A Quality-Based Criteria for Efficient View Selection

Rémy Alcouffe, Sylvie Chambon, Géraldine Morin, Simone Gasparini

A Quality-Based Criteria for Efficient View Selection

Rémy Alcouffe[1], Sylvie Chambon[1][0000−0001−8104−1637], Géraldine Morin[1][0000−0003−0925−3277], and Simone Gasparini[1][0000−0001−8239−8005]

University of Toulouse, IRIT – Toulouse INP, France
{name.surname}@irit.fr

Abstract. The generation of complete 3D models of real-world objects is a well-known problem. The accuracy of a reconstruction can be defined as the fidelity to the original model, but in the context of the 3D reconstruction, the ground truth model is usually unavailable. In this paper, we propose to evaluate the quality of the model through local *intrinsic* metrics, that reflect the quality of the current reconstruction based on geometric measures of the reconstructed model. We then show how those metrics can be embedded in a Next Best View (NBV) framework as additional criteria for selecting optimal views that improve the quality of the reconstruction. Tests performed on simulated data and synthetic images show that using quality metrics helps the NBV algorithm to focus the view selection on the poor-quality parts of the reconstructed model, thus improving the overall quality.

Keywords: Iterative 3D Reconstruction · View Selection · Accuracy.

1 Introduction

The 3D reconstruction of real-world objects has become a key tool in revolutionising industries by converting the real world into digital shapes. 3D reconstruction can be performed using algorithms based on images [9,26,30] or depth sensors like LiDAR (Light Detection and Ranging) [12,25,29]. Although 3D reconstruction is reliable and efficient, it is not immune to imperfections: it may generate incomplete models, or models with noise, non-uniform, or inaccurate geometry [27], as illustrated in Fig. 1. Usually, they are caused by insufficient data collected by the sensor, whether because some parts of the scene are missed during the acquisition (*e.g.* because of occlusions), or because the complex shape and properties of the object prevent a complete reconstruction.

To palliate these issues, an iterative process integrating data acquisition and reconstruction can be used to strategically determine the *Next-Best-View* (NBV) [5] to enhance the 3D reconstruction of an object. NBV finds relevance in a multitude of applications spanning robotics, autonomous navigation, and exploration [28], virtual reality [21], and industrial inspection [8]. In this iterative paradigm, the next view is determined by maximising the contribution of an additional view to the quality of the reconstructed model. The contribution of a

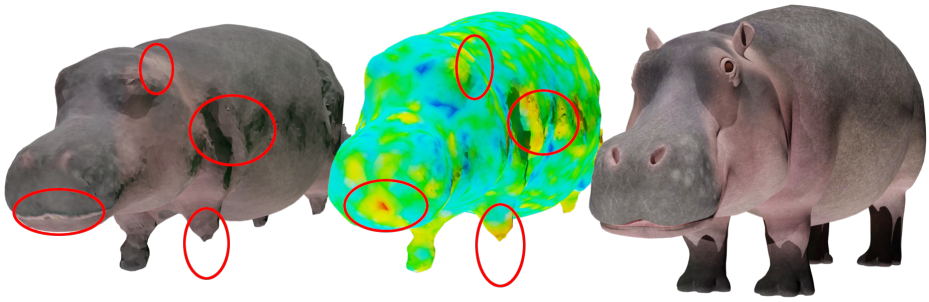


Fig. 1: Example of the assessment of a 3D reconstruction quality: the red circles highlight the poorly reconstructed parts (left model) of the original model (right model). The heat map (middle model) shows the response of the extrinsic metric MSDM2 [13], with the poorly reconstructed parts having a higher score (hot colors).

view to the reconstruction quality can be defined according to different metrics. Some methods focus on the completeness of the model, thus base their metrics on the visibility of the 3D model [5]: Massios et al. [17] control the geometry of the model via the computation of the patch normals, Li et al. [16] minimize the Mass Vector Chain and Maver et al. [18] explicitly detect possible occlusions in the partial model. Other methods compute the uncertainty of the reconstructed model and select the next view by minimizing the entropy [32] or covariance matrix of the observations [7,10]. In the optimisation step, multiple metrics may be considered simultaneously. Mendez *et al.* [19] combine uncertainty and coverage to improve the current reconstruction and the baseline vergence angle to explore new parts of the scene. The APORA algorithm [6] considers the uncertainty, the visibility, and the proximity to frontier voxels. At the same time, in [2] a utility function maximizes the information gain (*i.e.* the entropy of voxels in a given viewpoint), and the density of the final model while penalizing viewpoints that are too far away.

The accuracy of the final 3D model is usually evaluated w.r.t. a reference model (*i.e.* the ground truth) with different metrics that rely on the geometric distance (*e.g.* Hausdorff [11], chamfer [3], *accuracy* and *completeness* [27]) between the points of the reconstructed model and the ground truth model, or perceptual measures [14,13,31]. However, in general, the reference model is not available, thus making the task of assessing the accuracy a more challenging task.

In this paper, we propose to use local intrinsic quality metrics to assess the quality of a 3D model without a reference model. We show that these quality metrics can provide helpful insight to detect the regions of the model that are poorly reconstructed and thus require more acquisitions. We then show how the proposed metrics can be plugged into an existing NBV framework to guide the

acquisition process to select the views that improve and ensure the quality of the final reconstructed model.

After presenting the NBV problem in Section 2, we present a full NBV pipeline that we adapted to account for quality functions in Section 3. We review the existing metrics for the assessment of the quality of 3D models that we propose to use for the determination of the NBV as well as validation metrics in Section 4. We conduct experiments on simulated data (Section 5) and synthetic images (Section 6) data to show the advantage of using intrinsic metrics in an NBV Selection context.

2 Next-Best-View Selection

In some applications, the image acquisitions involved in reconstructing a 3D model are conducted in a series of successive steps. This means that a first set of acquisitions is made, then, the 3D model is estimated and it is decided to acquire new images to improve the 3D model. Other applications involve sensor-equipped platforms with constrained mobility [15]. As a consequence, the choice of subsequent viewpoints is crucial in efficiently improving the reconstruction by minimizing occlusions, enhancing feature visibility, and mitigating the uncertainty of the reconstruction.

As explained in the introduction, the NBV selection consists of the optimal choice of the novel viewpoint in a sequence of viewpoints from which an object or scene should be observed to optimize quality improvement of the 3D reconstruction.

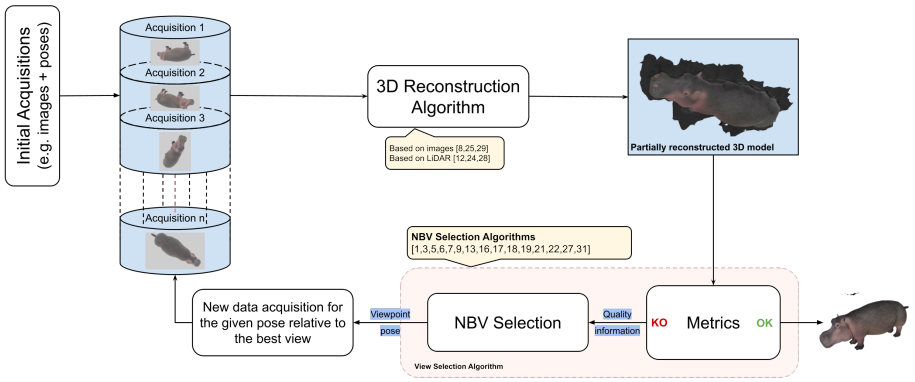


Fig. 2: NBV pipeline used in this paper.

Figure 2 presents a common NBV pipeline for an iterative reconstruction process. At each step, new data is acquired to generate the current reconstructed

model. If the model achieves a sufficient level of reconstruction, the algorithm terminates; otherwise, the metrics are then used to evaluate the contributions of novel views. Once an optimal view is selected, the corresponding data is acquired and provided to the reconstruction algorithm to generate a novel reconstruction. The NBV problem is an optimisation problem involving the greedy selection of the next viewpoint for a given metric. The core concept of this algorithm concerns the choice of the metric for selecting the best view, as discussed in the next section.

3 Selected NBV Algorithm

3.1 A Global Max-Flow Based Method

Pan *et al.* [23] proposed a new approach based on the APORA [6] algorithm. Starting from the observation that a local iterative selection may lead to poor performance in view evaluation and that the fixed resolution of the 3D volumetric occupancy map may cause sampling problems for small objects, a max-flow-based approach is proposed that tackles the problem of small object reconstruction by adding a multi-resolution dimension.

Even though this method was designed for reconstruction using a LiDAR sensor, it can be suitable for any kind of 3D reconstruction algorithm, as the max-flow term ensures a certain overlap between the acquisitions. Like in APORA [6] algorithm, Pan *et al.* [23] choose the NBV from a selected set of views around the object. Then, for each view, they compute two terms: I_{local} and I_{flow}, which will further be combined to compute the information gain.

The local view quality function is defined as :

$$I_{local}(v) = \sum_{\forall r \in R_v} \sum_{\forall x \in X_r} H(x) \times P(vis_x) \times P(obj_x) \tag{1}$$

With R_v, the set of rays from the view v, X_r the set of voxels x alongside the ray r and $H(x)$ the Shannon entropy function, applied to the uncertainty of the voxel x. $P(vis_x)$ represents the visibility of a voxel from a given view: if it is seen through a lot of voxels, the probability of being visible will decrease. $P(obj_x)$ represents the probability of a voxel being hit by a ray coming from v to be part of the object's surface. This probability relies on the fact that most of the objects are closed or nearly closed, this models the continuity of the surface of the underlying model.

To achieve global optimality, they decide to introduce a novel representation, and they model the NBV problem with a tripartite graph containing the view, the rays, and the voxels, with the term of the sum in (Eq. (1)) on the edges between rays and voxels. This representation is used to find a subset of views that covers all the voxels of the model and that maximizes the information gain. It is solved as a graph optimisation problem, using network flow modeling.

From this solution, they define the flow-network information gain I_{flow} for each view, based on their history of appearance in the solution of the max flow problem.

This measure will then be mixed up with the local information gain I_{local}, using a weighted sum:

$$I_{global}(v) = (1-\gamma) \times \frac{I_{local}(v)}{\sum_v I_{local}(v)} + \gamma \times \frac{I_{flow}(v)}{\sum_v I_{flow}(v)} \tag{2}$$

Finally, the NBV at a given iteration is defined as the view that maximizes the I_{global} function :

$$v^* = \underset{v \in V}{\operatorname{argmax}}\, I_{global}(v) \tag{3}$$

3.2 Problem adaptation for accuracy improvement

These criteria mostly focus on the improvement of the completeness of the model. As for the accuracy of the final model, this is usually solved by reducing the entropy or uncertainty of the points, which is mostly based on the number of times a point has been seen by a camera, and it is not correlated to the geometric quality of the reconstructed model. During the process of acquiring data to create a 3D model, different parts of the object usually require varying levels of detail or complexity, which in turn, requires more data to accurately capture their geometry. To overcome this challenge, we propose to integrate new metrics into the NBV process that consider the geometric quality of the reconstructed parts of the model. This approach ensures that the acquisition process is guided by two opposing objectives: completing the model and ensuring sufficient accuracy. To test our approach, we used the NBV pipeline proposed by Pan *et al.* [23] and their available code. Moreover, we wanted to show that the proposed metrics may be integrated into an existing View Selection pipeline.

We replaced their local information metric I_{local} with a quality function containing, for each voxel of the model, the value of the quality ranging between 0 (low quality) and 1 (high quality), multiplied by the visibility criterion previously used in the proposed pipeline:

$$I_{local}(v) = \sum_{\forall r \in R_v} \sum_{\forall x \in X_r} P(vis_x) \times (1 - quality) \tag{4}$$

With that novel local information function, the selected views will tend to have a greater score if they are looking to poor quality voxels.

The next section presents the metrics that can be used to take the accuracy into account and how they can be easily integrated.

4 Estimation of 3D Model quality

4.1 From Existing 3D Quality Metrics...

To assess the local quality of a 3D mesh, we perform a comparison to a reference mesh. We will refer to those metrics as *extrinsic* metrics because they rely on an

external reference model. As shown in our previous work [1], the most common extrinsic metrics [14,13,31,11] are based on the computation of the Euclidean distance between two meshes to assess their differences.

Other more complex metrics use the intrinsic parameters of both the reference mesh and the reconstructed one. DAME [31] uses the difference in dihedral angle to assess the quality of a reconstructed mesh. MSDM [14] uses several parameters such as contrast, structure, and curvature, computed on both meshes. It can produce local quality maps that are then used to generate a global quality score on the model. A second version of that metric, called MSDM2 [13], takes into account the multi-scale dimension of the problem. These two last metrics have been proven to be correlated with subjective scores.

These metrics are interesting because they provide a quantitative and qualitative map of local quality by highlighting the differences between the original and reconstructed meshes. The provided different local maps can be used as a ground truth map to estimate the quality of a reconstructed mesh and as validation metrics in our context of experiments on 3D reconstruction.

However, as the reference model is rarely available, the use of that kind of metric is not suitable to assess the quality of the reconstruction. A metric that depends only on the intrinsic geometric properties of the 3D model is therefore more appropriate. This type of metric is called *intrinsic* and will be introduced in the next section.

4.2 ...to a Quality Metric for NBV

In our previous work [1], we proposed different intrinsic metrics (*i.e.* metrics that rely only on the geometry and the intrinsic parameters of the underlying 3D model). These metrics detect the different defects and issues of a 3D model, as well as respond to sharp and salient features of the 3D object. They are suitable in the context of 3D reconstruction as they do not need a reference mesh to be computed and can be iteratively computed on the go.

Different geometric properties of a 3D model can be computed such as curvature, dihedral angles, saliency, or local roughness. They are usually used and combined to give a global quality score for the mesh. In our context, we use them to express a quality factor for each point of the 3D model. In our precedent work, we presented three intrinsic metrics suitable for the problem of quality assessment. The Plane Local Roughness (PLR) as proposed by Rodríguez-Cuenca *et al.* [24], estimates the roughness of the 3D model by evaluating the reprojection error on the least square fitting plane of the neighboring patch for every vertex. This metric examines whether the existing vertex neighborhood can be characterized as planar. This gives us information on the underlying surface of the reconstructed model. For instance, when reconstructing a 3D object, if a surface is mostly planar, the need to have a high density of points is reduced. The Quadric Local Roughness [1] metric is similar to PLR, as it estimates the roughness of the model by computing the reprojection error on the least square-fitting quadric of the neighboring patch. Those two first metrics allow us to evaluate locally the regularity of the surface that should respect a plane or a quadric. The

Mean Curvature introduced by Meynet *et al.* [20], assesses the mean curvature of a vertex by computing the derivative of the estimated least square fitting quadric of its neighborhood. As this metric is unbounded, it can not be used directly as a quality metric. However, it can give us information on high curvature areas that are either sharp features or defects of the object. These two kinds of surfaces are of interest for improving the reconstruction as they can correspond to parts of the model poorly reconstructed.

All these metrics are integrated in the NBV pipeline but we show in the first synthetic quality experiment that the metric needs to be monotonic as it will allow the camera to change positions. If a metric is not monotonic between every iteration and does not change, the NBV proposed by the algorithm will stay relatively close to the previous one as it was the best one according to the quality criterion. In that case, we need a quality function that is locally strictly monotonic. A way to achieve that is to multiply the quality factor by a new term that is correlated with the occupancy of each voxel in the Pan *et al.* [23] pipeline:

$$I_{local}(v) = \sum_{\forall r \in R_v} \sum_{\forall x \in X_r} P(vis_x) \times (1 - quality) \times H(x), \tag{5}$$

where the Shannon entropy function is

$$H(x) = -\operatorname{occ}(x) \ln(\operatorname{occ}(x)) - (1 - \operatorname{occ}(x)) \ln(1 - \operatorname{occ}(x)), \tag{6}$$

and $\operatorname{occ}(x)$ is the occupancy of the voxel x. The occupancy of the occupied voxels is a function that tends to one as the number of iterations increases.

5 Experiments on Simulated Data

5.1 Experimental Setup

To add a quality criterion to the pipeline described in Fig. 2, we need to modify the metrics used for the view selection algorithm. We introduce two new stages in the metrics estimation scheme (*cf.* Fig. 3): the first one computes the intrinsic quality of the model used to determine the NBV. The second one computes the extrinsic quality of the model to validate the choice of the view and to verify the increase of the reconstructed model quality at each iteration. To that end, it needs a reference model for the comparison and the computation of extrinsic quality metrics.

First, we need to assess whether the addition of a local quality metric to the NBV pipeline has an impact and guides the views to explore the poor quality areas of the model. To that end, we performed synthetic experiments on a chosen 3D model, with an arbitrary given quality score for each vertex.

To perform this experiment, we adapted the NBV pipeline provided by Pan *et al.* [23] as described in the Section 3. The pipeline also comes with a LiDAR emulator, that can acquire the 3D points seen by a camera from a ground truth model. For our first synthetic experiments, we used this emulator.

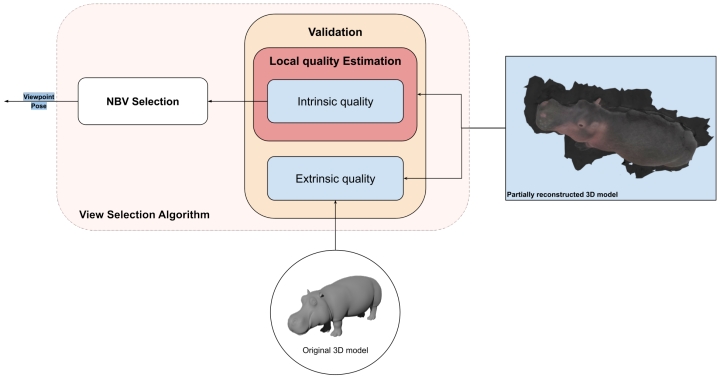


Fig. 3: A detailed view of the metric block of the global NBV pipeline. Note that the original 3D model is only used to render the images according to the selected viewpoints.

We created a small dataset composed of a single 3D geometric object (*cf*. the armadillo in Fig. 4). To simulate the quality, we defined an arbitrary metric: for each vertex of the model, we associate a quality value that is either 0.5 for vertices considered with "poor quality" and 1.0 for those of "good quality". We defined 20 different, randomly generated quality versions for that same 3D object. In each version, the positions of the quality patches are randomly chosen as well as their sizes, as can be seen in Fig. 4.

At each iteration, when performing a new acquisition, the quality term of the voxel is updated as follows:

$$qlt(v_i) = \begin{cases} qlt(v_i)^{0.9} & \text{if } v_i \in V_{seen} \\ 0.5 & \text{elsewhere.} \end{cases} \tag{7}$$

With v_i the voxel to update, $qlt(v_i)$ representing its quality in the working octomap, and V_{seen} the set of voxels that has already been seen during previous iterations. The quality of a voxel is defined to be updated only once during an iteration and not every time an acquired point falls into a voxel because it will help us in the validation process to compute the number of voxels of poor quality that have been modified at each iteration. We use the function $pow(x, 0.9)$ to ensure that the algorithm will not propose the same view at each iteration as the quality metric is not evolving on the ground truth model. We employ this function due to its asymptotic behavior approaching one as the number of iterations tends towards infinity. This update function allows us to see if the algorithm, is well suited for a quality metric that will improve itself and whether it can explore the poorly reconstructed parts of the model.

To validate the data we computed the completeness of the model at each iteration in two different ways. The first one corresponds to the number of voxels marked as *Seen* in the partially reconstructed model. The second one uses the

Fig. 4: Some examples of poor quality patches on the armadillo model (in red). The patches are placed on the model by randomly selecting a vertex and its N-ring neighbours, where N is a parameter that defines the size of the patch.

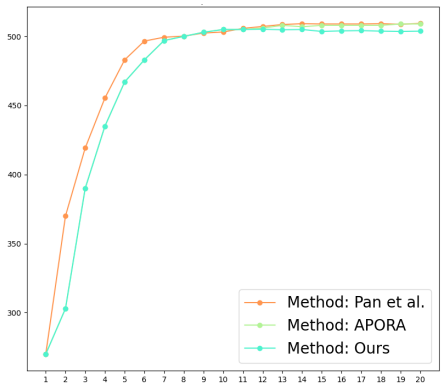
completeness defined by Seitz *et al.* [27] and is computed by comparing the partial point cloud and the original model. Then we introduced another metric that corresponds to the number of times the quality of voxels has been updated using the Eq. (7). This metric will show us if the NBV is focusing more on poor quality areas compared to standard methods.

5.2 Results and Analysis

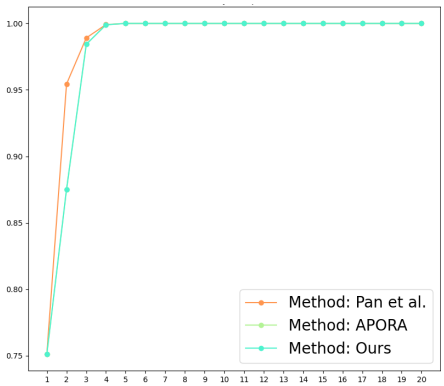
For this study, we compare the default reconstruction method proposed by Pan *et al.* [23], the APORA method proposed by Daudelin *et al.* [6], and our metric described in Section 3. We performed 20 iterations of the NBV algorithm.

As our method does not have any term that encourages the exploration of unknown areas of the object we needed to have an already complete 3D model. So if the first selected views do not see any "poor quality" voxels, our metric will show no results. To that end, for our method, the first 10 iterations will be computed using the APORA method. We chose the number of 10 iterations as the model will have sufficient completeness. We use APORA instead of the Pan *et al.* method, as it is deterministic so that different runs can be compared. Moreover, for every patch size and patch position, we will work on the same initial reconstructed model.

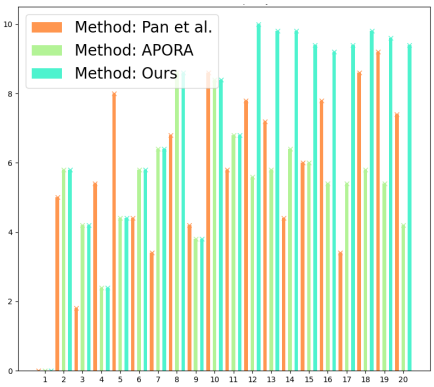
Figure 5 presents the average results on 5 different patch positions, for a given patch size. We used 10 iterations for the reconstruction of a complete model as the completeness metric and the number of *Occupied* voxels viewed at



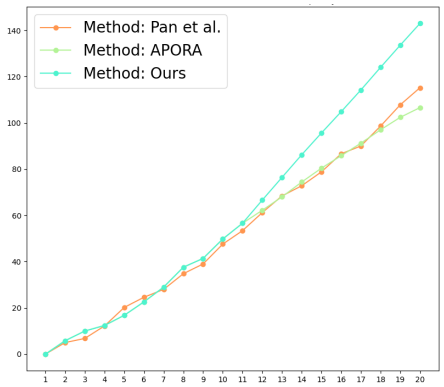
(a) Number of *Occupied* voxels viewed.



(b) Middleburry Completeness



(c) Number of voxels of bad quality that changed of state at each iteration

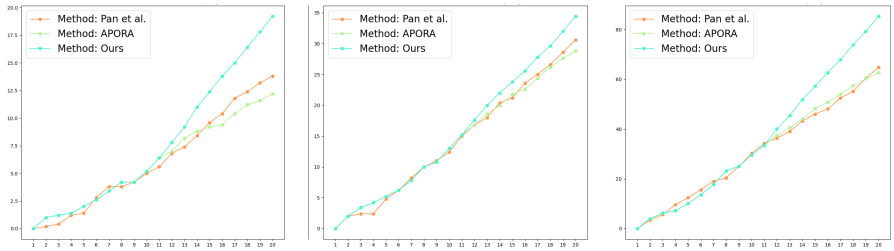


(d) Cumulative number of voxels of bad quality that changed of state

Fig. 5: Average values on different patch positions for a given patch size (here 10-ring neighborhood).

that stage seems constant. The maximum number of voxels viewed is achieved around the tenth iteration and changes very little afterward. Note that the blue and green curves (resp. APORA and Ours) are merged on the 10 first iterations as the APORA algorithm is used to perform the complete reconstruction on the first ten iterations of our method. Finally, the fourth graph, presenting the cumulative number of voxels whose quality changed during the reconstruction, shows a huge improvement in the number of "poor quality" voxels seen at each iteration. On average, for this size of the patch, the Pan *et al.* and the APORA method view only 115 poor-quality voxels while our method can view more than 140 voxels.

The same kind of results are achieved for even smaller patch sizes, see Fig. 6. We can denote that for the last ten iterations, the slope of the curve is almost equivalent to the number of voxels that are considered to have a "poor quality". This demonstrates that the algorithm selects views that show a patch of poor quality.



(a) 4-ring neighborhood (b) 6-ring neighborhood (c) 8-ring neighborhood

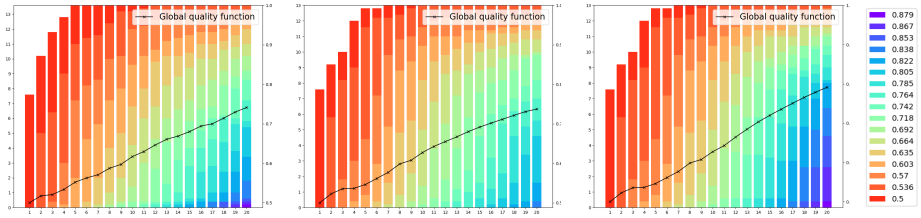
Fig. 6: Average cumulative number of voxels of "bad quality" that changed of state, on different neighborhood sizes.

Another advantage of our method is that it tends to view more times the same poor-quality patches. By using our function described in Eq. (7), and limiting the quality update of voxels to only one time per iteration, we can draw the map of the quality distribution of the octomap depicted in Fig. 7: at each iteration, it shows a stacked bar histogram of the distribution of the quality values in the octomap.

On average, our method allows us to have a wider range of quality states, as it is more likely to propose views that acquire the pre-defined poor-quality patches. We also defined a global quality function score:

$$\gamma = \sum_{v_i \in V^i_{qlt}} qlt(v_i), \tag{8}$$

where V^i_{qlt} represents the set of voxels known to have a "poor quality" at the iteration i. This function, plotted in black in Fig. 6, shows how our method,



(a) Pan *et al.* method (b) APORA method (c) Our method

Fig. 7: Average map representing the distribution of the quality values in the octomap: at each iteration (x-axis) the stacked bar histogram shows the evolution of the quality values. The colors range from red (poor quality) to purple (better quality).

designed to encourage the views containing poor quality voxels, contributes to the enhanced overall model quality compared to the other two methods.

The comparison of those 3 methods shows that there is a real interest in adding a quality criterion to an NBV pipeline to propose a novel view that will be focused on the poor-quality regions of the objects. The challenge is to apply this transformation to a real object using a real-life 3D reconstruction algorithm, with bigger uncertainty on the reconstruction and with the use of real quality metrics like the ones described in Section 4.2.

6 Experiments on Synthetic Images

6.1 Experimental Setup

We tested the proposed quality metrics in the context of a 3D reconstruction based on images. We relied on Meshroom [9], a 3D reconstruction framework for unordered sets of images. We considered textured 3D models and used Blender [4] to generate synthetic images from a given set of viewpoints. The NBV pipeline proposed by Pan *et al.* [23] was used as a reference and baseline for our comparisons. Since our goal is to show where the model quality can be improved, we used the reconstructed model obtained by Pan *et al.* after 20 iterations (that is, using 20 images) as a starting point. We call this reconstruction the *base model*; note so that the base model was mostly complete. Then, for each subsequent iteration, we ran the APORA algorithm [6] and its modified version that includes a term to take quality into account in the minimization process. The reconstruction (with Meshroom) provides the poses of the selected views. Pan *et al.* [23] focus on LiDAR for acquiring 3D but in photogrammetry more views are necessary since overlap is needed for the reconstruction. Thus, at each step, we consider the 4 best candidate views to render the relevant images and we add them to the set of input images for the reconstruction.

For this study, we changed the metric used in the NBV selection step. We used two different quality metrics presented earlier, the *Quadric Local Roughness* and the *Plane Local Roughness*. We used the mean version of them and they will be computed using their 50-nearest neighbors. Those metrics have a strong response on the areas of the objects that are considered as a default, but they also can have a high response regarding the geometry of the object as they are also sensitive to sharp and salient structures. While the response of the metrics in areas of poor quality vanishes if they are corrected, that is not the case for the areas with interesting geometry. To tackle this problem we propose to add the Shannon entropy term to the Local Information for the NBV Selection in Eq. (5). This term helps reduce the importance of the voxels marked as "poor quality" because they are just responding to the geometry of the model. The Shannon entropy $H(x)$ defined in Eq. (5) goes to 0 if the point has been seen many times (by many different views). Thus it will help to differentiate the metrics response corresponding to the sharp and salient features from the response corresponding to its defaults.

Finally, extrinsic metrics that we described in Section 4.1 are used for validation, to evaluate the quality of the reconstruction relative to ground truth. During our whole process, we will be able to follow the evolution of the quality metrics we use for the reconstruction, as well as other global and/or local *extrinsic* quality metrics such as Seitz *et al.* completeness and accuracy [27], Hausdorff distance [11], MSDM [14] and MSDM2 [13]. An example of local maps is given in Figure 8. The Seitz *et al.* [27] metrics are defined using a certain threshold. For our study, we used an accuracy ratio of 99% and a completeness distance corresponding to 1% of the diagonal of the object. Those thresholds are restrictive, but allow us to better identify improvement in accuracy.

6.2 Results and Analysis

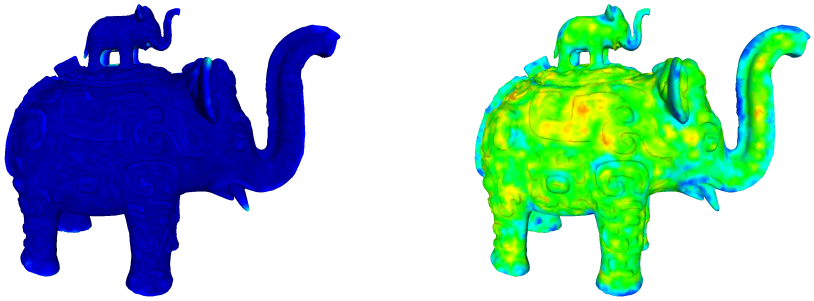
For our experiments, we have used a 3D model representing an Elephant from the Texture Quality Assessment Dataset [22]. This model has complex shapes but also has a good texture which is essential for the photogrammetry reconstruction pipeline.

From the base model, generated by 20 iterations of Pan *et al.* [23], we computed at each iteration the quality of the mesh, using either the *PLR* or *QLR*. For each iteration, the quality of the mesh is mapped to the octomap using a min function:

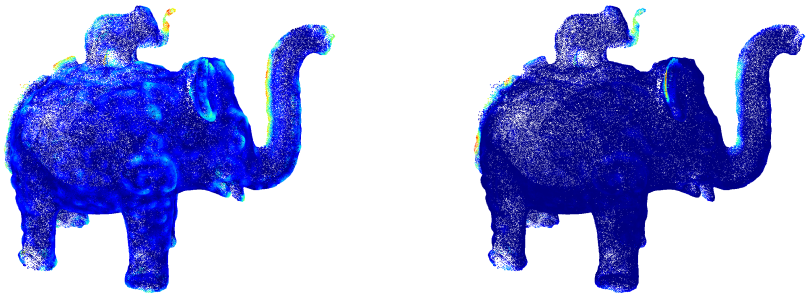
$$qlt_v(v) = min(qlt_p(x),\ \forall\ x\ \in\ v \subset X), \tag{9}$$

with $qlt_v(v)$ designing the quality of the voxel v in the octomap, $qlt_p(x)$ the quality of the point x in the reconstructed mesh, and X is the set of points of the reconstructed mesh. This function takes into account the fact that each voxel of the octomap can contain multiple points of the mesh. Since the points falling into one voxel might come from different parts of the mesh, the *min* function is to be preferred to *e.g. mean* and *median* to propagate the poor quality region in the octomap. An example of the octomap is shown in Fig. 9,

(a) Original model (b) Reconstructed model (c) Geometry



(d) Hausdorff distance (e) MSDM2 local map



(f) Mean PLR (g) Mean QLR

Fig. 8: Example of local maps for extrinsic and intrinsic metrics for the reconstructed model.

with the poor-quality voxel highlighted in red. Figure 9b shows the PLR metric has high responses on the back, on the trunk, and the back feet of the elephant. The views selected by the algorithm (*cf*. Figure 9.c-f) maximise the number of poor-quality voxels viewed, as they can be seen in the different selected views (trunk, feet, and back).

The choice of the views selected by the modified algorithm taking into account the quality shows an improvement of the accuracy of the model. For validation, Figure 10 presents the results of extrinsic metrics on the Elephant model, for the three considered methods, a basic method using the APORA algorithm, our method using the *Plane Local Roughness* and using the *Quadric Local Roughness*.

The methods embedding the quality term achieve a similar reconstruction completeness after 15 iterations, where 87% of points are at a distance less than 1% of the object diagonal. It is also worth noting that the methods with quality reach better completeness for the first 5 iterations, thus showing that the quality metrics also help for the completeness of the model.

For the accuracy of the reconstructed model (second graph from the left of Figure 10), the values have a similar evolution as the APORA method and are slightly lower (better accuracy). This is corroborated by the Hausdorff distance metric (middle graph of Figure 10), where the methods with the quality term converge more quickly to a threshold value. This value is mainly due to the poor reconstruction of the little elephant on the back of the big one, where the space between his legs is not well reconstructed.

Finally, the *Mesh Structural Distortion Measures* (last two graphs of Figure 10) confirm that the methods with the quality term have in general better results.

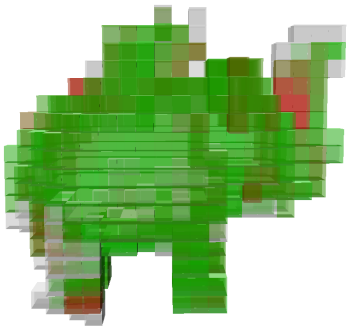
This shows that integrating the proposed quality metrics in an NBV pipeline provides a better selection of views. In the case of photogrammetry, both the accuracy and the completeness of the reconstructed model are improved.

The proposed quality metrics have, in general, a lower response on zones with high-density reconstructed points as they are computed on a 50-nearest-neighbor of each point. Hence they guide the view selection to focus on low-density areas of the object, which often correspond to poorly reconstructed zones of the object.

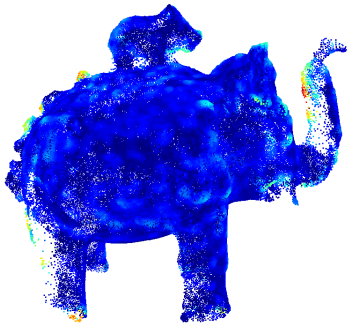
7 Conclusion

In this paper, we proposed the use and addition of local *intrinsic* metrics in the NBV Selection pipelines. In the first simulated experiment, we showed that using those metrics tends to encourage the views around the poor-quality regions of a 3D object. Then, we decided to test this method on real data, with an external photogrammetry reconstruction algorithm, using generated views from a reference 3D model. When using the local *intrinsic* metrics *PLR* and *QLR*, we show that it helps improve the accuracy of the model alongside its completeness and MSDM measures.

The metrics we proposed can detect the defaults of the 3D model, as well as the sharp and salient features of the object; The response of those metrics



(a) Octomap Quality



(b) Mean PLR on the reconstructed mesh

(c) First view selected

(d) Second view selected

(e) Third view selected

(f) Fourth view selected

Fig. 9: Local map of the mean PLR metric and its corresponding octomap representation, followed by the associated selected views.

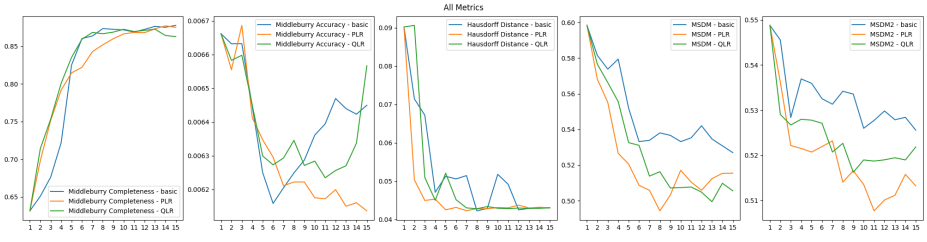


Fig. 10: The extrinsic metric response on 15 iterations for the 3 methods tested.

is defined according to a certain neighborhood, we can imagine developing a multiscale version of those metrics (like in MSDM2 [13]) to take into account the different sizes of features of the objects.

In the future, we would like to combine different metrics to be able to differentiate between salient features of the objects and actual defaults. This paper opens the way to the definition of new intrinsic quality metrics for NBV Selection algorithm, as it shows the interest of those metrics in a reconstruction process.

References

1. Alcouffe, R., Gasparini, S., Morin, G., Chambon, S.: Blind Quality of a 3D Reconstructed Mesh. In: 29th IEEE International Conference on Image Processing (ICIP 2022). pp. 3406 – 3410. IEEE, Bordeaux, France (Oct 2022). https://doi.org/10.1109/ICIP46576.2022.9897783
2. Almadhoun, R., Abduldayem, A., Taha, T., Seneviratne, L., Zweiri, Y.: Guided next best view for 3D reconstruction of large complex structures. Remote Sensing **11**(20), 2440 (2019). https://doi.org/10.3390/rs11202440, https://www.mdpi.com/2072-4292/11/20/2440
3. Borgefors, G.: Distance transformations in arbitrary dimensions. Computer Vision, Graphics, and Image Processing **27**(3), 321–345 (Sep 1984). https://doi.org/10.1016/0734-189x(84)90035-5
4. Community, B.O.: Blender - a 3d modelling and rendering package (2018), http://www.blender.org
5. Connolly, C.: The determination of next best views. In: IEEE International Conference on Robotics and Automation, ICRA. vol. 2, pp. 432–435 (1985)
6. Daudelin, J., Campbell, M.: An adaptable, probabilistic, next-best view algorithm for reconstruction of unknown 3-d objects. IEEE Robotics and Automation Letters (2017). https://doi.org/10.1109/LRA.2017.2660769
7. Dunn, E., Frahm, J.M.: Next best view planning for active model improvement. In: British Machine Vision Conference, BMVC (2009). https://doi.org/10.5244/C.23.53
8. Faria, M., Maza, I., Viguria, A.: Applying frontier cells based exploration and lazy theta* path planning over single grid-based world representation for autonomous inspection of large 3d structures with an UAS. Journal of Intelligent & Robotic Systems (2019). https://doi.org/10.1007/s10846-018-0798-4
9. Griwodz, C., Gasparini, S., Calvet, L., Gurdjos, P., Castan, F., Maujean, B., De Lillo, G., Lanthony, Y.: AliceVision Meshroom: An Open-Source 3D Reconstruction Pipeline. In: ACM Multimedia Systems Conference. p. 241–247 (2021). https://doi.org/10.1145/3458305.3478443
10. Haner, S., Heyden, A.: Covariance propagation and next best view planning for 3d reconstruction. In: European Conference on Computer Vision, ECCV. pp. 545–556. Lecture Notes in Computer Science, Springer (2012). https://doi.org/10.1007/978-3-642-33709-3_39
11. Hausdorff, F.: Set theory. Chelsea Publishing Company, 2nd edn. (1962)
12. Jun, C., Shiguang, Z., Xinyu, W.: Structured light-based shape measurement system. Signal Processing **93**(6), 1435–1444 (2013)
13. Lavoué, G.: A multiscale metric for 3d mesh visual quality assessment. Computer Graphics Forum **30**, 1427–1437 (2011). https://doi.org/10.1111/j.1467-8659.2011.02017.x

14. Lavoué, G., Drelie Gelasca, E., Dupont, F., Baskurt, A., Ebrahimi, T.: Perceptually driven 3d distance metrics with application to watermarking. In: SPIE Optics + Photonics. p. 63120L (2006). https://doi.org/10.1117/12.686964
15. Lee, I., Seo, J., Kim, Y., Choi, J., Han, S., Yoo, B.: Automatic pose generation for robotic 3-d scanning of mechanical parts. IEEE Transactions on Robotics (2020). https://doi.org/10.1109/TRO.2020.2980161
16. Li, Y., He, B., Bao, P.: Automatic view planning with self-termination in 3d object reconstructions. Sensors and Actuators A: Physical **122**(2) (2005). https://doi.org/10.1016/j.sna.2005.06.003
17. Massios, N.A., Fisher, R.B.: A best next view selection algorithm incorporating a quality criterion. In: British Machine Vision Conference, BMVC (1998). https://doi.org/10.5244/C.12.78
18. Maver, J., Bajcsy, R.: Occlusions as a guide for planning the next view. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI (1993). https://doi.org/10.1109/34.211463
19. Mendez, O., Hadfield, S., Pugeault, N., Bowden, R.: Next-best stereo: Extending next-best view optimisation for collaborative sensors. In: British Machine Vision Conference, BMVC (2016). https://doi.org/10.5244/C.30.65
20. Meynet, G., Digne, J., Lavoué, G.: PC-MSDM: A quality metric for 3d point clouds. In: International Conference on Quality of Multimedia Experience, QoMEX (2019). https://doi.org/10.1109/QoMEX.2019.8743313
21. Mortezapoor, S., Schönauer, C., Rüggeberg, J., Kaufmann, H.: Photogrammabot: An autonomous ROS-based mobile photography robot for precise 3d reconstruction and mapping of large indoor spaces for mixed reality. In: IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops, VRW (2022). https://doi.org/10.1109/VRW55335.2022.00033
22. Nehmé, Y., Dupont, F., Farrugia, J., Le Callet, P., Lavoué, G.: Textured mesh quality assessment: Large-scale dataset and deep learning-based quality metric. ACM Transactions on Graphics (2023). https://doi.org/10.1145/3592786, https://doi.org/10.1145/3592786
23. Pan, S., Wei, H.: A global max-flow-based multi-resolution next-best-view method for reconstruction of 3D unknown objects. IEEE Robotics and Automation Letters (2022). https://doi.org/10.1109/LRA.2021.3132430
24. Rodríguez-Cuenca, B., García-Cortés, S., Ordóñez, C., Alonso, M.C.: A study of the roughness and curvature in 3D point clouds to extract vertical and horizontal surfaces. In: IEEE International Geoscience and Remote Sensing Symposium, IGARSS. pp. 4602–4605 (2015). https://doi.org/10.1109/IGARSS.2015.7326853
25. Roldao, L., de Charette, R., Verroust-Blondet, A.: 3d surface reconstruction from voxel-based lidar data. In: IEEE Intelligent Transportation Systems Conference, ITSC. pp. 2681–2686 (2019). https://doi.org/10.1109/ITSC.2019.8916881
26. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2016)
27. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR. pp. 519–528 (2006)
28. Selin, M., Tiger, M., Duberg, D., Heintz, F., Jensfelt, P.: Efficient autonomous exploration planning of large-scale 3-d environments. IEEE Robotics and Automation Letters (2019). https://doi.org/10.1109/LRA.2019.2897343

29. Tachella, J., Altmann, Y., Mellado, N., McCarthy, A., Tobin, R., Buller, G., Tourneret, J., McLaughlin, S.: Real-time 3D reconstruction from single-photon lidar data using plug-and-play point cloud denoisers. Nature Communications **10**(1), 4984 (2019). https://doi.org/10.1038/s41467-019-12943-7
30. Tingdahl, D., Van Gool, L.: A public system for image based 3D model generation. In: International Conference on Computer Vision / Computer Graphics Collaboration Techniques and Applications, MIRAGE. vol. 6930/2011, pp. 262–273 (2011)
31. Váša, L., Rus, J.: Dihedral angle mesh error: a fast perception correlated distortion measure for fixed connectivity triangle meshes. Computer Graphics Forum **31**(5), 1715–1724 (2012)
32. Wenhardt, S., Deutsch, B., Angelopoulou, E., Niemann, H.: Active visual object reconstruction using d-, e-, and t-optimal next best views. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2007). https://doi.org/10.1109/CVPR.2007.383363