Baltic J. Modern Computing, Vol. 13 (2025), No. 1, pp. 157–165
http://doi.org/10.22364/bjmc.2025.13.1.08

# A Qualitative Comparison of the State-of-the-Art Next-Best-View Planners for 3D Scanning

Andrejs ARISTOVS[1], Evalds URTANS[2]

[1] Riga Technical University, Riga, Latvia
[2] Riga Technical University, Department of Artificial Intelligence and Systems Engineering, Riga, Latvia

andrejs.aristovs@edu.rtu.lv, evalds.urtans@rtu.lv

ORCID 0009-0008-1806-1772, ORCID 0000-0001-9813-0548

**Abstract.** This is a survey paper in which we review the state-of-the-art Next-Best-View planners with the focus on their application in solving an autonomous 3D scanning task. According to market reports, the 3D scanning market will continue to grow in response to the increasing demand for augmented and virtual reality solutions. Taking into account that the number of skilled 3D artists is limited and their labor is highly paid, an alternative way of creating high quality 3D models is 3D scanning existing objects. In many cases, 3D scanning is the only way to get photorealistic textures and high-definition models. Automated 3D scanning can be used as a way to preserve art, document changes in the environment, create detailed models of consumer products. Six next-best-view planners were compared using ROS in the Gazebo simulation environment. The MA-SCVP machine learning method achieved on average 93.1% coverage, that is 5.9% higher than ScanRL, 36% higher than SEE, and 1% higher than volumetric information gain methods. Maximum coverage with the MA-SCVP method was achieved after 12.2 views on average, versus 20 views for the volumetric information gain methods.

**Keywords:** Next-best view, 3D reconstruction, ROS, Machine Learning

## 1 Introduction

It is estimated that the 3D scanning market will reach a billion dollars by 2024 as discussed in Kari (2022), the main reason being the applications of AR and VR mostly in the marketing field. According to Boland (2020), photorealistic models create a sense of craving in consumers, improve conversion, and increase session length. In return an improvement in this metrics results in higher income and growth in customer satisfaction. In many AR/VR applications, photorealistic 3D models improve immersion and blend better in the scene. Such models can be created either by skilled 3D artist or by means of automated 3D scanning of real-world objects. So far only large companies such as IKEA have had the ability to digitize their products and create 3D assets of their inventories.

By using appropriate 3D scanning techniques, it is possible to democratize the creation of 3D models. In our experience, structured light scanning yields the best results, by controlling lighting, polarization and camera focus, high detail models with HDR textures can be achieved. The main hurdle is an intelligent way to plan the path with the intent of reducing the necessary number of view points. Structured light scanning creates high-resolution textured 3D models, but each new scan takes up to 10 seconds, depending on the number of patterns being projected. Most of this time is to make sure the system is static, dampen the vibrations, and adjust camera lenses focal length and aperture. With this in mind, we evaluated state-of-the-art NBV planners with the focus on minimal view coverage.

## 2 Related work

Depending on the target use for the NBV planner different metrics are used to evaluate planner's performance. As mentioned earlier for a system that uses structured light 3D scanning, most important metric is number of views needed to achieve threshold reconstruction quality.

### 2.1 Next-best view planner comparison techniques

Authors are aware of the last analysis of NBV planning methods carried out by Scott et al. (2003), in which the comparison metrics have been defined to evaluate different approaches. The evaluation criteria of this publication were reviewed as a basis for our literature analysis.

Many of the state-of-the-art NBV planners are iterations and improvements of previous methods, as the MA-SCVP method introduced in Pan et al. (2023) is an improvement of the SCVP method introduced in Pan et al. (2022) that additionally uses PC-NBV by Zeng et al. (2020) neural network to define the best view. NBV-Net 4-5 neural network architecture introduced in Vasquez-Gomez et al. (2021) (the numbers in the name stand for: 4 convolutional layers and 5 fully connected layers), is based on the previous paper Mendoza et al. (2019) NBV-Net network architecture that contained 3 convolutional layers and 5 fully connected layers. The authors also tested other NBV-Net configuration, like NBV-Net 3-3, NBV-Net 3-5, NBV-Net 4-3, and NBV-Net 5-4, with the conclusion that the NBV-Net 4-5 network achieves the best results. NBV-Net was the first 3D convolutional network architecture applied to solving 3D reconstruction. Multiple further solutions use a similar network architecture and use NBV-Net as ground truth.

An alternative to deep machine learning (ML) based methods are measurement direct methods (SEE Border et al. (2018) and SEE+ Border and Gammell (2022)), (PC-NBV Zeng et al. (2020)) - all using point cloud data to define the region of interest and the next best view.

After literature analysis, the most prominent NBV planners have been selected for further evaluation: MA-SCVP Pan et al. (2023), SEE Border and Gammell (2022), ScanRL Peralta et al. (2020) and volumetric information gain methods such as AE (Average Entropy), RSE (Rear Side Entropy), RSV (Rear Side Voxel), OA (Occlusion Aware) and PC (Proximity Count) by Delmerico et al. (2018) and UV (Unobserved Voxel) by Vasquez-Gomez et al. (2014) and volumetric information gain method defined in Kriegel et al. (2015).

### 2.2 3D model datasets

The ABC dataset introduced in Koch et al. (2018) contains more than 1 million CAD models, downloaded from the Onshape[3] platform. Shapenet Chang et al. (2015) dataset contains three millions of CAD models, 220 000 of which are categorized into 3135 classes. Thingi10K Zhou and Jacobson (2016) dataset contains 10 000 models intended for 3D printing. In general, large-scale datasets are used for training and testing ML algorithms. For 3D reconstruction tasks, smaller datasets with textured models created by 3D scanning are more common.

The models of the bunny, introduced in Turk and Levoy (1994), the dragon by Curless and Levoy (1996) and the armadillo and the Buddha by Krishnamurthy and Levoy (1996) are available on the Stanford University computer graphics laboratory website.[4] The 20 models introduced in Rodolà et al. (2013) are available on the Munich Technical University (TUM) computer vision group website.[5] The 80 detailed models created by means of structured light 3D scanning by Jensen et al. (2014) are available on the Image Analysis and Computer Graphics at the Technical University of Denmark (DTU) website.[6] Several models are available on the MIT Computer Science and Artificial Intelligence laboratory website.[7] The House3K dataset used in Peralta et al. (2020), which contains 3000 building models with textures, is available on the GitHub repository.[8] The Linemod dataset[9] used in Hinterstoißer et al. (2012) and the HomebrewedDB dataset[10] introduced in Kaskman et al. (2019) are available on the TUM websites.

## 3 Methodology

In order to evaluate different NBV planners, the Gazebo simulation environment was chosen, with foresight to later use NBV planners with ROS. Primarily due to authors familiarity with ROS as well as the capabilities to use both Gazebo simulation environment and real-world robotic systems. As different NBV planners were implemented in different environments, an approach only involving the view coordinates was used. By using only coordinates, multiple different environments can be utilized and the results are not software and setting dependent. The positions were transformed to be used

[3] https://www.onshape.com/

[4] http://graphics.stanford.edu/data/3Dscanrep/

[5] https://cvg.cit.tum.de/data/datasets/clutter

[6] https://roboimagedata.compute.dtu.dk/

[7] https://people.csail.mit.edu/tmertens/textransfer/data/

[8] https://github.com/darylperalta/Houses3K

[9] https://campar.in.tum.de/Main/StefanHinterstoisser

[10] https://campar.in.tum.de/personal/ilic/homebreweddb/

with Gazebo (the coordinates X, Y, Z were scaled and the rotations were converted to quaternions). The evaluation pipeline is presented in Figure 1.

The coverage percentage is calculated as a similarity between the ground-truth model and the resulting point cloud. Our approach differs from previously used methodologies as we use more comprehensive 3D test model dataset as well, compare multiple different approach performance combined with the focus on highest coverage percentage achievable in least views possible.

**Algorithm 1** Similarity calculation between pointclouds

```
**Input** cloudA, cloudB, threshold
**Output** cloudA similarity to cloudB
$tree = open3d.geometry.KDTreeFlann(cloudA)$
$num\_outlier = 0;\ num\_valid = 0$
**for** $pt$ **in** $cloudB.points$ **do**
    $dist = nearestDistance(tree, pt)$
    **if** $dist < threshold$ **then**
        $num\_valid = num\_valid + 1$
    **else**
        $num\_outlier = num\_outlier + 1$
**result** $= num\_valid/(num\_outlier + num\_valid)$
```

For the experiments, the threshold value has been set to 0.005 meters or 0.5 mm. CloudB is the point cloud of ground-truth created from the .ply model.

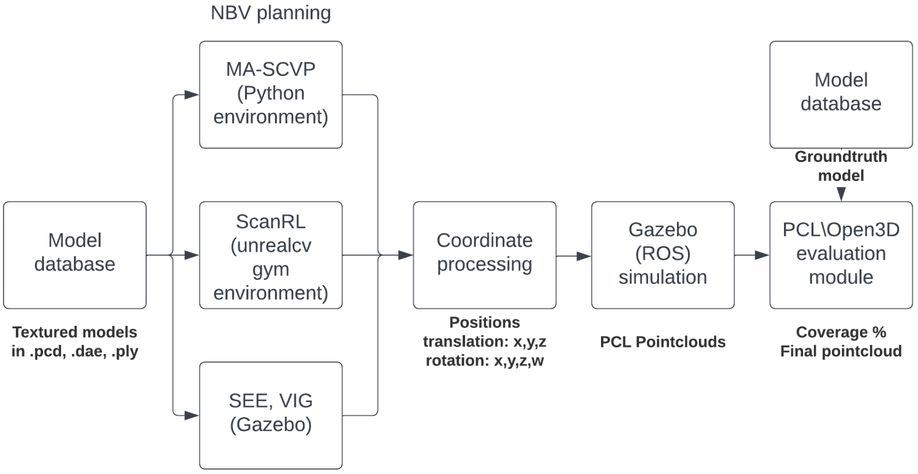
NBV planning
MA-SCVP (Python environment)
ScanRL (unrealcv gym environment)
SEE, VIG (Gazebo)
Model database
Textured models in .pcd, .dae, .ply
Coordinate processing
Positions translation: x,y,z rotation: x,y,z,w
Gazebo (ROS) simulation
PCL Pointclouds
Model database
Groundtruth model
PCL\Open3D evaluation module
Coverage % Final pointcloud

Fig. 1: NBV planner evaluation methodology

To evaluate the performance of the NBV planner, a testing dataset has been assembled. The dataset contains 9 models: bunny and dragon from the Stanford University dataset, can and cat from the LineMod dataset, and 5 models from the HomeBrewDB dataset: mug, minion, dog, stegosaur, and triceratops. Set composition was selected to include the most common test models as well as models with complicated geometries and occlusions. The models have been scaled down, and the geometry has been simplified. To improve reconstruction and perception, bright textures have been applied to the models. From the 3D models, ground truth point clouds have been created that contain on average 52 000 points. Simplified geometry models are not larger in size than 5 MB for the .ply mesh models and not larger than 12 MB for the .dae Gazebo models.

Fig. 2: Dataset of 3D models used for evaluation

## 4 Results

On average, with the same number of views, the MA-SCVP NBV planner achieved 9.2% higher coverage than the volumetric information gain methods. Until 7 views, the volumetric information gain metrics UV (Unobserved Voxel), AE (Average Entropy), or other volumetric information gain methods can achieve higher coverage than MA-SCVP, possibly due to MA-SCVP selecting views to optimize the local path rather than purely maximizing information gain. After about 7 views, MA-SCVP achieves higher coverage than the other evaluated methods.

MA-SCVP created on average a coverage set of 12.2 views, while ScanRL, SEE, and volumetric information gain methods were limited to 20 views. In 8 of the 9 models, MA-SCVP reached the highest overall coverage (with the exception of the bunny model, bunny the percentage of coverage shown in Figure 4). The bunny model was the only model where a larger number of views than those defined by MA-SCVP was beneficial and resulted in higher maximum coverage.

Measurement-direct approach SEE achieved on average a 36% lower maximum coverage than MA-SCVP, but gradually improved the quality of the model, where each next view increased the coverage. With volumetric information gain methods, in some cases, among the 20 views, some of the views were redundant and did not improve coverage.

Examples of the percentage of coverage after 5, 10 views and maximum achieved for the 2 selected models and some of the methods are presented in Table 1.

Table 1:

**Coverage after 5, 10 views and maximum achieved.**

| Model | Metric | MA-SCVP | ScanRL | SEE | UV | VG | AE | RSE | OA | PC | RSV |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Cat | C5 | **0.874** | 0.483 | 0.449 | 0.813 | 0.828 | 0.872 | 0.734 | 0.813 | 0.648 | 0.699 |
| Cat | C10 | **0.932** | 0.754 | 0.492 | 0.915 | 0.926 | 0.907 | 0.854 | 0.915 | 0.872 | 0.832 |
| Cat | Max | **0.957** | 0.866 | 0.619 | 0.942 | 0.945 | 0.925 | 0.943 | 0.942 | 0.934 | 0.938 |
| Bunny | C5 | 0.615 | 0.514 | 0.539 | **0.787** | 0.742 | 0.760 | 0.673 | 0.760 | 0.639 | 0.543 |
| Bunny | C10 | **0.852** | 0.774 | 0.628 | 0.843 | 0.828 | 0.835 | 0.813 | 0.825 | 0.818 | 0.814 |
| Bunny | Max | 0.862 | 0.870 | 0.675 | 0.873 | **0.895** | 0.877 | 0.881 | 0.883 | 0.891 | 0.844 |
| Dragon | R5 | 0.721 | 0.512 | 0.493 | 0.709 | 0.584 | **0.735** | 0.667 | 0.709 | 0.606 | 0.667 |
| Dragon | R10 | **0.857** | 0.703 | 0.535 | 0.782 | 0.789 | 0.769 | 0.812 | 0.782 | 0.751 | 0.788 |
| Dragon | Max | **0.876** | 0.832 | 0.651 | 0.849 | 0.877 | 0.843 | 0.874 | 0.849 | 0.878 | 0.873 |
| Triceratops | R5 | **0.892** | 0.558 | 0.519 | 0.847 | 0.838 | 0.841 | 0.743 | 0.847 | 0.821 | 0.743 |
| Triceratops | R10 | **0.948** | 0.816 | 0.576 | 0.902 | 0.882 | 0.919 | 0.864 | 0.902 | 0.893 | 0.881 |
| Triceratops | Max | **0.973** | 0.900 | 0.597 | 0.952 | 0.956 | 0.946 | 0.956 | 0.952 | 0.953 | 0.953 |

On most of the models, coverage similar to the cat model was achieved, where MA-SCVP achieved after 5 views the highest overall coverage or within 1-2% from the maximum coverage achieved by volumetric information gain methods. It is worth mentioning that since MA-SCVP defines a minimal coverage view set, the local path is optimized. Because of this limitation, the coverage in the first views might be lower than that achieved by information gain approaches.

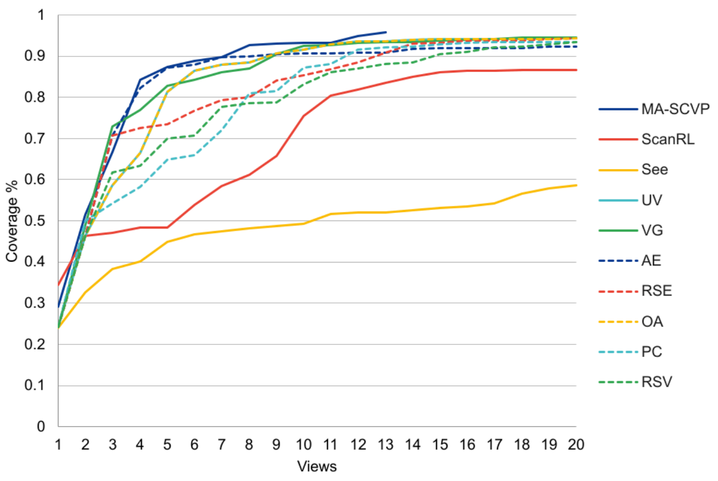

Fig. 3: Coverage for the cat model.

Th bunny model was the only model tested in which all volumetric information gain metrics as well as ScanRL achieved a higher maximum coverage than MA-SCVP. For the bunny model, MA-SCVP defined a set of 11 views, ScanRL achieved the maximum coverage in 17 views, and the volumetric information gain methods were limited to 20 views, but did not improve more than 1% in views 17 to 20. This points to the limitation of the smallest view set defined by MA-SCVP and the value of using more views to improve coverage. The SEE in the bunny model was tested up to 50 views and achieved 74.2% maximum coverage with an average 0.9% improvement per view.

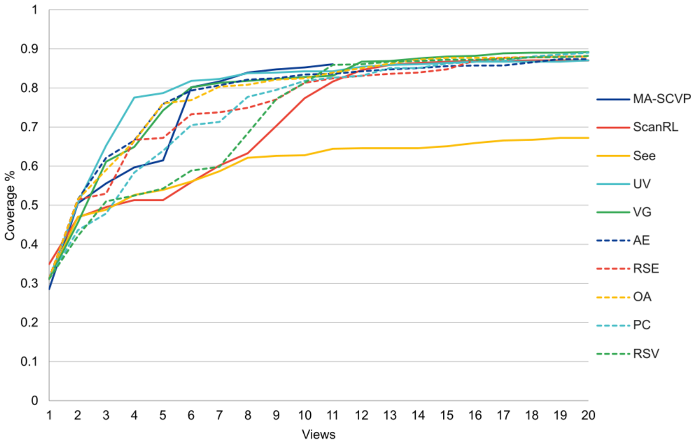

Fig. 4: Coverage for the bunny model.

## 5 Further research

As future work we define several possible directions.

Combining several NBV planning approaches to create hybrid NBV planners. The ML based MA-SCVP method is limited to a 32 view dataset. Combining its fast coverage in the initial views with a measurement-direct approach like SEE to define the next-best view based on the gaps in the resulting point cloud can lead to a higher overall coverage percentage.

ML model training on a larger dataset with models of higher geometric complexity. Most of the datasets do not include models with a high levels of occlusions and geometric complexity. Training ML models on a larger and more complex dataset can result in more robust NBV planners.

Neural networks with larger state and action spaces. Fixed view space is a limiting factor for neural network-based methods, as well as 32x32x32 voxel representation is not suitable for some objects, for example, plant leaves. Using higher-resolution state and action spaces might yield better results for more complex geometries.

Testing NBV planners on real-world objects with sensor sensor noise and positioning uncertainty.

## 6 Conclusions

In our experiments, the ML based MA-SCVP method achieved higher coverage with fewer views than other reviewed methods. By combining robotic platforms and ML NBV path planning, it is possible to optimize automated 3D asset requisition and achieve high resolution models in less amount of views.

This study emphasizes the importance of understanding the differences between NBV planners when applied to 3D reconstruction. Future research should explore hybrid methods, combine the strengths of the models discussed, and develop more adaptive strategies that can better handle a wider range of geometries.

## References

Boland, M. (2020). Artillery intelligence briefing. Accessed on 01.05.2023. `https://artillry.co/wp-content/uploads/2020/08/August-2020-ARtillery-Intelligence-Briefing.pdf`

Border, R., Gammell, J. D. (2022). The Surface Edge Explorer (SEE): A measurement-direct approach to next best view planning, *The International Journal of Robotics Research (IJRR)* . Submitted, Manuscript #IJR-22-4541, arXiv:2207.13684 [cs.RO].

Border, R., Gammell, J. D., Newman, P. (2018). Surface edge explorer (see): Planning next best views directly from 3d observations, *2018 IEEE International Conference on Robotics and Automation (ICRA)* pp. 1–8.

Chang, A. X., Funkhouser, T. A., Guibas, L. J., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F. (2015). Shapenet: An information-rich 3d model repository, *ArXiv* **abs/1512.03012**.

Curless, B., Levoy, M. (1996). A volumetric method for building complex models from range images, *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, Association for Computing Machinery, New York, NY, USA, p. 303–312. `https://doi.org/10.1145/237170.237269`

Delmerico, J., Isler, S., Sabzevari, R., Scaramuzza, D. (2018). A comparison of volumetric information gain metrics for active 3d object reconstruction, *Autonomous Robots* **42**.

Hinterstoißer, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G. R., Konolige, K., Navab, N. (2012). Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes, *Asian Conference on Computer Vision*.

Jensen, R. R., Dahl, A., Vogiatzis, G., Tola, E., Aanæs, H. (2014). Large scale multi-view stereopsis evaluation, *2014 IEEE Conference on Computer Vision and Pattern Recognition* pp. 406–413.

Kari, M. (2022). Augmented reality drives e-commerce growth. Accessed on 01.05.2023. `https://nordicgrowth.com/en/augmented-reality-drives-e-commerce-growth/`

Kaskman, R., Zakharov, S., Shugurov, I. S., Ilic, S. (2019). Homebreweddb: Rgb-d dataset for 6d pose estimation of 3d objects, *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)* pp. 2767–2776.

Koch, S., Matveev, A., Jiang, Z., Williams, F., Artemov, A., Burnaev, E., Alexa, M., Zorin, D., Panozzo, D. (2018). Abc: A big cad model dataset for geometric deep learning, *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 9593–9603.

Kriegel, S., Rink, C., Bodenmüller, T., Suppa, M. (2015). Efficient next-best-scan planning for autonomous 3d surface reconstruction of unknown objects, *Journal of Real-Time Image Processing* **10**, 611–631.

Krishnamurthy, V., Levoy, M. (1996). Fitting smooth surfaces to dense polygon meshes, *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, Association for Computing Machinery, New York, NY, USA, p. 313–324. https://doi.org/10.1145/237170.237270

Mendoza, M., Vasquez-Gomez, J. I., Taud, H., Sucar, L. E., Reta, C. (2019). Supervised learning of the next-best-view for 3d object reconstruction, *Pattern Recognit. Lett.* **133**, 224–231.

Pan, S., Hu, H., Wei, H. (2022). Scvp: Learning one-shot view planning via set covering for unknown object reconstruction, *IEEE Robotics and Automation Letters* **7**, 1463–1470.

Pan, S., Hu, H., Wei, H., Dengler, N., Zaenker, T., Bennewitz, M. (2023). One-shot view planning for fast and complete unknown object reconstruction.

Peralta, D., Casimiro, J., Nilles, A. M., Aguilar, J. A., Atienza, R., Cajote, R. (2020). Next-best view policy for 3d reconstruction, *arXiv preprint arXiv:2008.12664* .

Rodolà, E., Albarelli, A., Bergamasco, F., Torsello, A. (2013). A scale independent selection process for 3d object recognition in cluttered scenes, *International Journal of Computer Vision* **102**, 129–145.

Scott, W. R., Roth, G., Rivest, J.-F. (2003). View planning for automated three-dimensional object reconstruction and inspection, *ACM Comput. Surv.* **35**(1), 64–96. https://doi.org/10.1145/641865.641868

Turk, G., Levoy, M. (1994). Zippered polygon meshes from range images, *Proceedings of the 21st annual conference on Computer graphics and interactive techniques* .

Vasquez-Gomez, J. I., Sucar, L. E., Murrieta-Cid, R., Lopez-Damian, E. (2014). Volumetric next-best-view planning for 3d object reconstruction with positioning error, *International Journal of Advanced Robotic Systems* **11**.

Vasquez-Gomez, J. I., Troncoso, D., Becerra, I., Sucar, E., Murrieta-Cid, R. (2021). Next-best-view regression using a 3d convolutional neural network, *Machine Vision and Applications* **32**.

Zeng, R., Zhao, W., Liu, Y.-J. (2020). Pc-nbv: A point cloud based deep network for efficient next best view planning, *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* pp. 7050–7057.

Zhou, Q., Jacobson, A. (2016). Thingi10k: A dataset of 10,000 3d-printing models, *arXiv preprint arXiv:1605.04797* .

Received September 7, 2023 , revised September 30, 2024, accepted January 13, 2025