanner are shown. The dotted lines denote corresponding projections of one scene point.

Figure 2 shows the capture geometry. A scene point $\boldsymbol{X}$ has multiple correspondences in the images taken along the camera path and shows up exactly once in the range data captured along the scanner path. Camera and scan views are described by their projection matrices $\boldsymbol{P}_{Ci}$ and $\boldsymbol{P}_{Sj}$, where

the index i and j refer to a captured image and depth scan at specific time instances t_i and t_j , respectively.

4 Notations

The acquisition device consisting of the two sensors is assumed to move along a trajectory described by a position $\boldsymbol{T}_{\mathrm{D}}(t)$ and rotation $\boldsymbol{R}_{\mathrm{D}}(t)$ w.r.t. world coordinates.
Our video camera is described by the common pinhole model. A world point $\boldsymbol{X} = [X, Y, Z, 1]^{\mathrm{T}}$ is projected to image coordinates $\boldsymbol{p} = [u, v]^{\mathrm{T}}$ as follows [21]:

$$\boldsymbol{p} \sim \boldsymbol{P}_{Ci} \boldsymbol{X}$$

Here $\boldsymbol{P}_{Ci}$ is the projection matrix for view i captured at time t_i:

$$\boldsymbol{P}_{Ci} = \boldsymbol{K}\ [\boldsymbol{I}_{3x3}|\boldsymbol{0}_3]\ \boldsymbol{M}_{\mathrm{CD}}\ \boldsymbol{M}_{\mathrm{D}}(t_i)$$

where $\mathbf{K}$ denotes the intrinsic calibration matrix. $\mathbf{M}_{\mathrm{CD}}$ is the time invariant relative extrinsic calibration matrix and is given as:

$$\boldsymbol{M}_{\mathrm{CD}} = \begin{bmatrix} \boldsymbol{R}_{\mathrm{CD}}{}^{T} & -\boldsymbol{R}_{\mathrm{CD}}{}^{T}\boldsymbol{T}_{\mathrm{CD}} \\ 0_3^T & 1 \end{bmatrix}$$

where $\boldsymbol{R}_{\mathrm{CD}}$ and $\boldsymbol{T}_{\mathrm{CD}}$ represent the rotation and translation of the camera relative to the moving acquisition device. $\mathbf{M}_{\mathrm{D}}(t_i)$ denotes the time variant and global part of the extrinsic calibration matrix at capture time t_i :

$$\boldsymbol{M}_{\mathrm{D}}(t_i) = \begin{bmatrix} \boldsymbol{R}_{\mathrm{D}}(t_i)^{T} & -\boldsymbol{R}_{\mathrm{D}}(t_i)^{T}\boldsymbol{T}_{\mathrm{D}}(t_i) \\ 0_3^T & 1 \end{bmatrix}$$

Without loss of generality, we assume the position and rotation of the laser scanner being identical with the acquisition device. Then the laser scanner we use can be described as follows. A data tuple $\boldsymbol{s} = (a, d)$ with a being the sample's index within one scan and the measured depth d is mapped to a world point $\boldsymbol{X} = [X, Y, Z, 1]^{\mathrm{T}}$ as follows:

$$\boldsymbol{X} = \boldsymbol{P}_{Sj}\, \boldsymbol{f}(\boldsymbol{s})$$

with $\boldsymbol{P}_{\mathrm{S}j}$ as the inverse projection matrix for scan j at capture time t_j:

$$\boldsymbol{P}_{Sj} = \boldsymbol{M}_{D-}(t_j) = \begin{bmatrix} \boldsymbol{R}_{\mathrm{D}}(t_j) & \boldsymbol{T}_{\mathrm{D}}(t_j) \\ 0_3^T & 1 \end{bmatrix}$$

$\boldsymbol{f}$ is a nonlinear mapping from $\boldsymbol{s}$ to scanner coordinates:

$$\boldsymbol{X}_{Scanner} = \boldsymbol{f}(\boldsymbol{s}) = \begin{bmatrix} 0 \\ d \cdot \sin(\mu \cdot a) \\ d \cdot \cos(\mu \cdot a) \\ 1 \end{bmatrix}$$

with μ as an intrinsic calibration parameter for the field of view of the scanner.

5 Intrinsic calibration

As intrinsic calibration of the sensors can be separated from the extrinsic calibration procedure, the intrinsic calibration matrix $\boldsymbol{K}$ of the camera is determined using the calibration tool available at [15]. We assume that the images have been warped to eliminate radial and tangent distortions of real lenses. The scanner's internal parameter μ is calibrated using a 3D calibration pattern.

6 Extrinsic calibration

For the extrinsic calibration of our sensor device we choose the capture geometry for concentric mosaics. Mounted on a rotating camera crane the motion trajectory is restricted to a circle. We assume the center of rotation being identical with the origin of the world coordinate system and the rotation plane being the xy-plane. In this case the motion of the acquisition device represented by the translation $\boldsymbol{T}_{\mathrm{D}}(t)$ and rotation $\boldsymbol{R}_{\mathrm{D}}(t)$ has 5 degrees of freedom which are included in the calibration process. These parameters are the radius r of the camera path and the rotation speed ω of the camera crane as well as 3 parameters describing a local rotation. The relative position $\boldsymbol{T}_{\mathrm{CD}}$ and rotation $\boldsymbol{R}_{\mathrm{CD}}$ of the camera to the laser scanner give another 6 degrees of freedom. The whole mapping can be noted as follows:

$$\boldsymbol{p}_{ij} \sim \boldsymbol{K}\ [\boldsymbol{I}_{3x3}|\boldsymbol{0}_3]\ \underbrace{\boldsymbol{M}_{\mathrm{CD}}}_{\text{6DOF}}\ \underbrace{\boldsymbol{M}_{\mathrm{D}}(t_i)\boldsymbol{M}_{D-}(t_j)}_{\text{5DOF}}\boldsymbol{f}(\boldsymbol{s}) \qquad (1)$$

The objective of our joint calibration algorithm is to minimize the Euclidean distance between a scene point measured by the laser range finder $\boldsymbol{X}_{\mathrm{j}}$ and mapped to image coordinates $\boldsymbol{p}_{ij}$ in image i and the corresponding observation $\tilde{\boldsymbol{p}}_{ij}$. Rather than minimizing a geometric error this relates best to an application in image-based rendering as ghosting artifacts are minimized whenever views are interpolated, even if the geometry is not accurate.

Given the acquisition times t_i and t_j for the images and range scans together with correspondences $\boldsymbol{s}_j$ and $\tilde{\boldsymbol{p}}_{ij}$ between range samples and multiple positions in image coordinates, the objective function to minimize becomes:

$$\sum_i \sum_j \left\| \boldsymbol{p}_{ij} - \tilde{\boldsymbol{p}}_{ij} \right\|^2 \qquad (2)$$

In our case, and whenever the motion is uniform $\boldsymbol{M}_{\mathrm{D}}(t_i)$ and $\boldsymbol{M}_{\mathrm{D}\text{-}}(t_j)$ can be combined into one matrix $\boldsymbol{M}_{\mathrm{D}}(\Delta t) = \boldsymbol{M}_{\mathrm{D}}(t_i - t_j)$ describing the change in position and orientation during the time period Δt. This does not decrease the number of parameters to calibrate but may help to find a closed solution for initialization of the global optimization in future work. [4] could be used to determine $\boldsymbol{M}_{\mathrm{CD}}$ beforehand in order to reduce the degrees of freedom for optimization. But, though K is assumed to be known, factorization of (1) is still difficult. Therefore we choose to initialize a global nonlinear optimization by rough measurements of the calibration parameters in the real world rather than solving a linear calibration.

(2) is optimized using the Levenberg-Marquardt method [1]. A simple outlier removal strategy is implemented by deactivating 20% of the worst point correspondences and rerunning the algorithm.

7 Results

To evaluate the performance of the joint extrinsic calibration procedure we captured about 800 Mbytes of raw data containing 3200 laser scans and 640 images covering a rotation angle of the camera crane of about 90 degrees. 20 points in the depth data were chosen and about 10 correspondences in different images were established for each of these points manually. The algorithm converged after 500 iterations using the Levenberg-Marquardt method. The final RMS projection error is 0.94 pixel. Figure 3 shows the distribution of the projection error as difference between the projected scene point and manually determined correspondences in image coordinates. The RMS of the corresponding geometric error which was not minimized is 6.2 cm.

Figure 5 shows one original image captured with our video camera. The raw data captured with the laser scanner is also shown where the horizontal axis denotes the acquisition time and the vertical axis denotes the sample index corresponding to $\boldsymbol{s}$. Bright regions are near to the scanner.

Figure 4 shows a warped view onto the point cloud captured with the laser scanner after calibration. Note the white areas where no depth values could be measured because the laser was blocked from obstacles in the scene. In the lower part of the figure the axes of the global coordinate system are shown along with some frustums of the camera and laser scanner respectively.

Figure 6 shows an orthographically projected view of the calibrated point cloud from directly above. A comparison of the flatness of the wall with the shown reference line proves that the model fits the real world quite well.

A triangulation was performed on the geometry data and warped to the captured images. Figure 7 shows the result. For white areas no depth value could be obtained. Two novel views from extreme viewpoints generated by texture mapping and warping are shown in Figure 8.

Finally, Figure 9 shows an image predicted from two views captured about 12 cm to the left and right, respectively, covering a range of 61 captured frames. This shows how the proposed acquisition and calibration method is to be used for compression of image based scene representations. Similar to [14] disparity compensated prediction based on the reconstructed geometry will be used to improve the compression of concentric mosaics. Note that almost no ghosting is visible although both views used for interpolation are weighted equally.

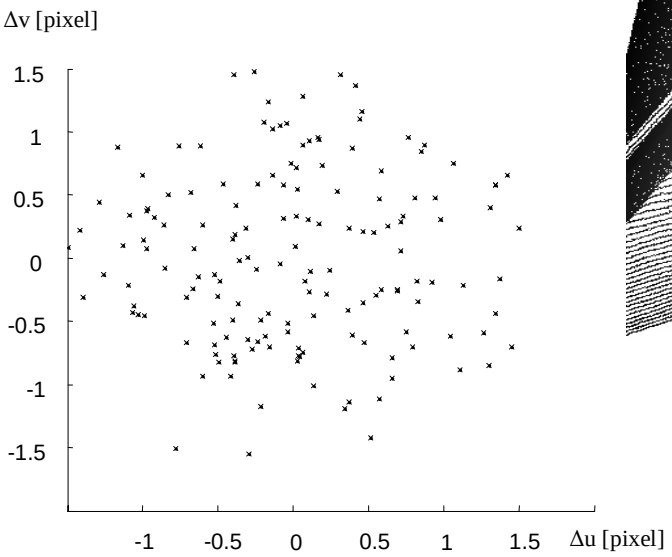

Figure 3: Distribution of the projection error after calibration for 20 scene points with approximately 10 correspondences in different images each.

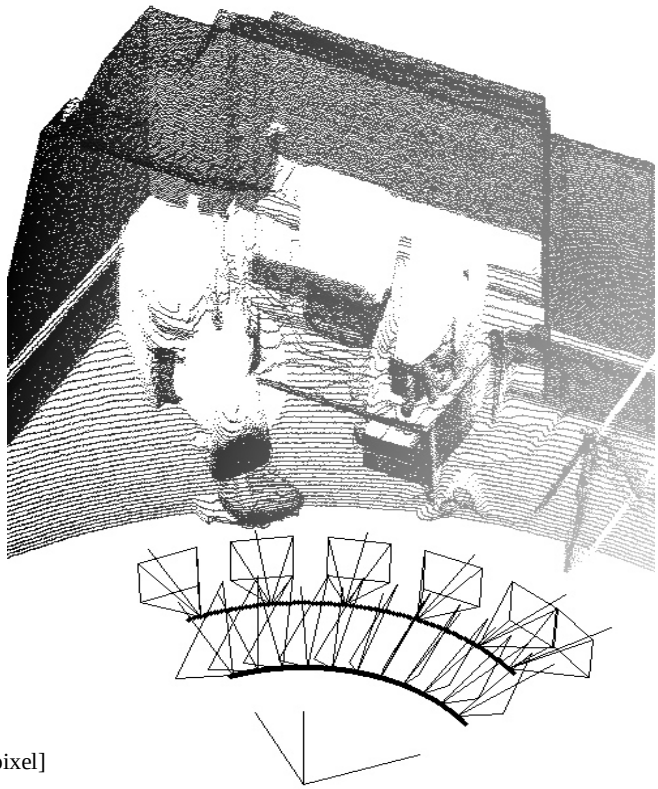

Figure 4: View onto the calibrated point cloud with example capture geometry.

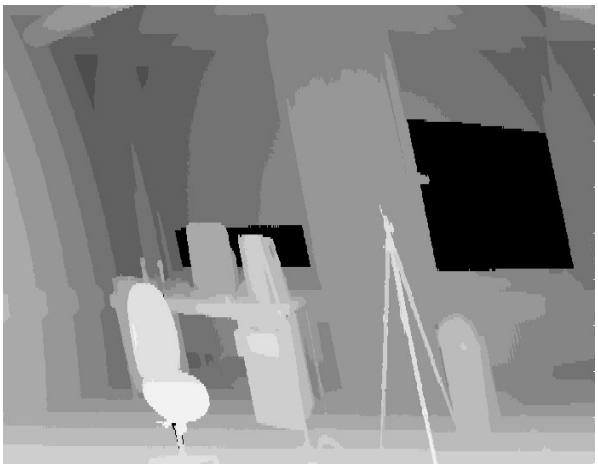

Figure 5: Examples of the acquired data. Left: One of the 640 captured images of the scene. Right: The captured depth panorama. Note the slant of the objects in the right image due to the local rotation of the laser scanner.

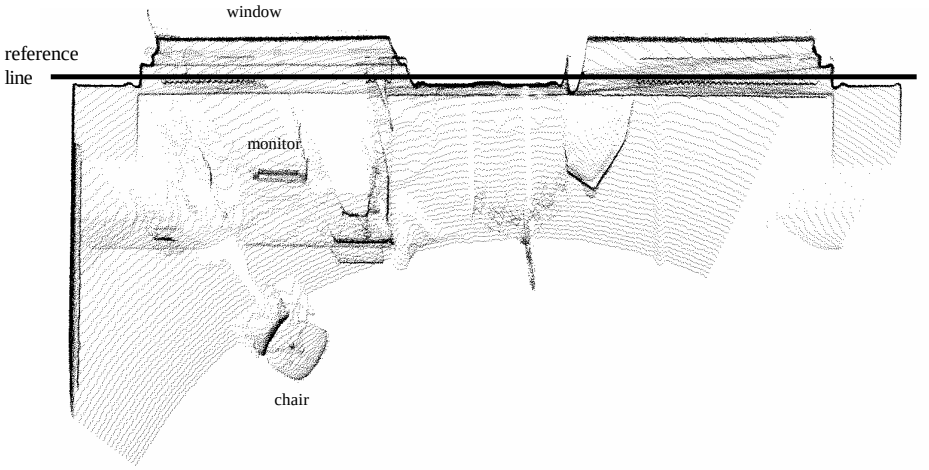

Figure 6: Calibrated point cloud viewed directly from above. Note the flatness of the wall for the whole capture range.

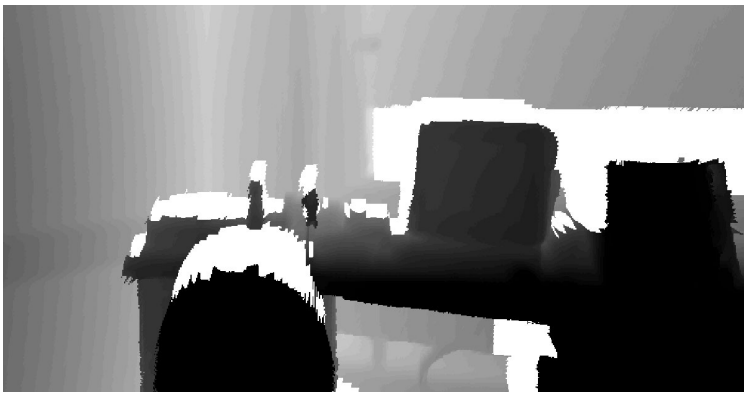

Figure 7: Triangulated geometry warped into a captured image (Part of Figure 5 - left). Darker areas are nearer to the camera.

Figure 8: Two novel views onto the scene at extreme viewpoints. Jagged edges are due to the low resolution of the range data compared to the spatial resolution of the captured images.

Figure 9: Detail of a view predicted from two captured views far apart. Left: View interpolated using the acquired geometry. Right: Original image part.

8 Lessons learned

Using a consumer DV camera inserts an unforeseen problem to the calibration process. Though the image quality is very good these cameras do not provide proper time stamping. Acquisition time capture has to be performed with a certain variable delay which then leads to a time shift finally showing up in the global extrinsic calibration as shift of the motion trajectories of the camera and the scanner. In this work we assume this delay to be compensated for perfectly by measuring it beforehand.
The spatial resolution of the laser scanner is approximately a twentieth of the resolution of the captured images (See Figure 10). Especially during the selection of corresponding points this leads to mismatches. This is the main reason for having that many outliers. The second reason for choosing to omit as much as 20% of the selected correspondences are vibrations due to the gear and the rotating mirror of the scanner. A third reason is that the laser beam expands extremely when shooting far into the scene. At edges this produces a heavy growing of near objects especially with objects having a reflective surface. To solve this crucial problem a 3D calibration pattern which allows for an accurate selection of scene points would be very useful. This pattern could consist of shots mounted on some device with several branches. The centers of the shots could be localized and interpolated very accurately in both the range and intensity data. This would allow for sub pixel accuracy also.

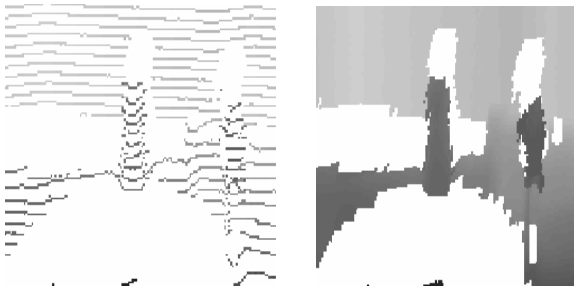

Figure 10: Left: Sparse depth samples warped to an image. Right: Interpolated depth image. The resolution of the laser scanner is about a twentieth of the resolution of the captured images. Reliable selection of matches is difficult.

9 Conclusions

In this paper, we present an acquisition device consisting of a low cost laser range finder and a standard consumer DV camera for image-based rendering. The proposed joint extrinsic calibration method takes acquired range and intensity data and uses scene points visible in only one laser scan but in many images to determine the relative position of the sensors as well as parameters of the motion and orientation of the devices. The algorithm is designed to support any trajectory. Results are shown for the acqui-

sition of concentric mosaics which prove the validity of the model. A RMSE of 0.95 pixel is achieved for the projection of a scene point into many images.

The device and the acquired data will be used for compression and rendering purposes in an image-based rendering system. Future work will also include a linear model for initialization of the global optimization procedure, range data fusion with geometry retrieved from the intensity data, and self calibration using reflectance data and feature extraction as well as super resolution of the depth data.

References

[1] D. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *SIAM J. Appl. Math. Vol. 11*, pp 431-441, 1963.

[2] http://www.cs.unc.edu/~ibr/projects/RTRENV/

[3] H. Xu, N. Gossett, and B. Chen, "PointWorks: Abstraction and Rendering of Sparsely Scanned Outdoor Environments," *Proceedings of the 2004 Eurographics Symposium on Rendering (EGSR'04)*, Norrköping, Sweden, Jun 21-23, 2004.

[4] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder," *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2004.

[5] Q. Zhang and R. Pless, "Fusing video and sparse depth data in structure from motion," *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2004.

[6] H. Baltzakis, A. Argyros, and P. Trahanias, "Fusion of laser and visual data for robot motion planning and collision avoidance," *Machine Vision and Application*, 12:431–441, 2003.

[7] http://www.sick.com/home/en.html

[8] P. Dias, "Three dimensional Reconstruction of Real World Scenes Using Laser and Intensity Data," PhD thesis, University of Aveiro, September 2003.

[9] http://www.rieglusa.com/

[10] http://bj.middlebury.edu/~schar/stereo/web/results.php

[11] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk "The Digital Michelangelo Project: 3D scanning of large statues," *SIGGRAPH 2000*, New Orleans, LA, 24-28 July, 2000.

[12] D. Cobzas, H. Zhang, and M. Jagersand, "A comparative analysis of geometric and image-based volumetric and intensity data registration algorithms," *Proceedings of IEEE International Conference on Robotics and Automation (ICRA 2002)*, Washington DC, 2002.

[13] S. E. H. Adelson, and J. R. Bergen, "The plenoptic function and the elements of early vision," Computational Models of Visual Processing, Chapter 1, Edited by Michael Landy and J. Anthony Movshon. *The MIT Press*, Cambridge, Mass. 1991.

[14] P. Ramanthan, "Compression and Interactive Streaming of Light Fields," Ph.D. Thesis, Stanford University, March 2005.

[15] http://www.vision.caltech.edu/bouguetj/calib_doc/index.html

[16] H. Aanæs, "Methods for Structure from Motion" Ph.D. Thesis, Technical University of Denmark, 2003.

[17] C. Zhang and T. Chen, "View-Dependent Non-Uniform Sampling for Image-Based Rendering", *ICIP 2004*, Singapore, Oct 2004.

[18] C. Matabosch, J. Salvi, and D. Fofi, "A New Proposal to Register Range Images," *QCAV2005*, May 2005.

[19] http://www.instop.es/CYRAX/Cyrax.htm

[20] C. Früh and A. Zakhor, "Data Processing Algorithms for Generating Textured 3D Building Façade Meshes From Laser Scans and Camera Images," *Proc. 3D Data Processing, Visualization and Transmission 2002*, Padua, Italy, June 2002.

[21] R. Hartley and A. Zisserman, "Multiple view geometry in computer vision," *Cambridge University Press*, 2000.