Calibrated Imagery for Quantitative Evaluation of IU Classification, Pose-Estimation, and Stereo Algorithms

Jonathan C. Owen, H. James de St. Germain, Stevan Stark,
Thomas C. Henderson, and **William B. Thompson**
Department of Computer Science
University of Utah
Salt Lake City, UT 84112
http://www.cs.utah.edu/projects/robot/

Abstract

Quantitative evaluation of general purpose computer vision algorithms requires testing data including imagery, camera calibration information, and any necessary information about the true nature of the scene relevant to whatever function the vision algorithm is intended to serve. This paper describes the creation of test data for the evaluation of recognition, pose estimation, and stereo surface reconstruction methods. Since the accuracy of pose estimation and stereo reconstruction must be measured with respect to scene geometry, special care is taken to gain knowledge about object shape and orientation with respect to the cameras which acquire the actual imagery.

1 Introduction

While the need for calibrated data with which to quantitatively evaluate the performance of computer vision algorithms has been reiterated for many years, such data is still not widely available. The creation of imagery for this purpose requires more than just camera calibration. Many computer vision methods are intended to recover some sort of geometric description of a scene. Other methods such as model-based object recognition use geometric information in order to recover more qualitative scene descriptions. If such systems are to be tested in a meaningful way, it is necessary to know the "true" geometries involved.

We are creating a data set consisting of imagery and sufficient collateral information to support the evaluation of computer vision methods for three sorts of generic tasks:

- *Model-based and exemplar-based object recognition.*

  Classify objects based on either geometric object models or on training images of the objects.

- *Model-based and exemplar-based pose estimation.*

  Determine object orientation relative to the camera, based on either geometric object models or on other images of the same object from known poses.

- *Depth reconstruction from binocular stereo.*

  Produce a depth map from a stereo image pair and camera calibration information.

The data set will contain geometric object models, information on camera calibration, images of isolated objects in a variety of known poses, and images of collections of objects with a portion of those objects in known poses. All imagery will be collected as stereo pairs. The variety of available data will allow for the evaluation of many different approaches to classification and pose estimation. Model-based methods can use the information of true object shape to drive recognition and orientation determination. This can be done for either isolated objects on a dark background, for which segmentation is straightforward, or for objects collected into a "jumble" in which segmentation and dealing with occlusion become major problems. Alternately, a portion of the images of isolated objects, together with information on the pose of those objects, can be used as training data for recognition and/or pose estimation algorithms, which are then tested on the remainder of the imagery. In either case, knowledge of the actual pose of objects in every image allows quantitative evaluation of pose-estimation algorithms. Stereo reconstruction of surface shape can also be evaluated quantitatively, since the actual surface geometry is known for some or all objects in each image.

2 Objects

The majority of objects used to create the data set are taken from the "Hard-Copy" Benchmark Suite [Thompson and Owen, 1994]. The full benchmark suite includes nine distinct objects, plus 2 duplicates. One set of objects can be assembled together to make a simplified robot gripper (Figure 1). The individual parts of the gripper are simple polyhedra with a few cylindrical holes. The remaining objects come from various versions of the Utah mini-Baja/formula SAE racing vehicle. Four suspension pieces are included (Figures 2–3 and 5–6). While relatively simple, they cover a representative range of machined fea-

This work was supported by the Advanced Research Projects agency under Army Research Office grant number DAAH04-93-G-0420 and by the Office of Naval Research under grant N00014-93-1-0950.

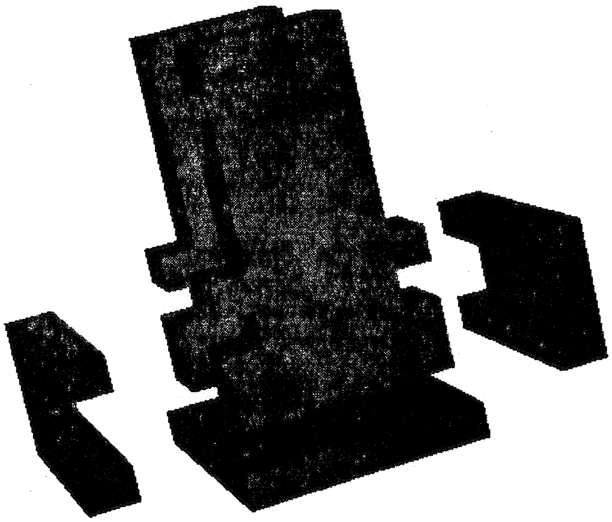

Figure 1: Simple gripper assembly.

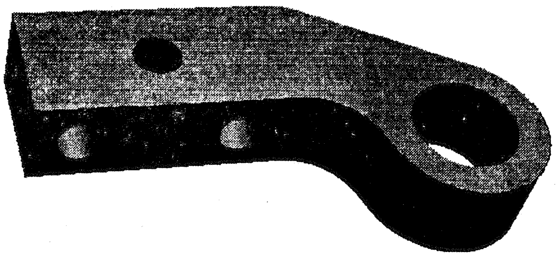

Figure 2: Steering arm.

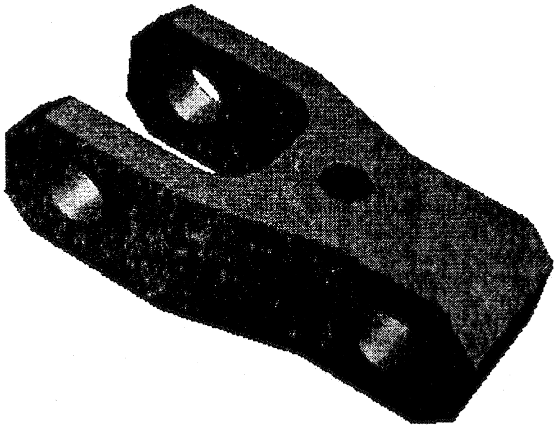

Figure 3: Upper suspension link

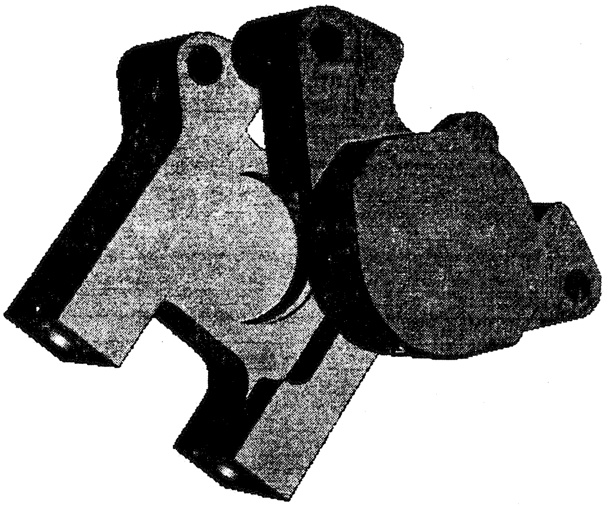

Figure 4: Brake caliper assembly.

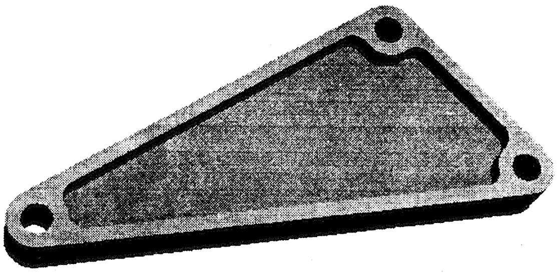

Figure 5: Shock absorber linkage.

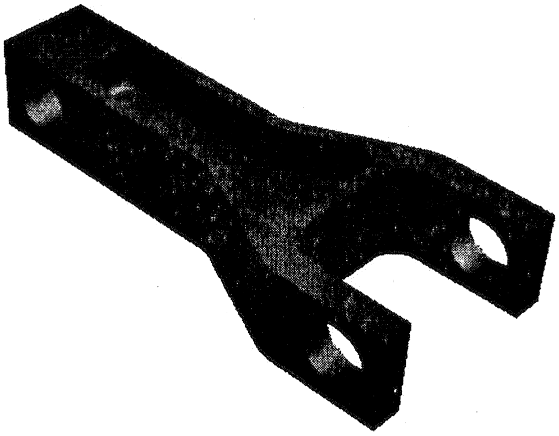

Figure 6: Lower suspension link.

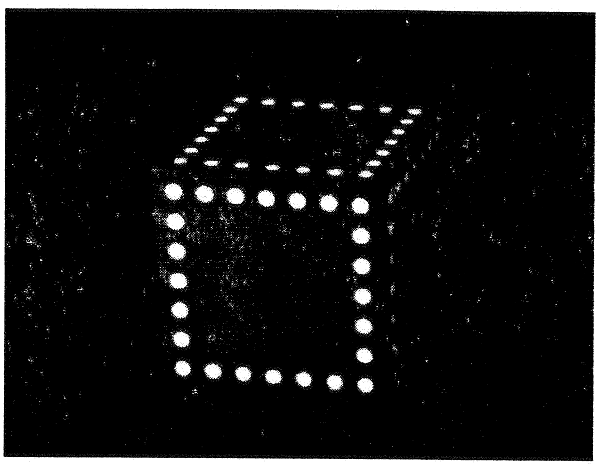

Figure 7: Image of camera calibration cube.

tures. In addition, the two halves of a disk brake caliper assembly are part of the test set (Figure 4).

To make sure we know the actual geometry of each object, all were designed and manufactured in-house using modern CAD technology and precision NC machining. Object geometry will be made available in the IGES, STL, and Alpha_1 CAD formats and as polygonal approximations using the newly emerging VRML standard. Because we manufacture the objects in the Hard-Copy Benchmark Suite under controlled conditions, we can also make available accurate replicas of the same parts as used in our calibrated imagery for use in the evaluation of active vision systems. The objects in the Benchmark Suite are machined aluminum and are thus highly specular. Since this causes difficulties for many current generation computer vision systems, all imagery will be shot twice, once with the metallic part(s) and once with the parts(s) "painted" by spraying them with a talc-like powder that results in a relatively matt finish.

Though substantial progress has been made in automated techniques for recovering surface shape from stereo imagery, most systems still cannot deal with large, visually homogeneous regions. These regions are common in man-made objects such as those described above. As a result, we will include additional stereo imagery of objects with distinct visual texturing over their surfaces. Since CAD models for these objects will not be available, actual surface geometry will be measured using a DIGIBOT II laser scanner.

3 Data Collection

Camera models are determined by imaging a calibration target originally supplied with a K^2T GRF-II structured light range finder. The target is a dark plastic cube 150 mm on a side, with 20 white dots inset around the outside of each face, 25 mm from the edge and 20 mm from each other (Figure 7). Dots can easily be located to sub-pixel accuracy in images of the cube. We use these measured image positions, together with the known 3–D locations of each dot center, to do camera

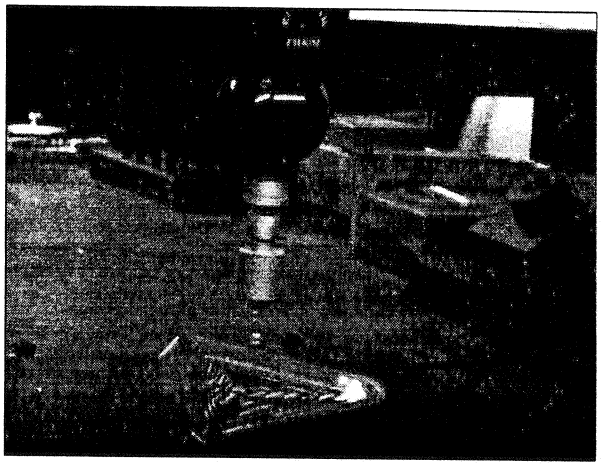

Figure 8: Measuring object pose with a CMM.

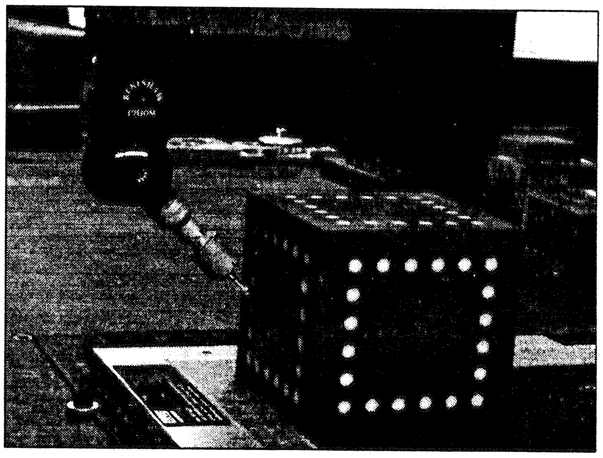

Figure 9: Calibration target in CMM workspace.

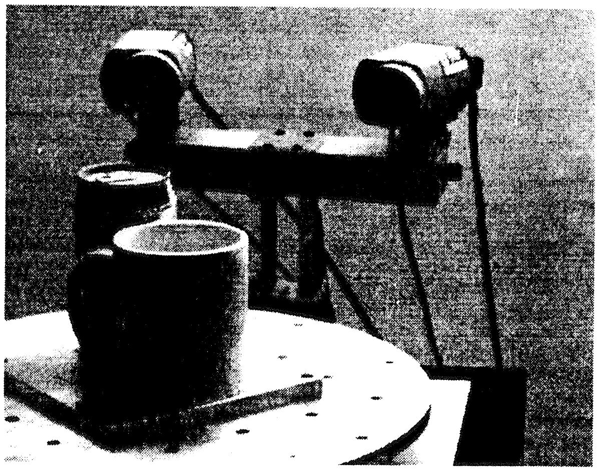

Figure 10: Acquiring a stereo image pair.

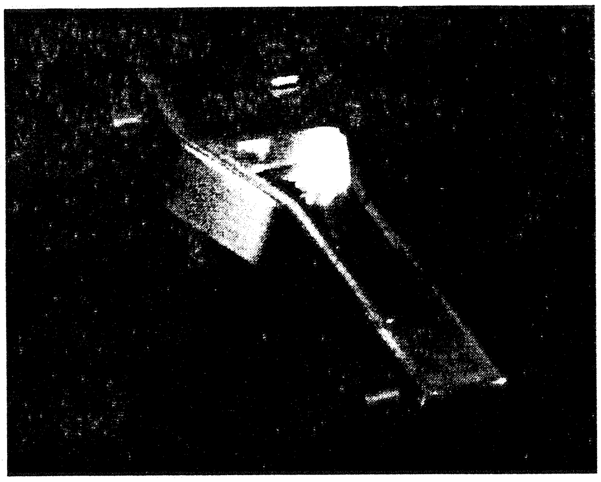

Figure 11: Lower link – orientation 1

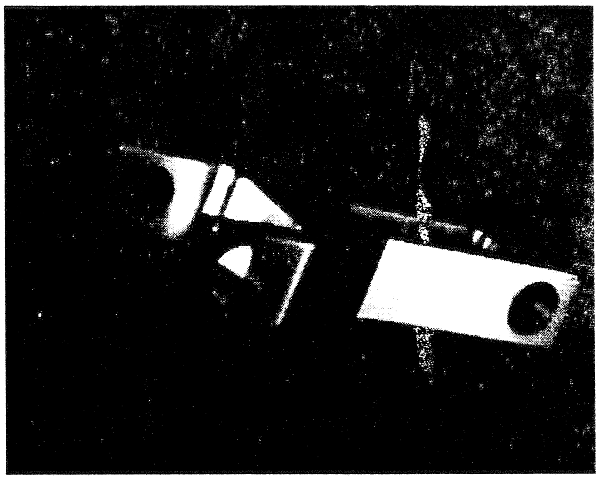

Figure 14: Lower link – orientation 2

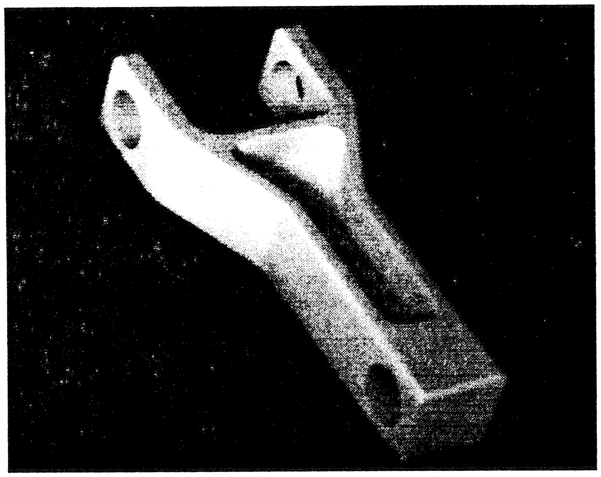

Figure 12: Painted lower link – orientation 1

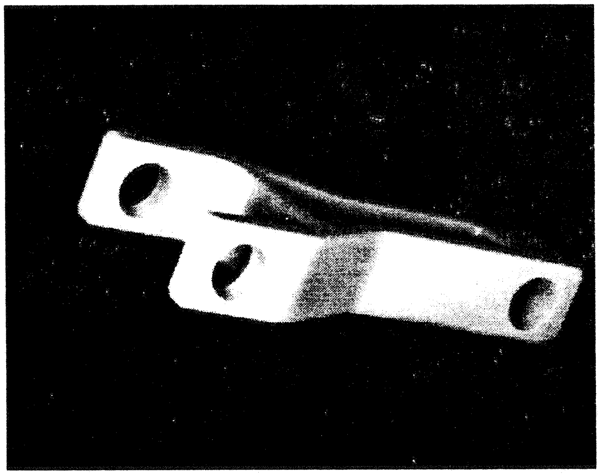

Figure 15: Painted lower link – orientation 2

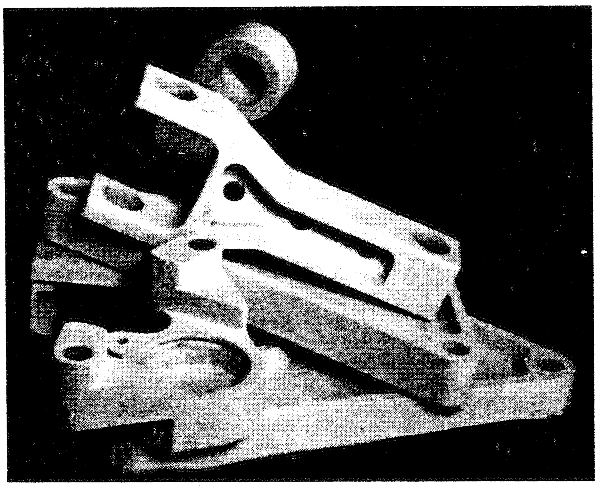

Figure 13: Painted jumble

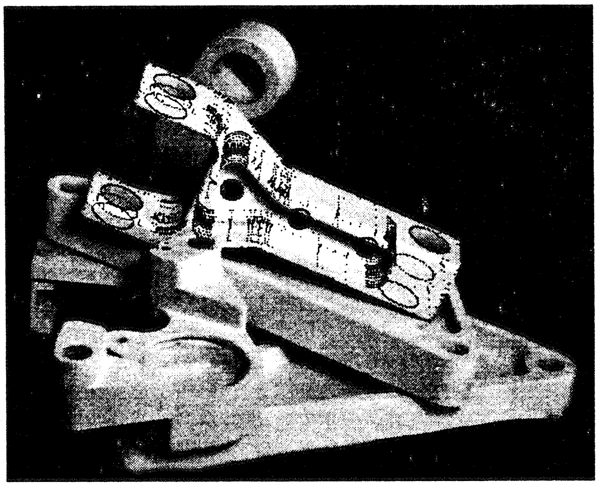

Figure 16: Model back projected onto image of jumble

calibrations using the methods described in [Tsai, 1987] and [Faugeras, 1993]. This produces a camera model in the coordinate system of the calibration cube.

Object pose is measured using a contact coordinate measuring machine (CMM) to measure the position and orientation of three or more planar faces (Figure 8). The transformation between the coordinate system used to specify the object model and the CMM coordinate system is found by associating these measured planes with corresponding planar surfaces in the model and then applying a technique similar to [Shum *et al.*, 1994] (Figure 9). The camera calibration cube is also measured using the CMM and the transformation between cube and CMM coordinates determined. A simple composition yields the transformations between the object and camera coordinate systems.

A slightly different approach is used to collect stereo data for objects for which there is no CAD model. In this case, all data collection is done within the workspace of the DIGIBOT II laser position digitizer (Figure 10). Cameras are calibrated as before, except that for technical reasons a smaller, 75 mm, 24 dots per face calibration cube was used with only two faces visible. The calibration target and all objects of interest are then scanned using the DIGIBOT, resulting in a 3–D point cloud for each. Planes are extracted from the point cloud corresponding to the calibration target and used to get the transformation between DIGIBOT and camera coordinate systems.

Two related methods are available for generating the "true" depth map associated with a stereo image pair. The most straightforward takes a dense sampling of 3–D points on the object surface and projects them through the camera model determined using the calibration procedure, using a software Z-buffer to save only the nearest point at each pixel. When a CAD model is available for an object, the points run through this projection are determined from the model using standard CAGD techniques. Otherwise, the points used are those obtained through laser scanning. Filtering is used to fill holes and provide some amount of noise immunity. The second approach uses rendering techniques driven by either a polygonal approximation derived from the CAD models or a triangulated mesh fit to measured surface points [Hoppe *et al.*, 1993]. While this has a number of theoretical and practical advantages over the previous method, these advantages are largely lost if there are any errors in the polygonal approximation.

4 Imagery

Figures 11–12 and 14–15 show images of one object in two different orientations, with and without the powdered spray. Figure 13 shows the same object in a jumble with other objects also part of the data set. A small amount of occlusion is present for the lower link. The two shock absorber linkages and the gripper assembly are subject to substantially more occlusion. Figure 16 shows the results of taking vertices in the polygonal approximation of the lower link derived from the CAD model of the link, back-projecting them through the pose and camera model transformations, and overlaying the

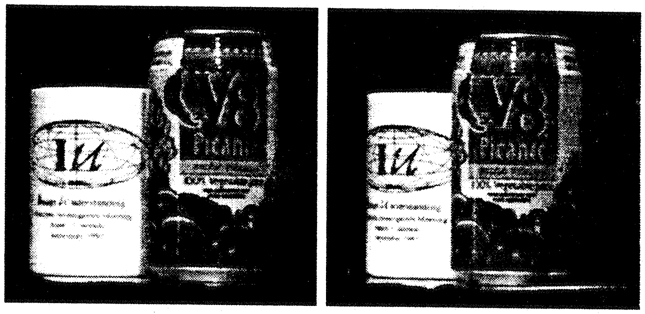

Figure 17: Stereo image pair.

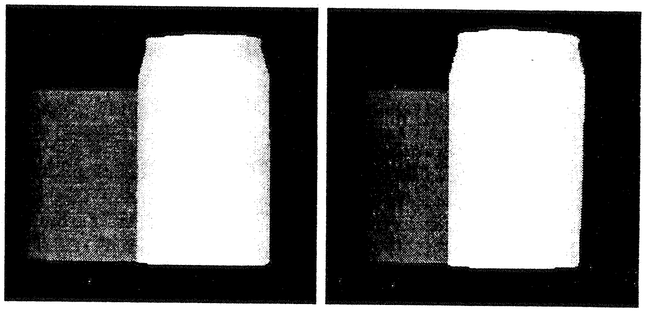

Figure 18: Intensity-coded range images corresponding to Figure 17

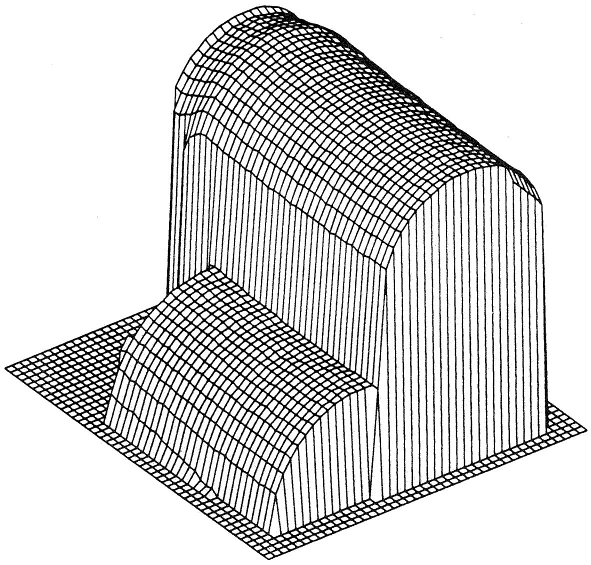

Figure 19: Depth map associated with right image in Figure 17

result on the image. Such back-projections are not intended to be part of the released data set, but instead are used to verify that the various calibration steps have been performed correctly.

Figure 17 shows a stereo image pair of two objects for which we do not have the original CAD models. Figure 18 shows an intensity coded range image for the left and right frames in Figure 17, generated by projecting data acquired with the laser scan back through each camera model. Figure 19 shows the synthetic range image for the right camera, displayed as a perspective plot.

Acknowledgments

Mohamed Dekhil and Alyosha Efros provided the camera calibration codes used in this project.

References

[Faugeras, 1993] O. Faugeras. *Three-Dimensional Computer Vision*. MIT Press, Cambridge, MA, 1993.

[Hoppe *et al.*, 1993] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle. Mesh optimization. In *Computer Graphics, SIGGRAPH '93*, volume 27, August 1993.

[Shum *et al.*, 1994] H. Y. Shum, K. Ikeuchi, and R. Reddy. Virtual reality modeling from a sequence of range images. In *Proc. ARPA Image Understanding Workshop*, pages 1189–1198, November 1994.

[Thompson and Owen, 1994] W. B. Thompson and J. C. Owen. "Hard-copy" benchmark suite for image understanding in manufacturing. In *Proc. ARPA Image Understanding Workshop*, pages 221–227, November 1994.

[Tsai, 1987] R. Y. Tsai. A versitile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, August 1987.