Benchmarking 3D Reconstructions from Next Best View Planning

Christoph Munkelt
Fraunhofer IOF
Albert-Einstein-Straße 7, 07745 Jena, Germany
christoph.munkelt@iof.fraunhofer.de

Michael Trummer, Joachim Denzler
Chair for Computer Vision
Friedrich Schiller University of Jena
Ernst-Abbe-Platz 2, 07743 Jena, Germany
{trummer,denzler}@informatik.uni-jena.de

Stefan Wenhardt
Chair for Pattern Recognition
Friedrich Alexander University of Erlangen
Martensstr. 3, 91058 Erlangen, Germany
wenhardt@informatik.uni-erlangen.de

Abstract

The problem of planning the Next Best View (NBV) still poses many questions. However, the achieved methods and algorithms are hard to compare, since researchers use their own test objects for planning and reconstruction and compute specific quality measures. Consequently, these numbers make different statements about different objects. Thus, the quality of the results and the performance of the methods are not easily comparable. In order to mend this lack of measure and comparability, this paper suggests a test object together with a reference benchmark. These allow comparison of reconstruction results from different NBV algorithms achieved with different techniques and various kinds of sensors.

1. Introduction

In the past years, a lot of work has been done in the area of active vision and sensor planning (cf. [3-7]). However, two major problems of all past activities can be stated. First, researchers use their own test objects. Second, the reconstruction quality is measured by different criteria or, if these are denoted the same, the criteria may be formalised in different ways. Because of these shortcomings in standardisation, the studied methods cannot be compared and the state of the art in this research area cannot be reviewed without further considerable effort, since the exact relative performance of the methods is unknown.

A solution to this problem is a common reference test object entailing challenges to all kinds of active vision systems, like different scanner hardware and different planning methods. This can be completed by a reference benchmark judging the quality of the reconstructed test object and, hence, judging the performance of the whole NBV planning system. Thus, the desired benchmark has to overcome the difficulty of being independent from the scanner hardware used (e.g. IR/laser scanners, intensity cameras with/without fringe projection) and from the actual planning methods.

The contribution of this paper is a discussion of different elements that a test object should consists of, in order to be a challenge for a wide range of algorithms. Additionally, we suggest formal criteria in order to judge reconstruction results in the context of NBV. The idea is to provide the community with a standardized test environment to make comparison of algorithms possible in the future.

The remainder of the paper is organised as follows. The next section gives a literature review showing different test objects previously used. In section 3, criteria for a common test object are proposed together with a suggestion of one sample of such a object. Section 4 discusses elements of a reconstruction benchmark. Section 5 shows first results using the test object with two different approaches for reconstruction and NBV planning, and illustrate the proposed benchmark. The paper ends with a summary and outlook in section 6.

2. Literature Review

The variety of test objects makes evaluation of the methods difficult. In this situation, as Scott [8] states, the field could benefit from standardized performance benchmarks and quantitative performance analysis.

To evaluate their algorithms, authors use different geometrical primitives (L-shaped block, wedge [9], cube, sphere, torus [10]), various small objects (cups [7,10], duck/temple/dino/bird/dog models [11,1]), as well as several busts [17]. Some attempts have been made to construct specific test objects, either as building blocks scene [14], or as simulation models [12,13]. However, from a benchmarking perspective, most of those approaches lack certain features which are needed in a view planning framework.

First of all, the complexity (regarding NBV planning) of most used test objects is rather low, especially in the sense that there are only small regions with (self-) occlusions (with the exception of the temple [1] and simulation models [12,13]). This leads to nearly 99% completeness within few views taken. While this might be adequate for a proof-of-concept, this is not a satisfying starting position for a detailed comparison between planning algorithms. Seitz et al. [1] state e.g. that the completeness numbers were not very discriminative. Therefore an adapted test object should contain a suitably large amount of concavities and occluded surface areas.

Second, most objects are chosen with a specific scanning technology in mind. But the idea to benefit from collaboration between different reconstruction approaches should result in a test object which is useful for a variety of methods *by design*. This includes surface texture as well as increasing complexity of individual details.

Third, while most of the planning algorithms consider grazing angle as a subjective quality measure [8], we think it is crucial to evaluate the *quality of the final reconstruction* as well as the surface coverage. Seitz et al. [1] do so by modeling an error, which 90% (user threshold) of the reconstructed points do not exceed. Girod et al. [17] use the difference volume between surface meshes.

Finally, different performance measures have been used. Most of the authors considered the number of views and completeness, while the latter was often not formalised. Only few authors take accuracy into account [1]. However, [8] proposes the following three measures for evaluating view planning algorithms: view plan quality (in terms of the quality of the reconstruction), view plan efficiency (in terms of views needed) and view plan computational efficiency (in terms of complexity and time). They also state that there is a lack of clarity with respect to the viewpoint planning objective, whether it is some global optimum, „acceptable“ viewpoints or a minimum number of views.

Although there is a clear demand for a reference benchmark, it is clear that a common minimal set of object details (posing special challenges) is hard to find. In the next sections, we present generic geometric details that a test object should consists of together with a discussion of an associated formal benchmark.

3. Reference Object in Detail

General benchmark criteria The objectives of a benchmark in the context of NBV planning can be summarized as follows. The primary goal is to evaluate the reconstruction quality a NBV algorithm achieves. As section 4 shows in detail, this is done by combining the results of the various object details introduced with the complete reconstruction. Therefore, the primary objective of the NBV algorithm in question is to maximize the completeness and accuracy of the reconstruction (similar to „view plan quality“ [8]). The secondary condition is to minimize the number of views – and time – needed to achieve the reconstruction quality (similar to „view plan efficiency“ [8]).

Derivation of a reference object In this section we want to discuss attributes, a NBV test object should possess. We then show how we realised those attributes using certain object details by presenting one specific reference object prototype (details denoted by their number in brackets, cf. figure 1 for an example object). A test object should not be symmetric (see details 1,2,4). This often leads to overly simplified, regularly spaced view plans. Furthermore, self occlusions are needed to challenge the planner (1,3,4,7,8). Additionally, when using active fringe projection systems, shadows should be cast onto the object (1). Curved surfaces (2,4,8) as well as sharp edges pose difficulties to different reconstruction approaches. To test the incorporation of the sensor model into the planner we need a detail which requires special sensor alignment to it (5). The access to holes is a difficult planning issue and has to be tested (5,6). Optionally, length errors could be tested (7,8). Finally, a scanner resolution estimate should be provided (9).

We now describe the proposed details. Their numbers reference them in fig. 1. Further views of the prototype object together with a description for POV-Ray are available at [15].

- *Basic object setup*. This subset examines basic planning and reconstruction capabilities. It consists from the following details:
  1. „Tripod“: Three small cuboidal elements, which occlude (and shade) parts of the „sinusoidal face“. Both number of points and completeness in its junctions are demanding.
  2. „Sinusoidal face“: Yields an asymmetric overall shape of the test object. Has a smooth, but curved surface. Varying surface normals complicate reconstruction.
  3. „Notch“: Used as a cavity for two details of the full benchmark. Its side faces need to be scanned from appropriate views.
  4. „Negative half sphere“: Constrained visibility and shadows challenge the reconstruction of its interior.
- *Full object setup*. All details combined pose a challenge to more sophisticated scanning and/or planning methods. The full benchmark extends the basic one by the following, additional details:
  5. „Slotted hole“: Scanning this concavity requires a certain alignment of stereo systems. Concealable.
  6. „Drill holes“: Three holes with proportions of diameter to depth from 2:1 through 1:1 to 1:2 act as prototypes for concavities of interest with increasing difficulty. Concealable.
  7. „Frustum of pyramid“: Common test detail with hard to scan side faces. Optionally, one could compare both the planarity of its five upper faces and the length deviation of their resulting intersections to their corresponding ground truth values. Pluggable.
  8. „Positive half sphere“: Test detail with calibrated radius. Sphere base is hard to scan. Pluggable.
  9. „Riffle plate“ - optional: Plate with a 2D array of miniature frustrums of pyramids. The size of their top face decreases. Can be used to determine scanning resolution in object space. Pluggable.

These details, surely, do not represent all objects in the world, but as a union of abstract challenges they cover a wide range of real-world objects.

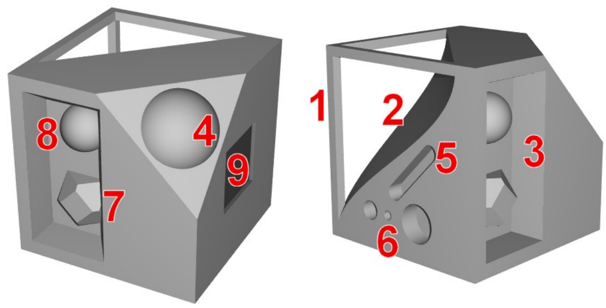

Figure 1. Proposed prototype of a test object consisting of the details described. The numbers relate to the individual details in the text.

Our Prototype The proposed test object has the overall shape of a cube, but it is not symmetric. It combines a set of object details covering the discussed requirements. By constraining its size to approximately 20 x 20 x 20 cm^3, the object satisfies the (often implicit) demand for complete enclosure into the measurement volume. To satisfy the need for a fine grained surface texture, laser labeling can optionally be applied. A common texture pattern for passive lighting techniques is available at [15]. Adaptation to different complexity levels can be achieved through a plug-in architecture of several details, which can be added as needed. The test object has five faces of interest, which can be scanned from the upper half sphere around it. While normally standing on its bottom side, it can be placed on any of its sides.

This test object is not intended to evaluate a scanners physical resolution. However, to reasonably rate the reconstruction precision, a relative accuracy (regarding special object details) of $30\mu m$ for smoothness as well as length deviation is targeted for.

4. Evaluation of the Planning Result

For benchmarking the reconstruction result of a NBV planning algorithm, we take the reconstruction quality and the *number of views* v (cf. view plan efficiency from [8], but avoiding interdependencies) into account. The reconstruction quality is represented by the *coverage* c, the *average distance of neighboring points* d_{av} within the reconstruction and the *average error* e_{av}. Further specification of these items will be carried out in the next section. The benchmark for the reconstruction of the whole reference object is given by the tuple

$$b=[c, d_{av}, e_{av}, v]. \tag{1}$$

The number of views is considered for the whole object only. For a certain object detail i the benchmark is

$$b_i=[c_i, d_{av,i}, e_{av,i}]. \tag{2}$$

The coverage $c_{(i)}$ describes, what percentage of the reference surface is covered by reconstructed points. This value is independent of the point density of the reconstruction and does *not* need any user threshold. For computation, an equidistant point cover (equidistant points created on the reference CAD surface) is created, holding the same number of points as the reconstruction. The same technique can be utilized to yield points from continuous surfaces. Then, for each reconstructed point, the closest point of this point cover is marked. Now, the coverage is given by the percentage of marked points within the point cover. This measure implicitly demands a homogeneous distribution of the reconstructed points on the object surface. As another important aspect, the reconstruction point density should be revealed in some way. This can be achieved by computing the average distance $d_{av(,i)}$ between a reconstructed point and its nearest neighbor within the set of all reconstructed points. In combination, these two values correspond to the *completeness* of the reconstruction. Reconstruction *accuracy* is modeled inversely by the modified Hausdorff distance $e_{av(,i)}$ (average error) of a reconstructed point from the reference surface (as also used in [2] for object matching).

Establishing the overall benchmark *b* The individual object details pose different challenges to NBV planning approaches. Therefore, we combine the values belonging to one criterion, weighting equally the value of the whole object and the mean value of the special details. As an example, to compute the overall coverage c for the basic benchmark, we combine the coverage c_{whole} of the whole basic reference object and the coverage values $c_i\ (i=1,\ldots,F)$ of the F details (basic: F=4) by the following formula:

$$c=\frac{1}{2}\left(c_{whole}+\frac{1}{F}\sum_{i=1}^{F}c_i\right). \tag{3}$$

We do this in an analogous manner for d_{av} and e_{av}. By these weighted combinations, we yield a global reconstruction benchmark as shown in equation (1).

5. Experiments

In order to verify and discuss the proposed test object together with the benchmark, we perform experiments with two different NBV planning approaches. As a first step of the verification of the desired characteristics, these tests are realised within a simulation environment. The first method is described in [13]. It uses active fringe projection (active illumination) and incorporates a rough prior shape model for planning the NBV. The second one (see [6]) is an information theoretic approach without active illumination. The simulation framework allows reconstructing points within the reference coordinate frame. Therefore, an initial fitting step (before applying the benchmark) can be omitted. In practice, this alignment could be done by using the ICP algorithm. Figure 2 shows the respective reconstructions of the negative half sphere of the two methods. The point sets triangulated for illustration.

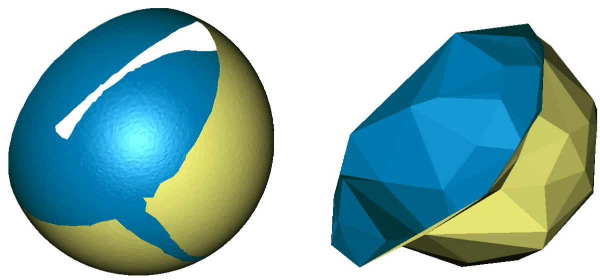

Figure 2. Reconstructions of negative half sphere, results from NBV planning approaches with ([13], left) / without ([6], right) active illumination. Whole object observed by 8 / 10 views.

Table 1 shows the benchmark results and demonstrates the comparison of the computed reconstructions. The results are compared by coverage, average minimal point distance and the average error of a reconstructed point as shown in equation (2). The values are computed as described in the previous section with the exception of the average error.

The possibility of incorporating the CAD model is not implemented, yet. As a temporary workaround, we generate a dense point cover (average minimal point distance $100\mu m$) of the object. The error distance of a reconstructed point then is calculated as the distance from the next point within this dense point cover. The benchmark result for the whole object is completed by the number of views (cf. equation (1)).

Table 1. Benchmark results. Tuples $[c_i, d_{av,i}, e_{av,i}]$ for object details and $[c, d_{av}, e_{av}, v]$ for the whole object (basic benchmark)

object detail	reconstruction according to [13]	reconstruction according to [6]
tripod	[69%,0.34mm,0.10mm]	[58%,6.59mm,1.43mm]
sinusoidal face	[73%,0.41mm,0.08mm]	[47%,7.18mm,1.58mm]
notch	[75%,0.51mm,0.18mm]	[51%,12.70mm,5.62mm]
neg. sphere	[77%,0.55mm,0.12mm]	[57%,7.34mm,2.42mm]
whole basic object	[70%,0.31mm,0.11mm]	[52%,7.52mm,1.50mm]
basic benchmark	**[71%,0.39mm,0.12mm,8]**	**[53%,7.98mm,2.13mm,10]**

The reconstructed point sets differ significantly in terms of completeness and accuracy. In both cases coverage is significantly below 100%. The reason for this is that coverage, as defined in section 4, incorporates the distributional homogeneity of the reconstruction. Hence, if the local point density is low (equivalently, the average minimal point distance is locally high) this can be seen as a hole in the reconstruction. In this sense the coverage is lower. The average minimal point distances differ notably between the two results. This is caused by a significantly larger number of reconstructed points from the reconstruction using active illumination. The larger average error of the reconstruction without active illumination is surely caused by the quality of the point feature detection without an active light source. There seems to be a contradiction between the coverage values and the illustrations in figure 2. The reconstruction result on the right does not show any holes in the mesh, but its coverage is lower than the other one. The reason is that our approach does not benchmark mesh surfaces, but point sets with regard to the reference model. This is further amplified by requiring distributional homogeneity of the reconstructed points. So, the contradiction is resolved when considering the inhomogeneous point distribution of the reconstruction result on the right.

6. Summary and Outlook

In this paper, we motivated the necessity of a common test object and a common benchmarking strategy for NBV planning in 3D reconstruction. Desired attributes of the test object were encouraged within a literature review. We proposed a test object and a benchmark for 3D reconstruction with special regard to NBV planning. The influence of the desired attributes on the design of the test object and the special object details was shown. By the proposed benchmark we were able to, quantitatively, compare results from the application of different NBV planning approaches.

However, the error value of this benchmark does not incorporate the maximal reconstruction error (maximal deviation of a reconstructed point from the reference model). This data could be important for some applications. Therefore, the way of regarding the maximal deviation needs further analysis.

The simulations were performed without initial fitting steps, since the coordinate frames of reference and reconstruction were the same. But in general, the problem of fitting the respective points to the reference model can cause problems, especially for single details with only a few points. Additionally, it must be ensured that the mapping of reconstructed points to a special object detail is unique. We will address this in future research.

As the next step, the simulation results have to be confirmed by results from experiments with a physical version of the test object. Further experiments with comparable reconstruction techniques will be conducted to prove effectiveness of the proposed benchmarking framework.

References

[1] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, R. Szeliski: "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms", *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 519-528, 2006

[2] M.-P. Dubuisson, A.K. Jain: "A modified Hausdorff distance for object matching", *Proceedings of 12th International Conference on Pattern Recognition*, pp. 566-568, 1994

[3] G. Olague, R. Mohr: "Optimal camera placement for accurate reconstruction", *Pattern Recognition Journal*, vol. 35, pp. 927-944, 2002

[4] F. Chaumette, S. Boukir, P. Bouthemy, D. Juvin: "Structure From Controlled Motion", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 5, pp. 492-504, 1996

[5] E. Marchand, F. Chaumette: "Active Vision for Complete Scene Reconstruction and Exploration", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 1, pp. 65-72, 1999

[6] S. Wenhardt, B. Deutsch, J. Hornegger, H. Niemann, J. Denzler: "An Information Theoretic Approach for Next Best View Planning in 3-D Reconstruction", *Proc. of the 18th Int. Conf. on Pattern Recognition*, vol. 1, pp. 103-106, 2006

[7] R. Pito: "A Solution to the Next Best View Problem for Automated Surface Acquisition", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 1016-1030, 1999

[8] W.R. Scott, G. Roth: "View Planning for Automated Three-Dimensional Object Reconstruction and Inspection", *ACM Computing Surveys*, vol. 35, no. 1, pp. 64-96, 2003

[9] D.R. Roberts, A.D. Marshall: "Viewpoint selection for complete surface coverage of three dimensional objects", *Proc. of British Machine Vision Conference*, vol. 1, pp. 740-750, 1998

[10] L.M. Wong, C. Dumont, M.A. Abidi: "Next Best View System in a 3-D Modeling Task", *Proc. of Int. Symp. on Computational Intelligence in Robotics and Automation (CIRA)*, vol. 1, pp. 306-311, 1999

[11] Y.F. Li, B. He, S. Chen, P. Bao: "A view planning method incorporating self-termination for automated surface measurement", *Measurement Science and Technology*, vol. 16, pp. 1865-1877, 2005

[12] J.E. Banta, L.M. Wong, C. Dumont, M.A. Abidi: "A Next-Best-View System for Autonomous 3-D Object Reconstruction", *IEEE Trans. on Systems, Man and Cybernetics*, vol. 30, no. 5, pp. 589-597, 2000

[13] C. Munkelt, P. Kühmstedt, J. Denzler: "Incorporation of a-priori information in planning the next best view", *Proc. of Vision, Modeling, and Visualization (VMV)*, pp. 261-268, 2006

[14] J. Maver, R. Bajcsy: "Occlusion as a Guide for Planning the Next Best View", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 5, pp. 417-433, 1993

[15] Reference Object and Benchmark for Next Best View Planning. Homepage: www.inf-cv.uni-jena.de/nbvbench

[16] M. Benz, J. Hartmann, T. Maier, E. Nkenke, K. Veit, A. Stellzig-Eisenhauer, F.W. Neukam, G. Häusler: "Optical 3d-metrology for medical applications", *Proc. of ICMP 2005 and BMT 2005*, Fachverlag Schiele und Schön, Berlin, 2005

[17] B. Girod, G. Greiner, H. Niemann: "Principles of 3D Image Analysis and Synthesis", *Kluwer Academic Publishers*, Boston – Dordrecht – London, pp. 166-180, 2000